Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 01: Fourier, heat diffusion, and shift-invariant operators

January 14, 2020

Lecturer: Matthew Hirn

1 Sparse Representations

Chapter 1 of A Wavelet Tour of Signal Processing [1].

Exercise 1. Read Chapter 1 (*Sparse Representations*) of *A Wavelet Tour of Signal Processing*. It gives a nice overview of the book and will give you a good perspective on computational harmonic analysis heading into the course.

Exercise 2. Read the appendices in *A Wavelet Tour of Signal Processing*, as we will not cover these in class. We will immediately need some of the material contained in them.

Remark 1.1. The integral we use in this course will be the Lebesgue integral, which is usually taught in a first year graduate course in real analysis. However, if these are unfamiliar to you, you may replace most if not all of the results with Riemann integrals from Calculus and assume that the generic functions f, g, h, etc. are Schwartz class functions. For more details on the Schwartz class and Fourier integrals, see [2].

2 The Fourier Kingdom

Chapter 2 of A Wavelet Tour of Signal Processing [1].

2.1 Linear time-invariant filtering

Section 2.1 of A Wavelet Tour of Signal Processing [1].

Fourier analysis originates with the work of Joseph Fourier, who was studying the heat equation:

$$\begin{aligned} \partial_t F &= \Delta F \\ F(u,0) &= f(u) \end{aligned}$$

Where $F : \mathbb{R}^d \times [0,\infty) \to \mathbb{R}$ and $f : \mathbb{R}^d \to \mathbb{R}$. This is a linear partial differential equation. In order to solve it, it helps to think about linear algebra. Suppose $\mathbf{A}$ is an $n \times n$ real valued, symmetric matrix, which maps vectors $x \in \mathbb{R}^n$ to other vectors $\mathbf{A}x \in \mathbb{R}^n$. Then from the

spectral theorem, we know that $\mathbf{A}$ has a complete set of orthonormal eigenvectors, $v_1, \ldots, v_n$, such that

$$\mathbf{A}v_k = \lambda_k v_k$$

for some $\lambda_k \in \mathbb{R}$. Since $\{v_k\}_{k \leq n}$ forms an ONB, it allows us to write, for any $x \in \mathbb{R}^n$,

$$x = \sum_{k=1}^{n} \langle x, v_k \rangle v_k \,,$$

which in turn makes evaluating $\mathbf{A}x$ very easy:

$$\mathbf{A}x = \sum_{k=1}^{n} \langle x, v_k \rangle \mathbf{A}v_k = \sum_{k=1}^{n} \lambda_k \langle x, v_k \rangle v_k \,.$$

Let us now try to apply the same ideas to the Laplacian, Δ. We may ask, what are the eigenfunctions of the Laplacian? If we consider complex valued functions, one can verify that

$$\Delta e^{i\omega \cdot u} = -|\omega|^2 e^{i\omega \cdot u}$$

for any $\omega \in \mathbb{R}^d$. Thus the function $e_\omega(u) = e^{i\omega \cdot u}$ is an eigenfunction of Δ for any ω. Let us formally define

$$\widehat{f}(\omega) = \langle f, e_\omega \rangle = \int_{\mathbb{R}^d} f(u) e^{-i\omega \cdot u} \, du \,.$$

This will be what we call the Fourier transform, but right now we see it as an analogue of basis coefficients in an ONB. Following the analogy, we may then be tempted to write:

$$f(u) = \int_{\mathbb{R}^d} \langle f, e_\omega \rangle e_\omega(u) \, d\omega = \int_{\mathbb{R}^d} \widehat{f}(\omega) e^{i\omega \cdot u} \, d\omega \,.$$

We this in hand, we then propose

$$F(u,t) = \int_{\mathbb{R}^d} e^{-|\omega|^2 t} \widehat{f}(\omega) e^{i\omega \cdot u} \, d\omega \,,$$

as the solution to the heat equation. One can verify that, formally, F indeed is the solution. Fourier analysis was then born by trying to understanding when all of this makes mathematical sense.

The reason Fourier analysis is used so often in signal processing, is that it turns out this analysis is not useful for just the Laplacian operator. In fact the Laplacian is just an example of a more general class operators, called *shift invariant* operators. Let us now work over $\mathbb{R}$ instead of $\mathbb{R}^d$; we will use t to denote a value in $\mathbb{R}$, since it is often useful to think of it as time. Let $f_\tau(t) = f(t - \tau)$ be the translation of f by τ; if t is time, then this is a time delay by τ. An operator L is shift invariant if it commutes with the time delay of any function,

$$(Lf_\tau)(t) = (Lf)(t - \tau)$$

As we shall see all linear, continuous shift invariant operators L are diagonalized by the complex exponentials $e_\omega(t) = e^{i\omega t}$. To see this, recall the *convolution* of two functions f, g:

$$f * g(t) = \int_{\mathbb{R}} f(u)g(t-u)\,du$$

Now let $\delta(t)$ be a Dirac (centered at zero), and $\delta_u(t) = \delta(t-u)$ be a Dirac centered at u. By definition this means $f * \delta(t) = f(t)$. We have:

$$f(t) = f * \delta(t) = \int_{\mathbb{R}} f(u)\delta(t-u)\,du = \int_{\mathbb{R}} f(u)\delta_u(t)\,du$$

Since L is continuous and linear,

$$Lf(t) = \int_{\mathbb{R}} f(u)L\delta_u(t)\,du$$

Let h be the impulse response of L, defined as

$$h(t) = L\delta(t)$$

Since L is shift invariant, we have

$$L\delta_u(t) = h(t-u)$$

and therefore

$$Lf(t) = \int_{\mathbb{R}} f(u)h(t-u)\,du = f * h(t) = h * f(t)$$

Thus *every* continuous, linear shift invariant operator is equivalent to a convolution with an impulse response h.

We can now use this fact to show our original goal, which was that the complex exponential functions $e_\omega(t) = e^{i\omega t}$ diagonalize L. This will in turn motivate the study of Fourier integrals. We have:

$$Le_\omega(t) = \int_{\mathbb{R}} h(u)e^{i\omega(t-u)}\,du = e^{it\omega}\underbrace{\int_{\mathbb{R}} h(u)e^{-i\omega u}\,du}_{\widehat{h}(\omega)} = \widehat{h}(\omega)e_\omega(t).$$

Thus $e_\omega(t)$ is an eigenfunction of L with eigenvalue $\widehat{h}(\omega)$, if $\widehat{h}(\omega)$ exists. The value $\widehat{h}(\omega)$ is the *Fourier transform* of h at the frequency ω. Since the functions $e_\omega(t) = e^{i\omega t}$ are the eigenfunctions of shift invariant operators, we would like to decompose any function f as a sum or integral of these functions. This will then allow us to write Lf directly in terms of the eigenvalues of L (as you do in linear algebra when you are able to diagonalize a matrix/operator on a finite dimensional vector space). We'll try to undestand when this is possible.

Exercise 3. Read Section 2.1 of *A Wavelet Tour of Signal Processing.*

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 02: The Fourier transform on $\mathbf{L}^1(\mathbb{R})$

January 14, 2020

Lecturer: Matthew Hirn

2.2 Fourier integrals

Section 2.2 of A Wavelet Tour of Signal Processing [1].

The Fourier transform is an operator $\mathcal{F}$ that maps a function $f(u)$ to another function $\widehat{f}(\omega)$, which is defined as:

$$\mathcal{F}(f)(\omega) = \widehat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{-i\omega t}\, dt \tag{1}$$

We will start by trying to understand what restrictions we need to place on f in order for this to make sense. In particular, if f is in some well defined space of functions, we will ask, does that imply $\widehat{f}$ is in some other well defined space of functions? We will start by considering the $\mathbf{L}^p$ spaces of functions. To that end, define:

$$\mathbf{L}^p(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} : \int_{\mathbb{R}} |f(t)|^p\, dt < +\infty \right\}, \quad 0 < p < \infty$$

The space $\mathbf{L}^p(\mathbb{R})$ is a Banach space with norm:

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(t)|^p\, dt \right)^{\frac{1}{p}}$$

The space $\mathbf{L}^2(\mathbb{R})$ is special, as it is in fact a Hilbert space with inner product

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t) g^*(t)\, dt$$

where we use $g^*(t)$ to denote the complex conjugate of $g(t)$. We also define $\mathbf{L}^\infty(\mathbb{R})$. Set:

$$\|f\|_\infty = \operatorname*{ess\,sup}_{t \in \mathbb{R}} |f(t)|$$

The value $\|f\|_\infty$ is the smallest number M, $0 \leq M \leq +\infty$, such that $|f(t)| \leq M$ for almost every $t \in \mathbb{R}$; if f is continuous, it is the smallest number M such that $|f(t)| \leq M$ for all $t \in \mathbb{R}$. It thus measures whether f is bounded or not. The space $\mathbf{L}^\infty(\mathbb{R})$ is the space of bounded functions:

$$\mathbf{L}^\infty(\mathbb{R}) = \{ f : \|f\|_\infty < +\infty \}$$

We then have:

Proposition 2.1. *If $f \in \mathbf{L}^1(\mathbb{R})$, then $\widehat{f} \in \mathbf{L}^\infty(\mathbb{R})$.*

Proof. Suppose $f \in \mathbf{L}^1(\mathbb{R})$. We have:

$$|\widehat{f}(\omega)| = \left| \int_{\mathbb{R}} f(t) e^{-i\omega t}\, dt \right| \leq \int_{\mathbb{R}} |f(t) e^{-i\omega t}|\, dt = \int_{\mathbb{R}} |f(t)|\, dt = \|f\|_1 < +\infty$$

□

Proposition 2.1 shows that $\mathcal{F} : \mathbf{L}^1(\mathbb{R}) \to \mathbf{L}^\infty(\mathbb{R})$ is a well defined map using the definition (1). Later on we will extend the Fourier transform to other $\mathbf{L}^p$ spaces for $1 \leq p \leq 2$, with particular interest in $\mathbf{L}^2(\mathbb{R})$. For now recall from Section 2.1 that we would like to write $f(t)$ in terms of $\widehat{f}(\omega)$. This requires a Fourier inversion formula. However, the above proposition only guarantees that $\widehat{f} \in \mathbf{L}^\infty(\mathbb{R})$, which will not help with convergence issues. We thus assume that $\widehat{f} \in \mathbf{L}^1(\mathbb{R})$ as well.

Theorem 2.2 (Fourier inversion)**.** *If $f \in \mathbf{L}^1(\mathbb{R})$ and $\widehat{f} \in \mathbf{L}^1(\mathbb{R})$ then*

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\omega) e^{i\omega t}\, d\omega, \quad \text{for almost every } t \in \mathbb{R} \tag{2}$$

To prove this theorem, we will need three standard results from graduate real analysis. We state them here, without proof.

Theorem 2.3. *Suppose $\{f_n\}_{n\in\mathbb{N}}$ converges to f in $\mathbf{L}^p$, meaning that*

$$\lim_{n\to\infty} \|f_n - f\|_p = 0$$

Then there exists a subsequency $\{f_{n_k}\}_{k\in\mathbb{N}}$ that converges to f almost everywhere,

$$\lim_{k\to\infty} f_{n_k}(t) = f(t) \quad \text{for almost every } t$$

Theorem 2.4 (Dominated Convergence Theorem)**.** *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions such that $\lim_{n\to\infty} f_n = f$. If*

$$\forall\, n \in \mathbb{N}, \quad |f_n(t)| \leq g(t) \quad \text{and} \quad \int_{\mathbb{R}} g(t)\, dt < +\infty$$

then $f \in \mathbf{L}^1(\mathbb{R})$ and

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(t)\, dt = \int_{\mathbb{R}} f(t)\, dt$$

Theorem 2.5 (Fubini's Theorem)**.** *Let $f(u,t)$ be a function of two variables $(u,t) \in \mathbb{R}^2$. If*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(u,t)|\, du \right) dt < +\infty$$

then

$$\begin{aligned}\iint_{\mathbb{R}^2} f(u,t)\,du\,dt &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(u,t)\,du\right)\,dt \\ &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(u,t)\,dt\right)\,du\end{aligned}$$

Proof of Theorem 2.2. Now we turn to the proof. Plugging in the formula of $\widehat{f}(\omega)$ into the right hand side of (2) yields

$$\frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega)e^{i\omega t} = \frac{1}{2\pi}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(u)e^{i\omega(t-u)}\,du\right)\,d\omega$$

However we cannot apply Fubini directly because the function $F(u,\omega) = f(u)e^{i\omega(t-u)}$ is not integrable in $\mathbb{R}^2$. Therefore we instead consider the following integral:

$$I_\varepsilon(t) = \frac{1}{2\pi}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(u)e^{-\varepsilon^2\omega^2/4}e^{i\omega(t-u)}\,du\right)\,d\omega$$

The Gaussian yields a new integrand $F_\varepsilon(u,\omega) = f(u)e^{-\varepsilon^2\omega^2/4}e^{i\omega(t-u)}$ which is integrable on $\mathbb{R}^2$, and for which $\lim_{\varepsilon\to 0} F_\varepsilon = F$. We can thus apply the Fubini theorem to $I_\varepsilon(t)$; we do so in two ways. For the first, we integrate with respect to u, giving:

$$I_\varepsilon(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega)e^{-\varepsilon^2\omega^2/4}e^{i\omega t}\,d\omega.$$

Since

$$\left|\widehat{f}(\omega)e^{-\varepsilon^2\omega^2/4}e^{i\omega t}\right| \le |\widehat{f}(\omega)|$$

and since $\widehat{f} \in \mathbf{L}^1(\mathbb{R})$, we can apply the dominated convergence theorem to obtain:

$$\lim_{\varepsilon\to 0} I_\varepsilon(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega)e^{i\omega t}\,d\omega \tag{3}$$

Now compute $I_\varepsilon(t)$ a second way by applying the Fubini theorem and integrating with respect to ω. We get that

$$I_\varepsilon(t) = \int_{\mathbb{R}} g_\varepsilon(t-u)f(u)\,du = f * g_\varepsilon(t)$$

where

$$\begin{aligned}g_\varepsilon(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\varepsilon^2\omega^2/4}e^{ix\omega}\,d\omega \\ &= \frac{1}{\varepsilon\sqrt{\pi}}\int_{\mathbb{R}} \frac{\varepsilon}{2\sqrt{\pi}}e^{-\varepsilon^2\omega^2/4}e^{ix\omega}\,d\omega \\ &= \frac{1}{\varepsilon\sqrt{\pi}}e^{-x^2/\varepsilon^2}\end{aligned}$$

To get the last line, we used the fact that that the Fourier transform of $\theta(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2}$ is equal to $\widehat{\theta}(\xi) = e^{-\sigma^2\xi^2/2}$. This is a useful identity that you should verify yourself and then remember. Another useful identity is that $\int_{\mathbb{R}} \theta(t)\, dt = 1$. From this latter formula we deduce that

$$\int_{\mathbb{R}} g_\varepsilon(x)\, dx = 1$$

Furthermore, we notice that

$$g_\varepsilon(x) = \varepsilon^{-1} g_1(\varepsilon^{-1}x), \quad g_1(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} \tag{4}$$

Thus the family $\{g_\varepsilon\}_{\varepsilon>0}$ is an approximate identity. For general approximate identities one can prove (see below):

$$\lim_{\varepsilon\to 0} \|f * g_\varepsilon - f\|_1 = 0$$

We now apply Theorem 2.3 to infer there exists a subsequence $\{f * g_{\varepsilon_k}\}_{k\in\mathbb{N}}$ with $\varepsilon_k \to 0$ as $k \to \infty$ such that $\lim_{k\to\infty} f * g_{\varepsilon_k} = f$ almost everywhere. On the other hand, using (3) we have

$$\frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega) e^{i\omega t}\, d\omega = \lim_{k\to\infty} I_{\varepsilon_k}(t) = \lim_{k\to\infty} f * g_{\varepsilon_k}(t) = f(t) \quad \text{for almost every } t$$

thus completing the proof. □

To complete the above proof of Theorem 2.2, we introduce the notion of an *approximate identity*. A family of functions $\{k_\lambda\}_{\lambda>0}$ is an approximate identity if:

1. Normalized: $\int_{\mathbb{R}} k_\lambda(t)\, dt = 1$ for every $\lambda > 0$
2. $\mathbf{L}^1$-boundedness: $\sup_{\lambda>0} \|k_\lambda\|_1 < \infty$
3. $\mathbf{L}^1$-concentration: For every $\delta > 0$,

$$\lim_{\lambda\to 0} \int_{|t|\geq\delta} |k_\lambda(t)|\, dt = 0$$

One can verify that the family of Gaussian functions $\{g_\varepsilon\}_{\varepsilon>0}$ from (4) is an approximate identity. More generally, one often constructs an approximate identity by dilating a single function $k \in \mathbf{L}^1(\mathbb{R})$ satisfying $\|k\|_1 = 1$, that is by setting $k_\lambda(t) = \lambda^{-1}k(\lambda^{-1}t)$. The following theorem is useful on its own, and completes the proof of Theorem 2.2.

Theorem 2.6. *Let $\{k_\lambda\}_{\lambda>0}$ be an approximate identity. Then*

$$\forall\, f \in \mathbf{L}^1(\mathbb{R}), \quad \lim_{\lambda\to 0} \|f - f * k_\lambda\|_1 = 0$$

To prove Theorem 2.6 we will need the following standard result from real analysis.

Theorem 2.7 ($\mathbf{L}^p$ continuity)**.** *Let $1 \leq p < \infty$. If $f \in \mathbf{L}^p(\mathbb{R})$, then*

$$\lim_{\tau \to 0} \|f - f_\tau\|_p = 0$$

where we recall $f_\tau(t) = f(t - \tau)$. By the definition of limit, this means for each $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|\tau| < \delta \Longrightarrow \|f - f_\tau\|_p < \varepsilon$$

Proof of Theorem 2.6. Since $f, k_\lambda \in \mathbf{L}^1(\mathbb{R})$, one can show that $\|f * k_\lambda\|_1 \leq \|f\|_1 \|k_\lambda\|_1 < \infty$, which means that $f * k_\lambda \in \mathbf{L}^1(\mathbb{R})$. Using the fact that $\int_{\mathbb{R}} k_\lambda = 1$ for all $\lambda > 0$, we have:

$$\begin{aligned}
\|f - f * k_\lambda\|_1 &= \int_{\mathbb{R}} |f(t) - f * k_\lambda(t)| \, dt \\
&= \int_{\mathbb{R}} \left| f(t) \int_{\mathbb{R}} k_\lambda(u) \, du - \int_{\mathbb{R}} f(t-u) k_\lambda(u) \, du \right| dt \\
&= \iint_{\mathbb{R}^2} |f(t) - f(t-u)| |k_\lambda(u)| \, du \, dt \\
&= \iint_{\mathbb{R}^2} |f(t) - f(t-u)| |k_\lambda(u)| \, dt \, du \quad \text{[Tonelli]} \\
&= \int_{\mathbb{R}} |k_\lambda(u)| \int_{\mathbb{R}} |f(t) - f(t-u)| \, dt \, du \\
&= \int_{\mathbb{R}} |k_\lambda(u)| \|f - f_u\|_1 \, du
\end{aligned}$$

Using Theorem 2.7 we know that for each $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|u| < \delta \Longrightarrow \|f - f_u\|_1 < \varepsilon$$

Also, using the properties of an approximate identity we have

$$K = \sup_{\lambda > 0} \|k_\lambda\|_1 < \infty$$

and there exists some $\lambda_0 > 0$ such that

$$\lambda < \lambda_0 \Longrightarrow \int_{|u| \geq \delta} |k_\lambda(u)| \, du < \varepsilon$$

So now assume that $\lambda < \lambda_0$ and continue the calculation we started above:

$$\begin{aligned}
\|f - f * k_\lambda\|_1 &= \int_{\mathbb{R}} |k_\lambda(u)| \|f - f_u\|_1 \, du \\
&= \int_{|u|<\delta} |k_\lambda(u)| \|f - f_u\|_1 \, du + \int_{|u|\geq\delta} |k_\lambda(u)| \|f - f_u\|_1 \, du \\
&\leq \int_{|u|<\delta} |k_\lambda(u)| \varepsilon \, du + \int_{|u|\geq\delta} |k_\lambda(u)| (\|f\|_1 + \|f_u\|_1) \, du \\
&\leq \varepsilon \int_{\mathbb{R}} |k_\lambda(u)| \, du + 2\|f\|_1 \int_{|u|\geq\delta} |k_\lambda(u)| \, du \\
&\leq \varepsilon K + 2\|f\|_1 \varepsilon
\end{aligned}$$

Taking $\varepsilon \to 0$ and $\lambda_0 \to 0$ completes the proof. □

Exercise 4. Prove that the assumptions of Theorem 2.2 (Fourier inversion) imply that f must be continuous and bounded.

Remark 2.8. Exercise 4 shows that our Fourier inversion theorem only holds for continuous functions. However, many signals that we encounter will have discontinuities. Thus we will need to extend the theory to discontinuous functions. This will be done by extending the Fourier transform to $\mathbf{L}^2(\mathbb{R})$ (more on this later).

Recall in Section 2.1 that for a linear shift invariant operator L with impulse response h, we wanted to write Lf in terms of the eigenvalues $\widehat{h}(\omega)$ of L by also being able to compute $\widehat{f}(\omega)$. The previous theorem gives us part of the solution; the other part is given by the *convolution theorem*, which is stated next.

Theorem 2.9 (Convolution thoerem)**.** *Let $f, g \in \mathbf{L}^1(\mathbb{R})$. Then the function $h = f * g \in \mathbf{L}^1(\mathbb{R})$ and*

$$\widehat{h}(\omega) = \widehat{g}(\omega)\widehat{f}(\omega)$$

Proof. See p. 37 of *A Wavelet Tour of Signal Processing.* □

Recall now that every bounded, linear shift invariant operator L can be written as $Lf = h * f$, where $h = L\delta$. Thus using the Fourier inversion theorem and the convolution theorem we have:

$$Lf(t) = h * f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{h * f}(\omega) e^{i\omega t} \, d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{h}(\omega)\widehat{f}(\omega) e^{i\omega t} \, d\omega$$

Thus at last we see that the sinusoids $e_\omega(t) = e^{i\omega t}$ diagonalize L, with eigenvalues $\widehat{h}(\omega)/2\pi$. To see this recall from before we wrote for a real valued symmetric matrix $\mathbf{A}$ with eigenvectors $\{v_k\}_k$ and eigenvalues $\{\lambda_k\}_k$,

$$\mathbf{A}x = \sum_{k=1}^{n} \langle x, v_k \rangle \mathbf{A} v_k = \sum_{k=1}^{n} \lambda_k \langle x, v_k \rangle v_k \, .$$

Property	Function	Fourier Transform
	$f(t)$	$\hat{f}(\omega)$
Inverse	$\hat{f}(t)$	$2\pi f(-\omega)$
Convolution	$f_1 \star f_2(t)$	$\hat{f}_1(\omega)\hat{f}_2(\omega)$
Multiplication	$f_1(t)\,f_2(t)$	$\frac{1}{2\pi}\hat{f}_1 \star \hat{f}_2(\omega)$
Translation	$f(t-u)$	$e^{-iu\omega}\hat{f}(\omega)$
Modulation	$e^{i\xi t}f(t)$	$\hat{f}(\omega-\xi)$
Scaling	$f(t/s)$	$\lvert s\rvert\hat{f}(s\,\omega)$
Time derivatives	$f^{(p)}(t)$	$(i\omega)^p\hat{f}(\omega)$
Frequency derivatives	$(-it)^p f(t)$	$\hat{f}^{(p)}(\omega)$
Complex conjugate	$f^*(t)$	$\hat{f}^*(-\omega)$
Hermitian symmetry	$f(t)\in\mathbb{R}$	$\hat{f}(-\omega)=\hat{f}^*(\omega)$

Figure 1: Summary of basic properties of the Fourier transform. Taken from Table 2.1 of *A Wavelet Tour of Signal Processing*.

The correspondence is $\lambda_k \leftrightarrow \widehat{h}(\omega)/2\pi$, $\langle x, v_k\rangle \leftrightarrow \widehat{f}(\omega)$, and $v_k \leftrightarrow e^{i\omega t}$.

The Fourier transform has several important properties that are listed in Figure 1.

Exercise 5. Verify all of the properties in Figure 1. No need to turn this one in, but it is important to do these verifications.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 03: The Fourier transform on $\mathbf{L}^2(\mathbb{R})$

January 16, 2020

Lecturer: Matthew Hirn

Remark 2.10. The Dirac $\delta(t)$ is not a function, and hence is not in $\mathbf{L}^1(\mathbb{R})$; it is a distribution, which we will not discuss in this course. The Dirac distribution has the property of being an identity under convolution, meaning that $f * \delta(u) = f(u)$ if $f \in \mathbf{L}^1(\mathbb{R})$ and if f is continuous. There is no $\mathbf{L}^1(\mathbb{R})$ function with this property, so the question is how to define the Dirac. The notion of an approximate identity $\{k_\lambda\}_{\lambda>0}$, defined above, can be used to define it. Indeed, we define $\delta(t) = \lim_{\lambda \to 0} k_\lambda(t)$, where we understand that the limit means *weak convergence*. By weak convergence, we mean that for any continuous function ϕ,

$$\lim_{\lambda \to 0} \int_{\mathbb{R}} \phi(t) k_\lambda(t)\, dt = \phi(0) =: \int_{\mathbb{R}} \phi(t)\delta(t)\, dt$$

We can also define a translated Dirac $\delta_\tau(t) = \delta(t - \tau)$, which is defined as the weak limit of a translated approximate identity. This means that

$$\phi * \delta(u) = \int_{\mathbb{R}} \phi(t)\delta(u - t)\, dt = \int_{\mathbb{R}} \phi(t)\delta(t - u)\, dt = \phi(u) \tag{5}$$

Note that these properties, in particular (5), follow from defining $\delta(t) = \lim_{\lambda \to 0} k_\lambda(t)$ in the weak sense, Theorem 2.6, and the fact that ϕ is continous.

Using this formalism, we define the Fourier transform of $\delta(t)$, $\widehat{\delta}(\omega)$, as:

$$\widehat{\delta}(\omega) = \int_{\mathbb{R}} \delta(t) e^{-i\omega t}\, dt = 1$$

A translated Dirac $\delta_\tau(t) = \delta(t - \tau)$ has Fourier transform calculated by evaluating $e^{-i\omega t}$ at $t = \tau$,

$$\widehat{\delta}_\tau(\omega) = \int_{\mathbb{R}} \delta(t - \tau) e^{-i\omega t}\, dt = e^{-i\omega\tau}$$

The *Dirac comb* is a sum of translated Diracs:

$$c(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT) \tag{6}$$

It is used to obtain a discrete sampling of an analogue signal, as we shall see later. Its Fourier transform is:

$$\widehat{c}(\omega) = \sum_{n=-\infty}^{+\infty} e^{-inT\omega}$$

Remarkably, $\widehat{c}(\omega)$ is also a Dirac comb, as the next theorem shows.

Theorem 2.11 (Poisson Formula)**.** *In the sense of distribution equalities,*

$$\sum_{n=-\infty}^{+\infty} e^{-inT\omega} = \frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta\left(\omega - \frac{2\pi k}{T}\right)$$

In other words, for every $\widehat{\phi} \in \mathbf{C}_0^\infty(\mathbb{R})$*, that is for every compactly supported infinitely differentiable function, one has*

$$\int_{\mathbb{R}} \widehat{\phi}(\omega) \left[\sum_{n=-\infty}^{+\infty} e^{-inT\omega}\right] d\omega = \int_{\mathbb{R}} \widehat{\phi}(\omega) \left[\frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta\left(\omega - \frac{2\pi k}{T}\right)\right] d\omega$$

Proof. See p. 41–42 of *A Wavelet Tour of Signal Processing.* □

Consider now the function

$$f(t) = \mathbf{1}_{[-1,1]}(t) = \begin{cases} 1 & -1 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We can compute the Fourier transform of this function:

$$\widehat{f}(\omega) = \int_{-1}^{1} e^{-i\omega t}\, dt = \frac{2\sin(\omega)}{\omega}$$

One can verify that this function is not integrable; we would expect this from Exercise 4 because $f(t)$ is not continuous. However, $\widehat{f}(\omega)$ is square integrable; that is $\widehat{f} \in \mathbf{L}^2(\mathbb{R})$. This motivates extending the Fourier transform to functions $f \in \mathbf{L}^2(\mathbb{R})$. Recall that $\mathbf{L}^2(\mathbb{R})$ is a Hilbert space with inner product

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t) g^*(t)\, dt.$$

We first have the following fundamental results:

Theorem 2.12 (Parseval)**.** *Let* $f, g \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$*. Then:*

$$\langle f, g \rangle = \frac{1}{2\pi} \langle \widehat{f}, \widehat{g} \rangle$$

Proof. See p. 39 of *A Wavelet Tour of Signal Processing.* □

Corollary 2.13 (Plancheral)**.** *Let* $f \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$*. Then:*

$$\|f\|_2 = \frac{1}{\sqrt{2\pi}} \|\widehat{f}\|_2$$

Note that in the previous theorems, the inner product and norm are computable because we assume $f, g \in \mathbf{L}^2(\mathbb{R})$, but the Fourier transform is only well defined because we assume $f, g \in \mathbf{L}^1(\mathbb{R})$ as well. We would like to remedy this be extending the Fourier transform to all functions $f \in \mathbf{L}^2(\mathbb{R})$, even those for which $f \notin \mathbf{L}^1(\mathbb{R})$. We do this with a *density argument*, which will define the Fourier transform of a function $f \in \mathbf{L}^2(\mathbb{R})$ as the limit of Fourier transforms of functions in $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$. A very useful inequality from real analysis, which we will need here, is *Hölder's inequality*:

$$\forall\, f \in \mathbf{L}^p(\mathbb{R}),\ g \in \mathbf{L}^q(\mathbb{R}),\ p, q \in [1, \infty],\ \frac{1}{p} + \frac{1}{q} = 1, \quad \|fg\|_1 \leq \|f\|_p \|g\|_q$$

Now to the density argument. The first thing to note is that $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ is dense in $\mathbf{L}^2(\mathbb{R})$. This means that given an $f \in \mathbf{L}^2(\mathbb{R})$, we can find a family $\{f_n\}_{n \geq 1}$ of functions in $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ that converges to f,

$$\lim_{n \to \infty} \|f - f_n\|_2 = 0$$

In fact it is easy to find a such a family. Define:

$$f_n(t) = f(t)\mathbf{1}_{[-n,n]}(t)$$

We have that $f_n \in \mathbf{L}^2(\mathbb{R})$ for all $n \geq 1$ since $|f_n(t)| \leq |f(t)|$ for all $t \in \mathbb{R}$. Furthermore, $f_n \in \mathbf{L}^1(\mathbb{R})$ since by Hölder's inequality we have:

$$\begin{aligned}
\|f_n\|_1 = \int_{\mathbb{R}} |f_n(t)|\, dt &= \int_{\mathbb{R}} |f(t)\mathbf{1}_{[-n,n]}(t)|\, dt \\
&\leq \left(\int_{\mathbb{R}} |f(t)|^2\, dt \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} |\mathbf{1}_{[-n,n]}(t)|^2 \right)^{\frac{1}{2}} \\
&= \|f\|_2 \left(\int_{-n}^{n} 1\, dt \right)^{\frac{1}{2}} \\
&= \sqrt{2n} \|f\|_2
\end{aligned}$$

We also have that

$$\|f - f_n\|_2 = \left(\int_{|t| > n} |f(t)|^2 \right)^{\frac{1}{2}} \to 0 \text{ as } n \to \infty$$

Now, since $f_n \to f$, the family $\{f_n\}_{n \geq 1}$ is also a *Cauchy sequence*, meaning that for all $\varepsilon > 0$ there exists an N such that if $n, m > N$, then $\|f_n - f_m\|_2 \leq \varepsilon$. Furthermore, since $f_n \in \mathbf{L}^1(\mathbb{R})$, its Fourier transform $\widehat{f_n}$ is well defined. The Plancheral formula (Corollary 2.13) then yields:

$$\|\widehat{f_n} - \widehat{f_m}\|_2 = \sqrt{2\pi} \|f_n - f_m\|_2$$

Thus since $\{f_n\}_{n \geq 1}$ is a Cauchy sequence, we see that $\{\widehat{f_n}\}_{n \geq 1}$ is a Cauchy sequence as well. Since $\mathbf{L}^2(\mathbb{R})$ is a Hilbert space, it is complete, which means that every Cauchy sequence converges to an element of $\mathbf{L}^2(\mathbb{R})$. Thus there exists an $F \in \mathbf{L}^2(\mathbb{R})$ such that

$$\lim_{n \to \infty} \|F - \widehat{f_n}\|_2 = 0$$

We define the Fourier transform of $f \in \mathbf{L}^2(\mathbb{R})$ as F, and from now on write $\widehat{f} = F$. Note that when $f \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ this definition of the Fourier transform (the $\mathbf{L}^2$ definition) coincides with the definition given in (1) (the $\mathbf{L}^1$ definition).

One can show the extension of the Fourier transform to $\mathbf{L}^2(\mathbb{R})$ satisfies the convolution theorem (Theorem 2.9), the Parseval formula (Theorem 2.12), the Plancheral formula (Corollary 2.13), and all properties in Figure 1. In particular, the Plancheral formula implies the following. Let $\mathcal{F}(f) = \widehat{f}$, so that $\mathcal{F}$ is the operator that maps a function f to its Fourier transform $\widehat{f}$. We see from the Plancheral formula and the extension of the Fourier transform to $\mathbf{L}^2(\mathbb{R})$ that $\mathcal{F} : \mathbf{L}^2(\mathbb{R}) \to \mathbf{L}^2(\mathbb{R})$, and furthermore that this linear operator is an isometry up to a factory $1/\sqrt{2\pi}$. The operator $\mathcal{F} : \mathbf{L}^2(\mathbb{R}) \to \mathbf{L}^2(\mathbb{R})$ is bijective, and thus is invertible; we therefore have Fourier inversion for $\mathbf{L}^2(\mathbb{R})$ functions as well.

Remark 2.14. To summarize the Fourier transform can be defined on $\mathbf{L}^1(\mathbb{R})$ in which case we have

$$\mathcal{F} : \mathbf{L}^1(\mathbb{R}) \to \mathbf{L}^\infty(\mathbb{R})$$

with $\|\widehat{f}\|_\infty \leq \|f\|_1$, or on $\mathbf{L}^2(\mathbb{R})$ where we have:

$$\mathcal{F} : \mathbf{L}^2(\mathbb{R}) \to \mathbf{L}^2(\mathbb{R})$$

with $\|f\|_2 = (1/\sqrt{2\pi})\|\widehat{f}\|_2$. It follows then from the Riesz-Thorin Theorem that the Fourier transform can be extended to $\mathbf{L}^p(\mathbb{R})$ for any $1 \leq p \leq 2$, where we have

$$\mathcal{F} : \mathbf{L}^p(\mathbb{R}) \to \mathbf{L}^q(\mathbb{R}), \quad \frac{1}{p} + \frac{1}{q} = 1, \;\; 1 \leq p \leq 2$$

and that

$$\|\widehat{f}\|_q \leq \left(\frac{1}{2\pi}\right)^{\frac{1}{p}} \|f\|_p \tag{7}$$

Equation (7) is called the *Hausdorff–Young Inequality*. Note that in general one only obtains equality for $p = q = 2$, and indeed $\mathcal{F}$ is not an isometry otherwise (up to the constant factor) and is not invertible. Indeed, we saw this for $\mathbf{L}^1(\mathbb{R})$, where in order to get Fourier inversion we had to assume that $\widehat{f} \in \mathbf{L}^1(\mathbb{R})$ as well.

Exercise 6. Read Section 2.2 of *A Wavelet Tour of Signal Processing.*

2.3 Regularity and Decay

Section 2.3.1 of A Wavelet Tour of Signal Processing [1].

The global regularity of f depends on the decay of $|\widehat{f}(\omega)|$ as $\omega \to \infty$. In particular, the smoother the function, the faster the decay of $|\widehat{f}(\omega)|$. The intuition is that smooth functions vary slowly, and thus can be well represented by low frequency modes $e^{i\omega t}$, i.e., those with small values of $|\omega|$. On the other hand, if f is irregular, then it must have sharp transitions

which require fast oscillations to capture. We make these intuitions precise with the following two results. First define $\mathbf{C}^n(\mathbb{R})$ as the space of functions with n continuous derivatives; $\mathbf{C}^0(\mathbb{R})$ is the space of continuous functions.

Theorem 2.15. *Let $1 \leq p \leq 2$ and $f \in \mathbf{L}^p(\mathbb{R})$. If there exits a constant C and $\epsilon > 0$ such that*

$$|\widehat{f}(\omega)| \leq \frac{C}{1 + |\omega|^{n+1+\epsilon}}$$

for some $n \in \mathbb{N}$, then $f \in \mathbf{C}^n(\mathbb{R}) \cap \mathbf{L}^\infty(\mathbb{R})$.

Proof. We know from Exercise 4 that if $\widehat{f} \in \mathbf{L}^1(\mathbb{R})$, then f is continuous and bounded. Notice for $n = 0$ we have:

$$\|\widehat{f}\|_1 = \int_{\mathbb{R}} |\widehat{f}(\omega)|\, d\omega \leq \int_{\mathbb{R}} \frac{C}{1 + |\omega|^{1+\epsilon}}\, d\omega < \infty$$

So indeed $f \in \mathbf{C}(\mathbb{R}) \cap \mathbf{L}^\infty(\mathbb{R})$. Now consider $n \in \mathbb{N}$ and $k \leq n$; define the function $F_k(\omega) = (i\omega)^k \widehat{f}(\omega)$. We see that:

$$\|F_k\|_1 \leq \int_{\mathbb{R}} \frac{C|\omega|^k}{1 + |\omega|^{n+1+\epsilon}}\, d\omega < \infty$$

It thus follows that $\mathcal{F}^{-1}(F_k)$ (i.e., the inverse Fourier transform of F_k) is continuous and bounded. But from Figure 1 we know that $\mathcal{F}^{-1}(F_k) = f^{(k)}(t)$, and so the proof is completed. □

Note in particular that if $\widehat{f}$ has compact support, then $f \in \mathbf{C}^\infty(\mathbb{R})$. In the other direction we have:

Theorem 2.16. *Let $f \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{C}^n(\mathbb{R})$ with $f^{(n)} \in \mathbf{L}^1(\mathbb{R})$. Then:*

$$|\widehat{f}(\omega)| \leq \frac{C}{|\omega|^n}$$

for some constant C[1].

Exercise 7. Prove Theorem 2.16.

Remark 2.17. Notice there is a gap between the two theorems relating regularity and decay. This cannot be avoided. Furthermore, we notice that the decay of $|\widehat{f}(\omega)|$ depends upon the *worst* singular behavior of f. Indeed as the function $f(t) = \mathbf{1}_{[-1,1]}(t)$ illustrates, the function is discontinuous and thus its Fourier decay is limited by Theorem 2.15. However, f has only two singular points. It is often much more desirable to characterize the local regularity of a function. However, the Fourier transform cannot do this since the sinusoids $e^{i\omega t}$ are global functions on $\mathbb{R}$. In order to remedy both of these points, we will need to introduce localized waveforms. We will see later that wavelets do the job.

[1]Thanks to Theodore Faust for pointing out a mistake in an earlier version of the statement of this theorem.

Exercise 8. Show that the Fourier transform of

$$f(t) = e^{-(a-ib)t^2}, \quad a > 0$$

is

$$\widehat{f}(\omega) = \sqrt{\frac{\pi}{a-ib}} \exp\left(-\frac{a+ib}{4(a^2+b^2)}\omega^2\right)$$

Exercise 9 (Riemann-Lebesgue Lemma)**.** Prove that if $f \in \mathbf{L}^1(\mathbb{R})$, then $\lim_{|\omega|\to\infty} \widehat{f}(\omega) = 0$. Hint: Start with $f \in \mathbf{C}^1(\mathbb{R})$ that have compact support, and use a density argument. This approach uses the standard fact from real analysis that compactly supported $\mathbf{C}^\infty(\mathbb{R})$ functions are dense in $\mathbf{L}^1(\mathbb{R})$. However, if you have not seen this before, it is unsatisfying to use it here to prove the exercise. In this case, consider instead the Gaussian function:

$$g(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

Define dilations of g as:

$$g_\sigma(u) = \sigma^{-1} g(\sigma^{-1}u), \quad \sigma > 0$$

Prove that $\{g_\sigma\}_{\sigma>0}$ forms an approximate identity. The functions $\{f * g_\sigma\}_{\sigma>0}$ are not compactly supported, but they can be used to prove the result. Figure out how and provide the proof.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 04: Uncertainty Principle and Introduction to DSP

January 21, 2020

Lecturer: Matthew Hirn

2.4 Uncertainty Principle

Section 2.3.2 of A Wavelet Tour of Signal Processing [1].

The previous section motivates the following question. Can we construct a function f that is well localized in both time and frequency, and if so, how well localized can it be simultaneously in both domains? We know that a Dirac $\delta(t)$ is well localized in space, but $\widehat{\delta}(\omega) = 1$ for all ω, and similarly $e_\xi(t) = e^{i\xi t}$ is not well localized in space, but $\widehat{e}_\xi(\omega) = \delta(\omega - \xi)$. From the previous section, we know that $|\widehat{f}(\omega)|$ decays quickly as $\omega \to \infty$ only if f is very regular. But if f is very regular, it cannot have sharp transitions and thus cannot decay too fast in space as $t \to \infty$.

Similarly, to adjust the spread of a function f while keeping its total energy constant, we can dilate by a factor $s > 0$ with suitable normalization:

$$f_s(t) = s^{-1/2} f(s^{-1}t)$$

If $s < 1$, then the spread of f is decreased, while if $s > 1$ the spread of f is increased. Regardless, the normalization $s^{-1/2}$ insures that $\|f_s\|_2 = \|f\|_2$. The Fourier transform of f_s is:

$$\widehat{f}_s(\omega) = \sqrt{s}\widehat{f}(s\omega)$$

We see that the dilation has the opposite effect on $\widehat{f}$. In particular, if $s < 1$, then the spread of $\widehat{f}$ is increased, while if $s > 1$, the spread of $\widehat{f}$ is decreased. We thus begin to see there is a trade-off between time and frequency localization.

Time and frequency localizations are limited by the (Heisenberg) uncertainty principle, which you may have seen in quantum mechanics as the uncertainty on the position and momentum of a free particle. We will use the framework of quantum mechanics to motivate the following discussion, although it will hold for general functions $f \in \mathbf{L}^2(\mathbb{R})$. The state of a one-dimensional particle is described by a wave function $f \in \mathbf{L}^2(\mathbb{R})$. The probability density function for the location of this particle to be at t is

$$\frac{1}{\|f\|^2}|f(t)|^2$$

while the probability density function for its momentum to be ω is

$$\frac{1}{2\pi\|f\|^2}|\widehat{f}(\omega)|^2$$

It follows that the average location of the particle is given by

$$u = \frac{1}{\|f\|^2} \int_{\mathbb{R}} t|f(t)|^2 \, dt$$

while its average momentum is:

$$\xi = \frac{1}{2\pi\|f\|^2} \int_{\mathbb{R}} \omega|\widehat{f}(\omega)|^2 \, d\omega$$

The variance around the average location u is

$$\sigma_t^2 = \frac{1}{\|f\|^2} \int_{\mathbb{R}} (t-u)^2|f(t)|^2 \, dt$$

and the variance around the average momentum is:

$$\sigma_\omega^2 = \frac{1}{2\pi\|f\|^2} \int_{\mathbb{R}} (\omega-\xi)^2|\widehat{f}(\omega)|^2 \, d\omega$$

The variances measure our uncertainty as to the location and momentum of the particle. In particular, the larger the variance, the less certain we are. As one may know from quantum mechanics, we cannot know the position and momentum of a particle simultaneously. The following theorem makes this statement precise

Theorem 2.18 (Uncertainty Principle)**.** *The temporal variance and the frequency variance of a function $f \in \mathbf{L}^2(\mathbb{R})$ must satisfy*

$$\sigma_t^2 \sigma_\omega^2 \geq \frac{1}{4}$$

We obtain equality if and only if there exists $(u, \xi, a, b) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that

$$f(t) = ae^{i\xi t - b(t-u)^2}, \quad \text{Real}(b) > 0 \tag{8}$$

Functions (8) are called Gabor functions.

Proof. The proof is relatively simple for functions $f \in \mathcal{S}(\mathbb{R})$, which are Schwartz class functions. The *Schwartz class* is an important class of functions to know, so we define it now. The space $\mathcal{S}(\mathbb{R})$ consists of all infinitely differentiable functions $f : \mathbb{R} \to \mathbb{C}$ such that $f^{(n)}(t)$ is rapidly decreasing for all $n \geq 0$, that is

$$\sup_{t\in\mathbb{R}} |t|^m |f^{(n)}(t)| < \infty, \quad \forall\, m, n \geq 0$$

An example of a Schwartz class function is the family of functions defined in (8). The Fourier transform, as defined for $\mathbf{L}^1(\mathbb{R})$ functions in (1), is also well defined for $f \in \mathcal{S}(\mathbb{R})$, and furthermore $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$.

Now to the proof. First, note that if the time and frequency averages of f are u and ξ respectively, then the time and frequency averages of $e^{-i\xi t}f(t+u)$ are zero. Thus it is sufficient to prove the theorem for $u = \xi = 0$. First note that if we write $f(t) = f_1(t) + if_2(t)$, then $f'(t) = f_1'(t) + if_2'(t)$,

$$|f(t)|^2 = f_1(t)^2 + f_2(t)^2$$

and

$$\frac{d}{dt}|f(t)|^2 = 2f_1(t)f_1'(t) + 2f_2(t)f_2'(t) = f^*(t)f'(t) + f(t)f'^*(t)$$

We then have using integration by parts:

$$\begin{aligned}
\|f\|^2 &= \int_{\mathbb{R}} |f(t)|^2\,dt \\
&= \underbrace{t|f(t)|^2\Big|_{-\infty}^{+\infty}}_{0 \text{ b/c } f\in\mathcal{S}(\mathbb{R})} - \int_{\mathbb{R}} t\frac{d}{dt}|f(t)|^2\,dt \\
&= -\int_{\mathbb{R}} t[f^*(t)f'(t) + f(t)f'^*(t)]\,dt
\end{aligned}$$

Taking the absolute value of both sides yields and using Hölder's inequality (Cauchy-Schwarz) we have:

$$\begin{aligned}
\|f\|^2 &= \left|\int_{\mathbb{R}} t[f^*(t)f'(t) + f(t)f'^*(t)]\,dt\right| \\
&\le 2\int_{\mathbb{R}} |t||f(t)||f'(t)|\,dt \\
&\le 2\left(\int_{\mathbb{R}} t^2|f(t)|^2\,dt\right)^{\frac{1}{2}}\left(\int_{\mathbb{R}} |f'(t)|^2\,dt\right)^{\frac{1}{2}} \\
&= 2\|f\|\sigma_t\left(\int_{\mathbb{R}} |f'(t)|^2\,dt\right)^{\frac{1}{2}}
\end{aligned}$$

Now use the Plancheral formula (Corollary 2.13) and the identity $\mathcal{F}(f')(\omega) = i\omega\widehat{f}(\omega)$ to obtain

$$\left(\int_{\mathbb{R}} |f'(t)|^2\,dt\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2\pi}}\left(\int_{\mathbb{R}} \omega^2|\widehat{f}(\omega)|^2\right)^{\frac{1}{2}} = \|f\|\sigma_\omega$$

Thus we obtain:

$$\|f\|^2 \le 2\|f\|\sigma_t\|f\|\sigma_\omega$$

from which the desired inequality follows.

For the second part, if $u = \xi = 0$, one can verify that equality holds for $f(t) = ae^{-bt^2}$. Now suppose equality holds. Then we must have equality when we applied the Cauchy-Schwarz inequality. But this can only happen if the two functions are equal, up to a constant, which in this case means that

$$f'(t) = \beta t f(t)$$

The solutions to this differential equation are $f(t) = ae^{\beta t^2/2}$. Setting $-b = \beta/2$ we obtain (8).

The proof can be extended to any $\mathbf{L}^2(\mathbb{R})$ function; see for example [3]. □

The uncertainty principle does not preclude a function having compact support in both time an frequency. However, this is also impossible.

Theorem 2.19. *Let $f \in \mathbf{L}^1(R) \cup \mathbf{L}^2(\mathbb{R})$. If $f \neq 0$ has a compact support, then $\widehat{f}(\omega)$ cannot be zero on a whole interval. Similarly, if $\widehat{f} \neq 0$ has compact support, then $f(t)$ cannot be zero on a whole interval.*

Proof. We prove the second statement. Suppose that $\widehat{f}$ has compact support, which is included in the interval $[-b, b]$. Then using the Fourier inversion formula, we have

$$f(t) = \frac{1}{2\pi} \int_{-b}^{b} \widehat{f}(\omega) e^{i\omega t} \, d\omega \tag{9}$$

Suppose by contradiction that $f(t) = 0$ for all $t \in [c, d]$. Set $t_0 = (c + d)/2$ and calculate the n^{th} derivative of f at t_0 as:

$$0 = f^{(n)}(t_0) = \frac{1}{2\pi} \int_{-b}^{b} \widehat{f}(\omega) \frac{d}{dt} e^{i\omega t} \Big|_{t=t_0} d\omega = \frac{1}{2\pi} \int_{-b}^{b} \widehat{f}(\omega) (i\omega)^n e^{i\omega t_0} \, d\omega$$

Now expand $e^{i\omega t}$ as an infinite Taylor series around t_0:

$$\forall\, t \in \mathbb{R}, \quad e^{i\omega t} = \sum_{n=0}^{\infty} \frac{(i\omega)^n}{n!} e^{i\omega t_0} (t - t_0)^n$$

Now go back to (9) and plug in the Taylor series for $e^{i\omega t}$,

$$f(t) = \sum_{n=0}^{\infty} \frac{(t - t_0)^n}{n!} \underbrace{\frac{1}{2\pi} \int_{-b}^{b} \widehat{f}(\omega) (i\omega)^n e^{i\omega t_0} \, d\omega}_{0} = 0$$

But now we have $f(t) = 0$ for all $t \in \mathbb{R}$, which implies that $\widehat{f}(\omega) = 0$ for all $\omega \in \mathbb{R}$; but this is a contradiction. □

Exercise 10. Read Section 2.3 of *A Wavelet Tour of Signal Processing.*

Exercise 11. Read Section 2.4 of *A Wavelet Tour of Signal Processing.*

Exercise 12. For any $A > 0$, construct a function f such that $\sigma_t(f) > A$ and $\sigma_\omega(f) > A$.

3 Discrete Revolution

Chapter 3 of A Wavelet Tour of Signal Processing [1].

3.1 Sampling Analog Signals

Section 3.1 of A Wavelet Tour of Signal Processing [1].

Signals $f : \mathbb{R} \to \mathbb{C}$ must be discretized to be stored on a computer. In practice we can only keep a finite amount of information, which means that we can only keep a finite number of samples from f. We will return to this setting in a bit. For now we consider a discrete, countably infinite number of samples from f, given by:

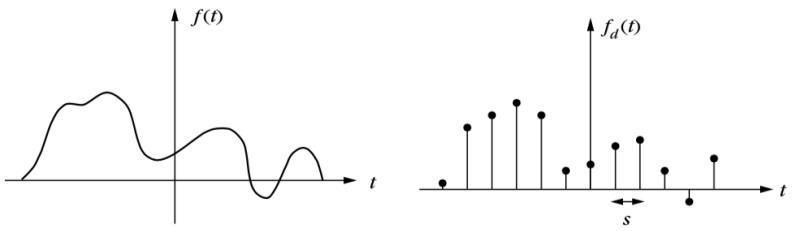
$$\text{Samples } = \{f(ns)\}_{n\in\mathbb{Z}}, \quad s^{-1} = \text{sampling rate} \tag{10}$$

In particular $s = 1$ means we sample every integer, $s = 2$ means we sample every other integer, while $s = 1/2$ means we sample every half integer, and so on.

Assume that f is continuous, so that (10) is well defined. We want to know when we can recover $f(t)$ for all $t \in \mathbb{R}$ from the samples $\{f(ns)\}_{n\in\mathbb{Z}}$. We represent these discrete samples as a sum of weighted Diracs:

$$f_d(t) = \sum_{n\in\mathbb{Z}} f(ns)\delta(t - ns)$$

The signal $f_d : \mathbb{R} \to \mathbb{C}$ is defined for all $t \in \mathbb{R}$ but only takes nonzero values at $t = ns$ for $n \in \mathbb{Z}$. It is thus a discrete sampling of f; see Figure 2.



(a) Function $f(t)$ (b) Discrete function samples $f_d(t)$

Figure 2: A continuous function and its discrete sampled version. Taken from Figure 3.1 of *A Wavelet Tour of Signal Processing*.

The Fourier transform of $f_d(t)$ is:

$$\widehat{f_d}(\omega) = \sum_{n\in\mathbb{Z}} f(ns)e^{-ins\omega}$$

Notice this is a Fourier series; we'll come back to this point later. We first compute $\widehat{f_d}(\omega)$ a second way, which will illuminate the relationship between $\widehat{f}(\omega)$ and $\widehat{f_d}(\omega)$.

Theorem 3.1. *The Fourier transform of $f_d(t)$ is:*

$$\widehat{f_d}(\omega) = \frac{1}{s}\sum_{k\in\mathbb{Z}} \widehat{f}\left(\omega - \frac{2k\pi}{s}\right)$$

Proof. Define the Dirac comb (see also (6)) as:

$$c(t) = \sum_{n \in \mathbb{Z}} \delta(t - ns)$$

We can rewrite $f_d(t)$ as the multiplication of $f(t)$ with $c(t)$:

$$f_d(t) = f(t)c(t)$$

Using the convolution theorem (Theorem 2.9), we have:

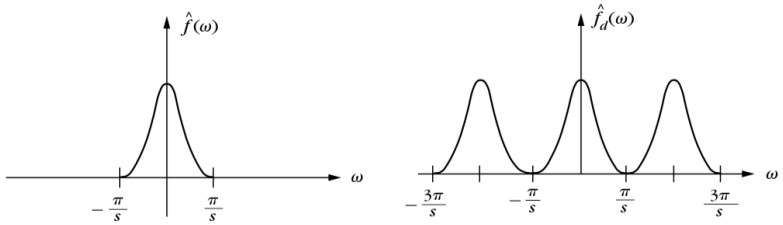
$$\widehat{f}_d(\omega) = \frac{1}{2\pi} \widehat{f} * \widehat{c}(\omega)$$

But the Poisson Formula (Theorem 2.11) proves:

$$\widehat{c}(\omega) = \frac{2\pi}{s} \sum_{k \in \mathbb{Z}} \delta\left(\omega - \frac{2\pi k}{s}\right)$$

The theorem then follows immediately. □

Theorem 3.1 proves that the Fourier transform $\widehat{f}_d(\omega)$ is obtained by making the Fourier transform $\widehat{f}(\omega)$ $2\pi/s$ periodic. Thus sampling f "periodizes" its frequency response. Figure 3 illustrates the point. The main point here is that if $\operatorname{supp} \widehat{f} \subseteq [-\pi/s, \pi/s]$, then $f(t)$ can be recovered from $f_d(t)$; if $\widehat{f}$ is supported outside of $[-\pi/s, \pi/s]$ then aliasing may occur, in which case we cannot recover $f(t)$ from $f_d(t)$. The next theorem makes precise the first point.



(a) Fourier transform $\widehat{f}(\omega)$ of the signal from Figure 2

(b) Fourier transform $\widehat{f}_d(\omega)$ of the sampled signal from Figure 2

Figure 3: The Fourier transforms of $\widehat{f}(\omega)$ and $\widehat{f}_d(\omega)$. Taken from Figure 3.1 of *A Wavelet Tour of Signal Processing*.

Theorem 3.2 (*Whittaker–Nyquist–Kotelnikov–Shannon Sampling Theorem*)**.** *If* $\operatorname{supp} \widehat{f} \subseteq [-\pi/s, \pi/s]$, *then*

$$f(t) = f_d * \phi_s(t) = \sum_{n \in \mathbb{Z}} f(ns)\phi_s(t - ns)$$

where

$$\phi_s(t) = \frac{\sin(\pi t/s)}{\pi t/s}$$

Proof. If $n \neq 0$, then the support of $\widehat{f}(\omega - 2n\pi/s)$ does not intersect with $\widehat{f}(\omega)$ since $\widehat{f}(\omega) = 0$ for $|\omega| > \pi/s$. Thus by Theorem 3.1 (see also Figure 3)

$$\widehat{f}_d(\omega) = \frac{\widehat{f}(\omega)}{s}, \quad |\omega| \leq \frac{\pi}{s}$$

The Fourier transform of $\phi_s(t)$ is

$$\widehat{\phi}_s(\omega) = s\mathbf{1}_{[-\pi/s,\pi/s]}(\omega)$$

Therefore

$$\widehat{f}(\omega) = \widehat{\phi}_s(\omega)\widehat{f}_d(\omega)$$

Now apply the inverse Fourier transform both sides:

$$f(t) = \phi_s * f_d(t) = \phi_s * \sum_{n\in\mathbb{Z}} f(ns)\delta(t - ns) = \sum_{n\in\mathbb{Z}} f(ns)\phi_s(t - ns)$$

□

If the support of $\widehat{f}(\omega)$ is not included in $[-\pi/s, \pi/s]$ then *aliasing* can occur, which is what happens when the supports of $\widehat{f}(\omega - 2k\pi/s)$ overlap for several k. In this case $\widehat{f}(\omega) \neq \widehat{\phi}_s(\omega)\widehat{f}_d(\omega)$, and the sampling theorem (Theorem 3.2) does not apply and we cannot recover $f(t)$ from $f_d(t)$. Indeed, the Fourier transform of $f_d * \phi_s(t)$ may be very different than the Fourier transform of $f(t)$, in which case $f_d * \phi_s(t)$ will look very different than $f(t)$. See Figure 4 for an illustration.

A *bandlimited* signal is a function f such that $\operatorname{supp} \widehat{f} \subseteq [-R, R]$ for some $R > 0$. The sampling theorem (Theorem 3.2) proves that such signals can be sampled with a discrete set of samples for an appropriate sampling rate $s = \pi/R$. However, by Theorem 2.15, such signals must necessarily be $\mathbf{C}^\infty$. We will want to be able to process other signals as well. We can do so by first filtering f with some filter h (or a family of filters), which computes $f * h(t)$. If $\operatorname{supp} \widehat{h} \subseteq [-R, R]$ then $f * h(t)$ is bandlimited as well, with the same frequency range. We can thus sample $f * h(t)$ according to Theorem 3.2. In general we are going to need more than one filter, and each filter will need to be localized in some part of the frequency axis. This will lead us to Gabor filters (windowed Fourier) and wavelets, amongst other filter families.

Exercise 13. Read Section 3.1 of *A Wavelet Tour of Signal Processing.*

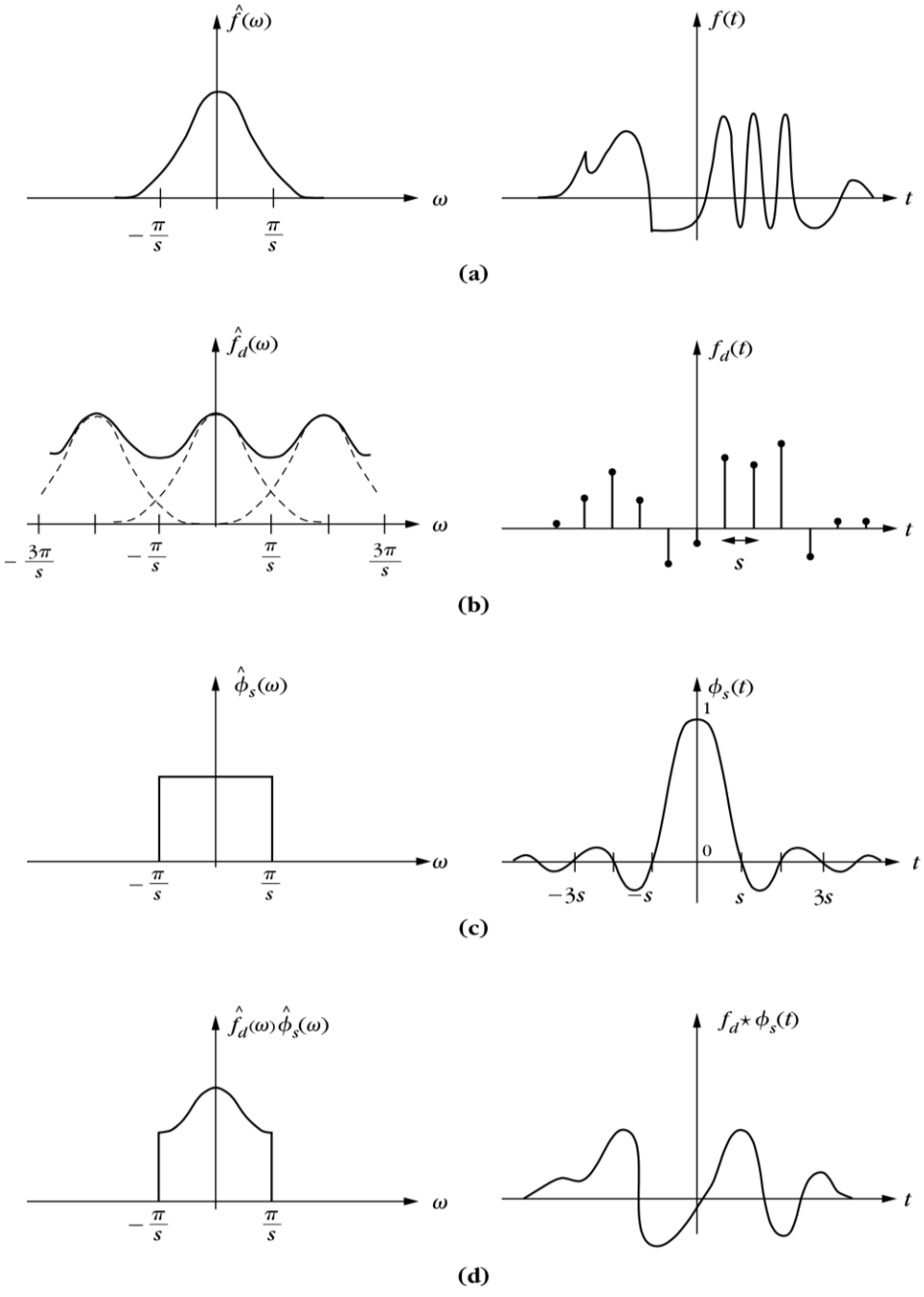


Figure 4: (a) Signal f and its Fourier transform $\widehat{f}$. (b) Aliasing produced by an overlapping of $\widehat{f}(\omega - 2k\pi/s)$ for different k, shown with dashed lines. (c) Low pass filter ϕ_s and its Fourier transform. (d) The filtering $f * \phi_s(t)$ which creates a low frequency signal that is different from f. Notice that non-differentiable singular points are smoothed out, and that the high frequency oscillations on the positive horizontal axis are replaced with a single bump. Taken from Figure 3.2 of *A Wavelet Tour of Signal Processing*.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 05: Fourier Series

January 23, 2020

Lecturer: Matthew Hirn

3.2 Fourier Series

Section 3.2.2 of A Wavelet Tour of Signal Processing.

Let the sampling rate be $s = 1$ now, which gives samples $\{f(n)\}_{n\in\mathbb{Z}}$ of a signal $f(t)$. Previously we defined

$$f_d(t) = \sum_{n\in\mathbb{Z}} f(n)\delta(t-n)$$

and observed that

$$\widehat{f_d}(\omega) = \sum_{n\in\mathbb{Z}} f(n)e^{-in\omega}$$

This is a *Fourier series*. Clearly $\widehat{f_d}(\omega)$ is 2π periodic, and thus it is uniquely determined by its restriction to $[-\pi,\pi]$. This motivates defining Fourier series on $\boldsymbol{\ell}^1$ and $\boldsymbol{\ell}^2$, which will allow us to represent $f_d(t)$ as a sequence $a = (a[n])_{n\in\mathbb{Z}} \in \boldsymbol{\ell}^p$ with $a[n] = f(n)$. Define

$$\boldsymbol{\ell}^p = \left\{ a = (a[n])_{n\in\mathbb{Z}} : a[n] \in \mathbb{C} \text{ and } \sum_{n\in\mathbb{Z}} |a[n]|^p < \infty \right\}, \quad 0 < p < \infty$$

and

$$\boldsymbol{\ell}^\infty = \left\{ a = (a[n])_{n\in\mathbb{Z}} : a[n] \in \mathbb{C} \text{ and } \sup_{n\in\mathbb{Z}} |a[n]| < \infty \right\}$$

Define the Fourier transform of $a \in \boldsymbol{\ell}^p$ as:

$$\mathcal{F}(a)(\omega) = \widehat{a}(\omega) = \sum_{n\in\mathbb{Z}} a[n]e^{-in\omega}, \quad \omega \in [-\pi,\pi]$$

The Fourier transform of $a \in \boldsymbol{\ell}^1$ is a Fourier series; it is analogous to the Fourier transform of $f \in \mathbf{L}^1(\mathbb{R})$. The spaces $\mathbf{L}^p[-\pi,\pi]$ are defined the same as $\mathbf{L}^p(\mathbb{R})$, except that the domain $\mathbb{R}$ is replaced with $[-\pi,\pi]$, and we normalize the norm so that for $A \in \mathbf{L}^p[-\pi,\pi]$ we have

$$\|A\|_p = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |A(\omega)|^p \, d\omega \right)^{1/p}$$

For $\mathbf{L}^2[-\pi,\pi]$ we have the inner product defined as:

$$\langle A, B \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega)B^*(\omega) \, d\omega$$

It is easy to see that $\mathcal{F} : \boldsymbol{\ell}^1 \to \mathbf{L}^\infty[-\pi, \pi]$ and with a little more work (see Theorem 3.3 below) that $\mathcal{F} : \boldsymbol{\ell}^2 \to \mathbf{L}^2[-\pi, \pi]$, which mirrors our results for the Fourier transform defined on $\mathbf{L}^1(\mathbb{R})$ and $\mathbf{L}^2(\mathbb{R})$. Further developing the parallel story, Theorem 3.3 below shows that the family of functions $\{e_n\}_{n\in\mathbb{Z}}$ with

$$e_n(\omega) = e^{-in\omega}$$

is an orthonormal basis for $\mathbf{L}^2[-\pi, \pi]$. It follows that $\mathcal{F} : \boldsymbol{\ell}^2 \to \mathbf{L}^2[-\pi, \pi]$ is a bijection, and hence invertible.

Theorem 3.3. *The family of functions $\{e_n\}_{n\in\mathbb{Z}}$ is a an orthonormal basis for $\mathbf{L}^2[-\pi, \pi]$.*

Proof. The proof that $\{e_n\}_{n\in\mathbb{Z}}$ are orthonormal is by direct calculation of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-in\omega}e^{im\omega}\,d\omega = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

Now we must show that linear expansions of $\{e_n\}_{n\in\mathbb{Z}}$ are dense in $\mathbf{L}^2[-\pi, \pi]$. This means we need to show the following: Let $A \in \mathbf{L}^2[-\pi, \pi]$, $N > 0$ and define the partial Fourier series of A as:

$$S_N(\omega) = \sum_{n=-N}^{N} \langle A, e_n\rangle e^{-in\omega}, \quad \langle A, e_n\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\omega)e^{in\omega}\,d\omega$$

We need to show that for each $\varepsilon > 0$, there exists $N > 0$ such that

$$\|A - S_N\|_2 < \varepsilon$$

which will mean that $\lim_{N\to\infty} S_N = A$ (in the $\mathbf{L}^2$ sense), which we can write as

$$A(\omega) = \sum_{n\in\mathbb{Z}} \langle A, e_n\rangle e^{-in\omega}$$

To prove this we will use some facts about periodic functions; the proofs of these results can be found in [2]. To start, define a trigonometric polynomial $P(\omega)$ as any function

$$P(\omega) = \sum_{n\in\mathbb{Z}} a_n e^{-in\omega}$$

with only a finite number of coefficients a_n being non-zero. The degree of P is defined as the largest value $|n|$ such that $a_n \neq 0$. One fact we will need is that any function $\phi \in \mathbf{C}[-\pi, \pi]$ with $\phi(-\pi) = \phi(\pi)$ can be uniformly approximated by trigonometric polynomials. That is, for each such ϕ and each $\varepsilon > 0$ there exists P such that

$$|\phi(\omega) - P(\omega)| \leq \varepsilon, \quad \forall\ -\pi \leq \omega \leq \pi$$

A second fact we will need is that for $A \in \mathbf{L}^2[-\pi,\pi]$ and $\varepsilon > 0$, we can find a function $\phi \in \mathbf{C}[-\pi,\pi]$ with $\phi(-\pi) = \phi(\pi)$ such that

$$\|\phi\|_\infty \leq \|A\|_\infty$$

and

$$\|A - \phi\|_2 \leq \varepsilon^2$$

We also remark that $A \in \mathbf{L}^2[-\pi,\pi]$ implies that $A \in \mathbf{L}^p[-\pi,\pi]$ for any since the length of $[-\pi,\pi]$ is finite.

Now for the remainder of the proof. Since the family $\{e_n\}_{n\in\mathbb{Z}}$ is orthonormal, we must have

$$A - S_N \perp e_n, \quad \forall\, |n| \leq N$$

from which it follows that $A - S_N \perp P_N$, where P_N is any trigonometric polynomial of degree N. Taking $P_N = S_N$, this in turn gives:

$$\|A\|_2^2 = \|A - S_N + S_N\|_2^2 = \|A - S_N\|^2 + \|S_N\|^2$$

We also note that for any trigonometric polynomial P_N, we have

$$\|A - S_N\|_2 \leq \|A - P_N\|_2 \tag{11}$$

with equality only when $P_N = S_N$. Indeed:

$$A - P_N = A - S_N + \underbrace{(S_N - P_N)}_{\widetilde{P}_N}$$

which implies that

$$\|A - P_N\|_2^2 = \|A - S_N\|_2^2 + \|\widetilde{P}_N\|_2^2$$

from which the inequality (11) follows.

We now complete the proof. First consider a $\phi \in \mathbf{C}[-\pi,\pi]$ with $\phi(-\pi) = \phi(\pi)$. Given $\varepsilon > 0$, we can find a trigonometric polynomial P_M with degree M such that

$$|\phi(\omega) - P_M(\omega)| < \varepsilon, \quad \forall\ -\pi \leq \omega \leq \pi$$

Therefore:

$$\begin{aligned}\|\phi - P_M\|_2^2 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |\phi(\omega) - P_M(\omega)|^2\, d\omega \\ &\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} \varepsilon^2\, d\omega \leq \varepsilon^2\end{aligned}$$

Thus we have $\|\phi - P_M\| \leq \varepsilon$. But since partial Fourier series are the best approximation (11), we then conclude that

$$\|\phi - S_N(\phi)\|_2 \leq \|\phi - P_M\|_2 \leq \varepsilon, \quad \forall\, N \geq M$$

Now let us return to the case of general $A \in \mathbf{L}^2[-\pi, \pi]$. For $\varepsilon > 0$, approximate A with a $\phi \in \mathbf{C}[-\pi, \pi]$, $\phi(-\pi) = \phi(\pi)$, such that $\|A - \phi\|_2 \leq \varepsilon$. Approximate ϕ with a trigonometric polynomial P_M, as before, to obtain:

$$\|A - P_M\|_2 \leq \|A - \phi\|_2 + \|\phi - P_M\|_2 \leq 2\varepsilon$$

Now again use the best approximation inequality (11) to conclude that:

$$\|A - S_N\|_2 \leq 2\varepsilon, \quad \forall\, N \geq M$$

□

Theorem 3.3 proves that any periodic function $A \in \mathbf{L}^2[-\pi, \pi]$ can be written as

$$A(\omega) = \sum_{n \in \mathbb{Z}} c_n e^{-in\omega} \tag{12}$$

with

$$c_n = \langle A, e_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\omega) e^{in\omega}\, d\omega \tag{13}$$

In particular, if we start with $a \in \boldsymbol{\ell}^2$ and compute its Fourier series:

$$\widehat{a}(\omega) = \sum_{n \in \mathbb{Z}} a[n] e^{-in\omega}$$

then we must have

$$a[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widehat{a}(\omega) e^{in\omega}\, d\omega$$

which is a type of Fourier inversion formula for Fourier series. Additionally, if $a[n] = f(n)$ for some signal f, then from the beginning of this section we see that $\widehat{f_d}(\omega) = \widehat{a}(\omega)$ and

$$f(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widehat{f_d}(\omega) e^{in\omega}\, d\omega$$

which gives Fourier inversion for the samples $f(n)$ from $\widehat{f_d}(\omega)$. We have a similar distributional version, which requires defining one more distributional Fourier transform. Recall that for $\delta_\tau(t) = \delta(t - \tau)$, we defined the Fourier transform as $\widehat{\delta}_\tau(\omega) = e^{-i\omega\tau}$. Given this, if we set $e_\xi(t) = e^{i\xi t}$, it makes sense to define its Fourier transform as $\widehat{e}_\xi(\omega) = 2\pi\delta(\omega - \xi)$. Indeed, e_ξ is a perfect harmonic vibrating at frequency ξ. Using this fact, if we compute the inverse Fourier transform of $\widehat{f_d}(\omega)$ in the distributional sense, it is clear we will get back $f_d(t)$. More generally, if we start $A \in \mathbf{L}^2[-\pi, \pi]$ and write it as in (12), and compute the distributional inverse Fourier transform of A, we will get

$$\mathcal{F}^{-1}(A)(t) = \sum_{n \in \mathbb{Z}} \langle A, e_n \rangle \delta(t - n)$$

Of course this is equivalent to computing the Fourier series inversion of A and getting a sequence $(\langle A, e_n\rangle)_{n\in\mathbb{Z}} \in \boldsymbol{\ell}^2(\mathbb{R})$, but it is sometimes convenient to use one over the other.

We also have the following version of the Plancheral formula:

$$\|a\|^2 = \sum_{n\in\mathbb{Z}} |a[n]|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\widehat{a}(\omega)|^2\, d\omega = \|\widehat{a}\|^2$$

Define the convolution of $a, b \in \boldsymbol{\ell}^1$ as:

$$a * b[n] = \sum_{m\in\mathbb{Z}} a[m]b[n-m]$$

We have a convolution theorem for $\boldsymbol{\ell}^1$ sequences as well:

Theorem 3.4. *Let $a, b \in \boldsymbol{\ell}^1$. Then $a * b \in \boldsymbol{\ell}^1$ and*

$$\widehat{a * b}(\omega) = \widehat{a}(\omega)\widehat{b}(\omega)$$

Remark 3.5. (see also [4]) In this section we started with a sequence $a \in \boldsymbol{\ell}^2$ and defined its Fourier transform as the Fourier series with coefficients $a[n]$. We saw the resulting Fourier series defines a periodic function $A(\omega)$ on $\mathbf{L}^2[-\pi, \pi]$. This was motivated by discrete sampling of a function $f \in \mathbf{L}^2(\mathbb{R})$. In many (pure) harmonic analysis texts, though, one starts with a periodic function A on $[-\pi, \pi]$ and computes its Fourier coefficients as the inner products of A with e_n, as in (13). In Theorem 3.3 we proved that for any $A \in \mathbf{L}^2[-\pi, \pi]$, the partial Fourier series

$$S_N(A)(\omega) = \sum_{|n|\le N} \langle A, e_n\rangle e^{-in\omega}$$

converges to A in $\mathbf{L}^2$ norm as $N \to \infty$, i.e.,

$$\lim_{N\to\infty} \|S_N(A) - A\|_2 = 0$$

In fact, the same is true for any $A \in \mathbf{L}^p[-\pi, \pi]$, for $1 < p < \infty$, that is:

$$\lim_{N\to\infty} \|S_N(A) - A\|_p = 0$$

Note this is not true for $p = 1$ or $p = \infty$! The Plancheral formula above shows that

$$\|A\|_2^2 = \sum_{n\in\mathbb{Z}} |\langle A, e_n\rangle|^2$$

that is, if we know the amplitudes of the Fourier coefficients of A, then we can deduce its $\mathbf{L}^2$ norm. It turns out the same is not true for $2 < p < \infty$; knowing the amplitudes is not enough, the phases play a critical role. Let us consider a sequence $a \in \boldsymbol{\ell}^2$. Let ε_n, $n \in \mathbb{Z}$,

be a sequence of independent Bernoulli random variables, meaning they take value $+1$ with probability $1/2$ and value -1 with probability $1/2$. Then the random Fourier series

$$\sum_{|n|\leq N} \varepsilon_n a[n] e^{-in\omega}$$

converges to a function that, for $2 < p < \infty$, belongs to each $\mathbf{L}^p[-\pi, \pi]$, almost surely (i.e., for almost all choices of the sequence ε_n). But it does not converge for all random sequences! If we want to estimate the $\mathbf{L}^p$ norm of A, breaking it down into its individual Fourier coefficients goes too far. The right approach is to break the frequencies down into dyadic packets. These are defined as:

$$\Delta_j A(\omega) = \sum_{2^j \leq |n| < 2^{j+1}} \langle A, e_n \rangle e^{-in\omega}, \quad j \in \mathbb{N}$$

Then using Littlewood-Paley theory, we have that the norm $\|A\|_p$ is equivalent to

$$|\langle A, e_0 \rangle| + \left\| \left(\sum_{j=0}^{\infty} |\Delta_j A|^2 \right)^{1/2} \right\|_p$$

This means, in particular, if $\sum_{j=0}^{\infty} \Delta_j A \in \mathbf{L}^p[-\pi, \pi]$, then so does $\sum_{j=0}^{\infty} \varepsilon_j \Delta_j A$ for all choices of $\varepsilon_j = \pm 1$. This type of idea, breaking the frequencies down into dyadic packets, is at the heart of wavelet analysis, which we will get to later. We will also rely on this type of Littlewood-Paley analysis when we want to analyze $\mathbf{C}^\alpha$ functions with wavelets.

Exercise 14. Read Section 3.2 of *A Wavelet Tour of Signal Processing.*

Exercise 15. A rectifier computes $g(t) = |f(t)|$ for recovering the envelope of modulated signals.

(a) Show that if $f(t) = h(t)\sin(\omega_0 t)$ with $h \in \mathbf{L}^1(\mathbb{R})$, $h(t) \geq 0$ and $\omega_0 > 0$, then $g(t) = |f(t)|$ satisfies

$$\widehat{g}(\omega) = \frac{2}{\pi} \sum_{n=-\infty}^{+\infty} \frac{\widehat{h}(\omega - 2n\omega_0)}{4n^2 - 1}$$

Hint: Let $A(t)$ be a 2π periodic function. By Theorem 3.3 we can write

$$A(t) = \sum_{n \in \mathbb{Z}} c_n e^{int}$$

with

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(t) e^{-int}\, dt$$

Compute these Fourier coefficients c_n.

(b) Suppose that $\widehat{h}(\omega) = 0$ for $|\omega| \geq \omega_0$. Find ϕ such that $h(t) = \phi * g(t)$.

Exercise 16. An *interpolation function* $f(t)$ satisfies $f(n) = \delta(n)$ for any $n \in \mathbb{Z}$.

(a) Prove that

$$\sum_{n\in\mathbb{Z}} \widehat{f}(\omega + 2n\pi) = 1 \quad \Longleftrightarrow \quad f \text{ is an interpolation function}$$

(b) Suppose that

$$f(t) = \sum_{n\in\mathbb{Z}} a[n]\theta(t-n), \quad a \in \boldsymbol{\ell}^2, \ \theta \in \mathbf{L}^2(\mathbb{R})$$

Find $\widehat{a}(\omega)$ as a function of $\widehat{\theta}(\omega)$ so that $f(t)$ is an interpolation function. Relate $\widehat{f}(\omega)$ to $\widehat{\theta}(\omega)$, and give a sufficient condition on $\widehat{\theta}$ to guarantee that $f \in \mathbf{L}^2(\mathbb{R})$.

Exercise 17. Let $g \in \boldsymbol{\ell}^1$ and set $h[n] = (-1)^n g[n]$. Relate $\widehat{h}(\omega)$ to $\widehat{g}(\omega)$. If g is a low pass filter (meaning that $\widehat{g}(\omega)$ is concentrated around 0), then what kind of filter is h? (i.e., where is its support concentrated?)

Exercise 18. Let $b \in \boldsymbol{\ell}^1$. A *decimation* of b computes a signal $a \in \boldsymbol{\ell}^1$ with $a[n] = b[Mn]$ for $M > 1$ ($M \in \mathbb{Z}$).

(a) Show that

$$\widehat{a}(\omega) = \frac{1}{M} \sum_{k=0}^{M-1} \widehat{b}(M^{-1}(\omega - 2k\pi))$$

(b) Give a sufficient condition on $\widehat{b}(\omega)$ to recover b from a and give the interpolation formula that recovers $b[n]$ from a.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 06: Finite Length Signals, DFT, and FFT

January 28, 2020

Lecturer: Matthew Hirn

3.3 Finite Length Signals

In practice we cannot store an infinite number of samples $\{f(n)\}_{n\in\mathbb{Z}}$ of a signal f; instead we can only keep a finite number of samples, say $\{f(n)\}_{0\leq n<N}$. We thus must amend our definition of the Fourier transform as well as convolution, which will lead to the Discrete Fourier Transform (DFT) and circular convolution. One thing that will arise is that regardless of whether the original signal f is periodic, we will be forced to think of the finite sampling $(f(n))_{0\leq n<N}$ as a discrete periodic signal with period N. This will lead to border effects which must be accounted for. However, the circular convolution theorem and Fast Fourier Transform will allow for fast computations of convolution operators.

Let $x, y \in \mathbb{C}^N$, which are vectors of length N, e.g., N samples of a signal f such that $x[n] = f(n)$ for $0 \leq n < N$. The inner product between x and y is:

$$\langle x, y\rangle = \sum_{n=0}^{N-1} x[n]y^*[n]$$

We must replace the sinusoids $e^{it\omega}$ ($t \in \mathbb{R}$) and $e^{in\omega}$ ($n \in \mathbb{Z}$), which are continuous in the frequency variable ω, with discrete counterparts. The variable ω is replaced with an index k with $0 \leq k < N$:

$$e_k[n] = \exp\left(\frac{2\pi ikn}{N}\right), \quad 0 \leq n, k < N \tag{14}$$

The *Discrete Fourier Transform (DFT)* of x is defined as:

$$\widehat{x}[k] = \langle x, e_k\rangle = \sum_{n=0}^{N-1} x[n]\exp\left(-\frac{2\pi ikn}{N}\right), \quad 0 \leq k < N$$

The following theorem shows that the set of vectors $\{e_k\}_{0\leq k<N}$ is an orthogonal basis for $\mathbb{C}^N$.

Theorem 3.6. *The family of vectors $\{e_k\}_{0\leq k<N}$ as defined in (14) is orthogonal basis for $\mathbb{C}^N$.*

Thus the DFT is a bijection and hence invertible. Since $\|e_k\|^2 = N$ for all k, it follows from Theorem 3.6 that x can be represented in the orthogonal basis $\{e_k\}_{0\leq k<N}$ as:

$$x[n] = \sum_{k=0}^{N-1} \frac{\langle x, e_k\rangle}{\|e_k\|^2} e_k[n] = \frac{1}{N}\sum_{k=0}^{N-1} \widehat{x}[k]\exp\left(\frac{2\pi ikn}{N}\right)$$

This gives the inverse DFT.

We would like a convolution theorem for the DFT similar to the convolution theorem for $\mathbf{L}^1(\mathbb{R})$ functions and $\boldsymbol{\ell}^1(\mathbb{R})$ sequences. We define convolution of $x, y \in \mathbb{C}^N$ by first extending them to signals $x_0, y_0 \in \boldsymbol{\ell}^1(\mathbb{R})$ defined as:

$$x_0[n] = \begin{cases} x[n] & 0 \leq n < N \\ 0 & n < 0 \text{ or } n \geq N \end{cases}$$

We then define the convolution of x and y as the convolution of x_0 and y_0, and keep only the values with a chance of being nonzero:

$$x * y[n] = \sum_{m \in \mathbb{Z}} x_0[m] y_0[n-m], \quad 0 \leq n < 2N-1$$

In practice, there will be many times when you want to compute such convolutions. Indeed, if x and y are discrete samplings of non-periodic signals f and g, respectively, computing $x * y$ will give a discrete approximation for $f * g$. However, for computational reasons, we will often not want to compute discrete convolutions directly (more on this in a bit). Indeed, it will be better to compute such convolutions "in frequency," which will require a convolution theorem for the DFT. However, the discrete sinusoids $\{e_k\}_{0 \leq k < N}$ are not eigenvectors of discrete convolution operators $Lx = x * h$. The vectors e_k are periodic, but the standard convolution is not; indeed, it extends the vectors x, y to a twice longer vector $x * y$. We therefore define a periodic version of convolution, which is called *circular convolution*.

To define circular convolution, rather than extending x and y with zeros, we will extend them with a periodization over N samples:

$$x_p[n] = x[n \mod N], \quad n \in \mathbb{Z}$$

The circular convolution is defined as:

$$x \circledast y[n] = \sum_{m=0}^{N-1} x_p[m] y_p[n-m]$$

Note that $x \circledast y \in \mathbb{C}^N$. One then has the following circular convolution theorem:

Theorem 3.7. *If $x, y \in \mathbb{C}^N$, then*

$$\widehat{x \circledast y}[k] = \widehat{x}[k]\widehat{y}[k]$$

The key to this theorem, and the DFT more generally, is that since the discrete sinusoids e_k are periodic vectors with period N, the DFT treats all vectors $x \in \mathbb{C}^N$ as periodic vectors with period N. This manifests in the convolution theorem by requiring us to utilize circular convolutions. However, it also means that when computing DFTs, we always need to think of x as a periodic vector with period N. In particular, seemingly "smooth" vectors such as $x[n] = n$ actually have very sharp transitions once made periodic, since $x[N-1] = N-1$ and $x[N] = x[0] = 0$; see Figure 5.

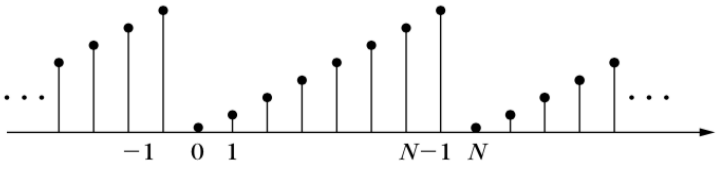


Figure 5: Periodization of the ramp vector $x[n] = n$ on $\mathbb{R}^N$. Taken from Figure 3.3 of *A Wavelet Tour of Signal Processing*.

Remark 3.8. Notice as well, we now have the following correspondences:

- Signals $f \in \mathbf{L}^2(\mathbb{R})$ and $\widehat{f} \in \mathbf{L}^2(\mathbb{R})$, otherwise no restrictions.
- Discrete, but infinite samplings $a \in \boldsymbol{\ell}^2(\mathbb{Z})$ with $a[n] = f(n)$, and $\widehat{a} \in \mathbf{L}^2[-\pi, \pi]$ a 2π periodic Fourier series in which

$$\widehat{a}(\omega) = \sum_{n \in \mathbb{Z}} \widehat{f}(\omega - 2\pi n)$$

- Discrete, finite samplings $x \in \mathbb{C}^N$ which must be considered as N-periodic to do any frequency calculation (e.g., Theorem 3.7). In particular, if

$$x[n] = \sum_{p \in \mathbb{Z}} a[n - pN]$$

then $\widehat{x} \in \mathbb{C}^N$ with

$$\widehat{x}[k] = \widehat{a}(2\pi k/N) \qquad \text{(see Exercise 22)}$$

Thus a discrete, but infinite sampling of f in time/space periodizes its Fourier transform, possibly leading to aliasing. A finite, discrete sampling in frequency also periodizes the signal in time/space, leading to possible border effects. We must account for both of these in practice.

The circular convolution theorem will be very important for opening up fast algorithms for computing $x \circledast y$. This will be made possible by the *Fast Fourier Transform (FFT)*, which we now describe. To motivate the algorithm, recall the DFT:

$$\widehat{x}[k] = \sum_{n=0}^{N-1} x[n] \exp\left(-\frac{2\pi i k n}{N}\right), \quad 0 \le k < N$$

and observe that it requires N^2 (complex) multiplications and additions (N for each k). The FFT algorithm reduces this to $O(N \log_2 N)$.

The FFT algorithm works through a divide and conquer approach; in these notes I will describe the radix-2 decimation in time (DIT) algorithm. This is a recursive algorithm.

Given x we divide the DFT summation into two sums, one for the even indices of x and one for the odd indices of x:

$$\begin{aligned}\widehat{x}[k] &= \sum_{n=0}^{N/2-1} x[2n] \exp\left(\frac{-2\pi i k(2n)}{N}\right) + \sum_{n=0}^{N/2-1} x[2n+1] \exp\left(\frac{-2\pi i k(2n+1)}{N}\right) \\ &= \sum_{n=0}^{N/2-1} x[2n] \exp\left(\frac{-2\pi i k n}{N/2}\right) + e^{-2\pi i k/N} \sum_{n=0}^{N/2-1} x[2n+1] \exp\left(\frac{-2\pi i k n}{N/2}\right)\end{aligned}$$

The second line looks like the sum of two DFTs of length $N/2$ signals. Indeed, define $x_{\mathrm{e}}, x_{\mathrm{o}} \in \mathbb{R}^{N/2}$ as:

$$\begin{aligned}x_{\mathrm{e}}[n] &= x[2n], \quad 0 \le n < N/2 \\ x_{\mathrm{o}}[n] &= x[2n+1], \quad 0 \le n < N/2\end{aligned}$$

and notice that we have

$$\begin{aligned}\widehat{x}_{\mathrm{e}}[k] &= \sum_{n=0}^{N/2-1} x[2n] \exp\left(\frac{-2\pi i k n}{N/2}\right), \quad 0 \le k < N/2 \\ \widehat{x}_{\mathrm{o}}[k] &= \sum_{n=0}^{N/2-1} x[2n+1] \exp\left(\frac{-2\pi i k n}{N/2}\right), \quad 0 \le k < N/2\end{aligned}$$

This allows us to recover $\widehat{x}[k]$ for $0 \le k < N/2$ as:

$$\widehat{x}[k] = \widehat{x}_{\mathrm{e}}[k] + e^{-2\pi i k/N} \widehat{x}_{\mathrm{o}}[k], \quad 0 \le k < N/2 \tag{15}$$

For the frequencies $N/2 \le k < N$, we use the fact that the DFT is periodic and observe that

$$\widehat{x}_{\mathrm{e}}[k + N/2] = \widehat{x}_{\mathrm{e}}[k] \quad \text{and} \quad \widehat{x}_{\mathrm{o}}[k + N/2] = \widehat{x}_{\mathrm{o}}[k]$$

We thus obtain:

$$\begin{aligned}0 \le k < N/2, \quad \widehat{x}[k + N/2] &= \widehat{x}_{\mathrm{e}}[k] + e^{-2\pi i (k+N/2)/N} \widehat{x}_{\mathrm{o}}[k] \\ &= \widehat{x}_{\mathrm{e}}[k] - e^{-2\pi i k/N} \widehat{x}_{\mathrm{o}}[k]\end{aligned} \tag{16}$$

Putting together (15) and (16) we obtain $\widehat{x}[k]$ for all $0 \le k < N$. Notice that already we have reduced computations. Indeed, the one step algorithm proceeds by first dividing x into even and odd indices signals, computing the two length $N/2$ DFTs, and recombining as above. The two length N DFTs cost $2(N/2)^2 = N^2/4$ multiplications and additions, and the combination costs N additions and multiplications. Thus we have replaced N^2 complex multiplications and additions with $N^2/2 + N$ complex multiplications and additions, which is already better for $N \ge 3$.

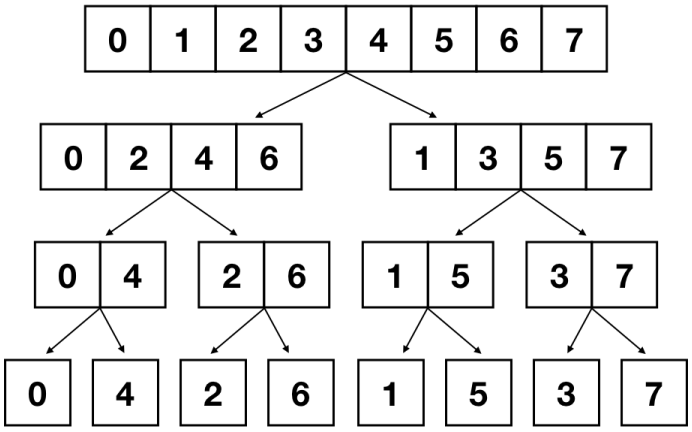


Figure 6: Recursive subdivision scheme of the FFT algorithm for $N = 8$.

Now for simplicity, suppose $N = 2^p$. The $O(N \log_2 N)$ FFT algorithm is obtained by recursively subdividing the original signal x, according to the same procedure as outlined above into "even" and "odd" components, until we have N signals of length one; Figure 6 illustrates the idea for $N = 8$. The algorithm then "computes" N length one DFTs - notice that these just return the value of each length one signal, so no computation is actually performed. The algorithm then forms $N/2$ length two DFTs the next level up by combining the pairs of length one DFTs that have the same parent signal, and multiplying the "odd" length one signal by the appropriate complex value before combination. At each level we incur a cost of $O(N)$, and there are $p = \log_2 N$ levels; thus the total cost of the algorithm is $O(N \log_2 N)$.

The FFT algorithm is remarkable for turning an $O(N^2)$ calculation into an $O(N \log N)$ calculation with no loss of accuracy. For this reason it is a pillar of digital signal processing. However, it is fundamentally an algebraic property of the DFT, based on symmetries. As such, the algorithm is "fragile," and in particular, if you do not uniformly sample your signal f, you cannot apply the FFT algorithm. That however is a discussion for another day (or class).

The FFT algorithm allows us to compute convolutions $x * y$ fast. Suppose the $x, y \in \mathbb{C}^N$ as usual; if we compute $x * y$ directly it will cost us $O(N^2)$ calculations. In order to calculate the non-circular convolution faster, we can use the circular convolution Theorem 3.7, which will allow us to leverage the FFT algorithm. The main idea is that instead of computing $x * y$ directly, we compute $\widehat{x}$ and $\widehat{y}$, each costing $O(N \log N)$ calculations, then we compute the multiplication $\widehat{x}[k]\widehat{y}[k]$ for $0 \leq k < N$, costing $O(N)$ calculations, and then we compute the inverse Fourier transform of $(\widehat{x}[k]\widehat{y}[k])_{0 \leq k < N}$ with another FFT, which costs $O(N \log N)$ calculations; the total run time of the algorithm if $O(N \log N)$, which in practice (depending

upon the exact FFT algorithm you use) will be better for $N \geq 32$. One thing that we have not addressed though, is that the convolution theorem for finite length signal applies to circular convolution. If we do not account for this, we will run into border effects since we will be computing $x \circledast y$ instead of $x * y$. To fix this issue, we zero pad x and y by defining:

$$x_0[n] = \begin{cases} x[n] & 0 \leq n < N \\ 0 & N \leq n < 2N \end{cases}$$

The signal $x_0 \in \mathbb{C}^{2N}$ and is just the signal x but with N zeros appended to the back of it. One can then verify that:

$$x_0 \circledast y_0[n] = x * y[n], \quad 0 \leq n < 2N$$

Thus we apply the fast FFT based algorithm to x_0 and y_0 (rather than x and y) to obtain $x * y$ in $O(N \log N)$ time.

Exercise 19. Read Section 3.3 of *A Wavelet Tour of Signal Processing.*

Exercise 20. Read Section 3.4 of *A Wavelet Tour of Signal Processing.*

Exercise 21. Let $\widehat{x}[k]$ be the DFT of a finite signale $x \in \mathbb{C}^N$. Define a signal $y \in \mathbb{C}^{2N}$ by:

$$\widehat{y}[N/2] = \widehat{y}[3N/2] = \widehat{x}[N/2]$$

and

$$\widehat{y}[k] = \begin{cases} 2\widehat{x}[k] & 0 \leq k < N/2 \\ 0 & N/2 < k < 3N/2 \\ 2\widehat{x}[k-N] & 3N/2 < k < 2N \end{cases}$$

Prove that y is an interpolation of x that satisfies $y[2n] = x[n]$ for all $0 \leq n < N$.

Exercise 22. We want to compute numerically the Fourier transform of $f(t)$. Let $a[n] = f(n)$ for $n \in \mathbb{Z}$ be the countably infinite discrete sampling of f and let $x \in \mathbb{C}^N$ be the periodization of a over the period of length N:

$$x[n] = \sum_{p \in \mathbb{Z}} a[n - pN]$$

(a) Prove that the DFT of x is related to the Fourier series of a and to the Fourier transform of f by the following formula:

$$\widehat{x}[k] = \widehat{a}(2\pi k/N) = \sum_{\ell \in \mathbb{Z}} \widehat{f}\left(\frac{2\pi k}{N} - 2\pi\ell\right)$$

(b) Suppose that $|f(t)|$ and $|\widehat{f}(\omega)|$ are negligible when $t \notin [-t_0, t_0]$ and $\omega \notin [-\omega_0, \omega_0]$. Relate N to t_0 and ω_0 so that one can compute an approximate value of $\widehat{f}(\omega)$ for all $\omega \in \mathbb{R}$ by interpolating the samples $\widehat{x} \in \mathbb{C}^N$. Is it possible to compute exactly $\widehat{f}(\omega)$ with such an interpolation formula?

Exercise 23. We are going to implement the FFT and fast convolution algorithms:

(a) Implement the DFT algorithm (programming language of your choice). Record the runtime for many values of N, and plot it as a function of N. Do you see the quadratic scaling? Turn in your code and plot(s).

(b) Make precise the $O(N \log N)$ FFT algorithm described above and implement it on your own for $N = 2^p$. Test the algorithm for accuracy by comparing its outputs to the outputs of your DFT algorithm. Test the algorithm for speed by comparing the runtime for numerous values of N to the runtimes your recorded for the DFT. For which value of N does your FFT algorithm become faster? Turn in your code, at least one output showing that the DFT and FFT codes produce the same results, and a plot of the FFT runtimes as a function of N (you can combine this plot with the DFT plot).

(c) Using either your own FFT and inverse FFT code, or built in code (in Matlab or Python, for example) since you are not required to write your own inverse FFT code, implement an algorithm to compute $x*y$ (for $x, y \in \mathbb{C}^N$) in $O(N \log N)$ time. Verify the accuracy by comparing against convolution code that computes $x * y$ directly (either your own code, or built in code), and compare the runtimes. For which N is your $O(N \log N)$ convolution code faster?

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 07: Time-Frequency Analysis and Windowed Fourier

January 30, 2020

Lecturer: Matthew Hirn

4 Time Meets Frequency

Chapter 4 of A Wavelet Tour of Signal Processing [1].

4.1 Time Frequency Atoms

Section 4.1 of A Wavelet Tour of Signal Processing [1].

A linear *time frequency transform* correlates the signal $f(t)$ with a dictionary of waveforms that are concentrated in time and frequency; these waveforms are called time frequency atoms. Denote a general dictionary of time frequency atoms by:

$$\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}, \quad \phi_\gamma \in \mathbf{L}^2(\mathbb{R}), \ \ \|\phi_\gamma\|_2 = 1$$

where Γ is a (multi)-index set. The time frequency transform of $f \in \mathbf{L}^2(\mathbb{R})$ in the dictionary $\mathcal{D}$ computes

$$\Phi f(\gamma) = \langle f, \phi_\gamma \rangle = \int_{\mathbb{R}} f(t) \phi_\gamma^*(t) \, dt$$

Recall that the Fourier transform of f is:

$$\widehat{f}(\omega) = \langle f, e_\omega \rangle = \int_{\mathbb{R}} f(t) e^{-i\omega t} \, dt \, , \quad e_\omega(t) = e^{i\omega t}$$

It is not a perfect analogue for the time frequency transform Φ since $e_\omega \notin \mathbf{L}^2(\mathbb{R})$, but both transforms analyze f by testing the signal against a family of waveforms. Let us now explore time-frequency transforms in more detail.

Recall the definitions of the time mean u, frequency mean ξ, time variance σ_t^2, and frequency variance σ_ω^2 of a function $f \in \mathbf{L}^2(\mathbb{R})$, first defined when we studied the uncertainty principle in Section 2.4. Apply them to the dictionary $\mathcal{D}$ for each time frequency atom ϕ_γ, and denote the corresponding quantities by

$$u_\gamma, \ \omega_\gamma, \ \sigma_t(\gamma), \ \sigma_\omega(\gamma)$$

The waveform ϕ_γ is essentially supported in time on an interval of length $\sigma_t(\gamma)$, centered at u_γ, while its Fourier transform $\widehat{\phi}_\gamma$ is essentially supported in frequency on an interval

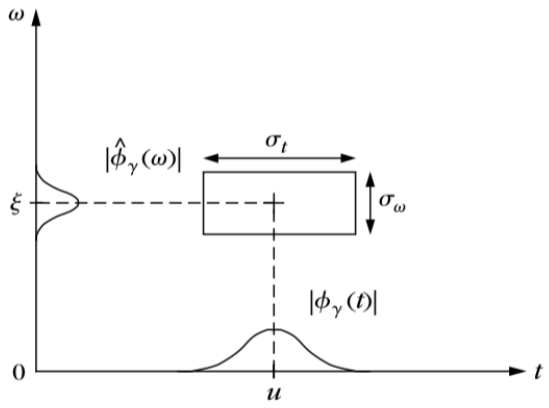


Figure 7: Heisenberg box representing the essential time frequency support of ϕ_γ

of length $\sigma_\omega(\gamma)$, centered at ξ_γ. Thus the joint time frequency support of ϕ_γ in the time frequency plane (t, ω) is given by a Heisenberg box centered at (u_γ, ξ_γ) having time width $\sigma_t(\gamma)$ and frequency width $\sigma_\omega(\gamma)$; see Figure 7.

The Parseval formula (Theorem 2.12) proves that:

$$\Phi f(\gamma) = \int_{\mathbb{R}} f(t)\phi_\gamma^*(t)\, dt = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(\omega)\widehat{\phi}_\gamma^*(\omega)\, d\omega$$

Thus we see that $\Phi f(\gamma)$ only depends upon the values of $f(t)$ and $\widehat{f}(\omega)$ in the Heisenberg box of ϕ_γ. In particular, $\Phi f(\gamma)$ only measures the frequencies of f in a neighborhood of ξ_γ, and it only measures these frequencies in a neighborhood of the time u_γ. Because of the uncertainty principle (Theorem 2.18), we know that

$$\sigma_t(\gamma)\sigma_\omega(\gamma) \geq \frac{1}{2}$$

Thus it is impossible to measure precisely the frequency response $\widehat{f}(\omega_0)$ at the time t_0. The best we can do is measure the time frequency response of f in a Heisenberg box of area $1/2$. Theorem 2.18 proves that the time frequency atoms that achieve this optimal time frequency localization are given by Gabor functions; we will come back to this point shortly when we introduce the windowed Fourier transform.

For pattern recognition and machine learning tasks, it often important to construct time frequency representations that behave well with respect to translations of the signal $f(t)$ (and in 2D, rotations as well). Define $f_u(t) = f(t-u)$ as the translation of f by u, and notice that:

$$\Phi f_u(\gamma) = \int_{\mathbb{R}} f(t-u)\phi_\gamma^*(t)\, dt = \int_{\mathbb{R}} f(t)\phi_\gamma^*(t+u)\, dt = \langle f, \phi_{-u,\gamma} \rangle,$$

where $\phi_{u,\gamma}(t) = \phi_\gamma(t-u)$. This motivates the construction of translation invariant dictionaries. A translation invariant dictionary is obtained by starting with a family of generators

$\{\phi_\gamma\}_{\gamma\in\Gamma}$, and augmenting this family with all translates of each time frequency atom ϕ_γ:

$$\mathcal{D} = \{\phi_{u,\gamma}\}_{u\in\mathbb{R},\gamma\in\Gamma}$$

Set:

$$\overline{\phi}_\gamma(t) = \phi^*_\gamma(-t)$$

The resulting time frequency transform with a translation invariant dictionary is given by:

$$\Phi f(u,\gamma) = \langle f, \phi_{u,\gamma}\rangle = \int_{\mathbb{R}} f(t)\phi^*_\gamma(t-u)\,dt = f * \overline{\phi}_\gamma(u)$$

It thus corresponds to a filtering of f by the time-frequency waveforms $\{\overline{\phi}_\gamma\}_{\gamma\in\Gamma}$.

Exercise 24. Read Section 4.1 of *A Wavelet Tour of Signal Processing.*

4.2 Windowed Fourier Transform

Section 4.2 of A Wavelet Tour of Signal Processing [1].

The Fourier transform $\widehat{f}(\omega)$ tells us every frequency in the signal $f(t)$, but it does not tell us *when* such frequencies are present. For example, in music we hear the time variation of the sound frequencies. Similarly, images with vastly different patterns in them may correspond to different frequencies, localized not over time but space; see the picture of the castle in Figure 8 for an example.

A natural way to account for these localized structures is to localize the Fourier transform with a window function. Let g be a real symmetric window $g(t) = g(-t)$, which has support localized around $t = 0$ (e.g., a Gaussian $g(t) = \frac{1}{\sqrt{2\pi}\sigma}e^{-t^2/2\sigma^2}$). Translations of this window by $u \in \mathbb{R}$, and modulations of this window by the frequency $\xi \in \mathbb{R}$, yield a *Gabor type dictionary*:

$$\mathcal{D} = \{g_{u,\xi}\}_{u,\xi\in\mathbb{R}}, \quad g_{u,\xi}(t) = g(t-u)e^{i\xi t}$$

The window is normalized so that $\|g\|_2 = 1$, which implies that $\|g_{u,\xi}\|_2 = 1$ for all $(u,\xi) \in \mathbb{R}^2$. The resulting *windowed Fourier transform* (also known as the short time Fourier transform, or Gabor transform) is:

$$Sf(u,\xi) = \langle f, g_{u,\xi}\rangle = \int_{\mathbb{R}} f(t)g(t-u)e^{-i\xi t}\,dt$$

Notice that $Sf(u,\xi)$ computes a localized version of the Fourier transform of $f(t)$, in which the Fourier integral is localized around u by the window $g(t-u)$.

The energy density of the windowed Fourier transform is the *spectrogram*:

$$P_S f(u,\xi) = |Sf(u,\xi)|^2 = \left|\int_{\mathbb{R}} f(t)g(t-u)e^{-i\xi t}\,dt\right|^2$$

Figure 8: Picture of a castle, taken from Wikipedia. Different regions of the picture have different patterns, such as the sky, the trees, and the castle itself. These patterns have different frequency responses, which are spatially localized.

The spectrogram removes the phase of $Sf(u,\xi)$ and measures the energy of f in a time frequency neighborhood of (u,ξ) specified by the Heisenberg box of $g_{u,\xi}$. The size of these Heisenberg boxes is in fact independent of (u,ξ), as we now show.

First note that since $g(t)$ is even, $g_{u,\xi}$ is centered at u. The variance around u is:

$$\sigma_t^2 = \int_{\mathbb{R}} (t-u)^2 |g_{u,\xi}(t)|^2 \, dt = \int_{\mathbb{R}} t^2 |g(t)|^2 \, dt$$

The Fourier transform $\widehat{g}$ of g is real and symmetric because g is real and symmetric. We also compute the Fourier transform of $g_{u,\xi}$ as (set $e_\xi(t) = e^{i\xi t}$):

$$\begin{aligned}
\widehat{g}_{u,\xi}(\omega) &= \widehat{g_u \cdot e_\xi}(\omega) \\
&= (2\pi)^{-1} \widehat{g}_u * \widehat{e}_\xi(\omega) \\
&= (2\pi)^{-1} (e_{-u} \cdot \widehat{g}) * 2\pi\delta_\xi(\omega) \\
&= e^{-iu(\omega-\xi)} \widehat{g}(\omega-\xi)
\end{aligned}$$

It follows that $\widehat{g}_{u,\xi}$ is centered at ξ, and

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_{\mathbb{R}} (\omega-\xi)^2 |\widehat{g}_{u,\xi}(\omega)|^2 \, d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} \omega^2 |\widehat{g}(\omega)|^2 \, d\omega$$

These calculations show that the Heisenberg boxes of $g_{u,\xi}$ centered at (u,ξ) with an area $\sigma_t \sigma_\omega$ that is independent of the location (u,ξ). Thus the windowed Fourier transform has

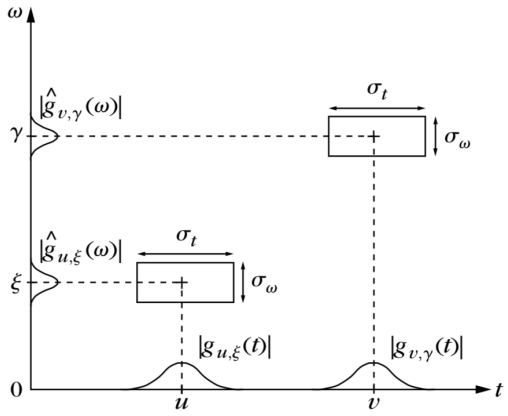


Figure 9: Heisenberg boxes of the windowed Fourier time frequency atoms.

the same resolution across the time frequency plane; see Figure 9. This is one of its defining properties; other time-frequency transforms that we will encounter (e.g., wavelets), will utilize Heisenberg boxes of different dimensions depending on their location in the time-frequency plane.

Exercise 25. Read Section 4.2 of *A Wavelet Tour of Signal Processing*, up to but not including Section 4.2.1.

4.2.1 Parseval for Windowed Fourier

The approach in this section and the next section follows the treatment in [5, Chapter 3]. For a more in depth treatment of the windowed Fourier transform and time frequency analysis, [5] is an excellent resource.

Recall that for a window $g \in \mathbf{L}^2(\mathbb{R})$ the windowed Fourier transform of $f \in \mathbf{L}^2(\mathbb{R})$ is defined as:

$$S_g f(u, \xi) = \int_{\mathbb{R}} f(t) g(t-u) e^{-i\xi t}\, dt$$

Here we write $S_g f$ rather than Sf to emphasize the dependence upon the window choice g. Up till now we have been a little sloppy in that, while we know the windowed Fourier transform is well defined pointwise, we do not know if this transform maps f into some nice functional class. To that end, the next theorem is an analogue of Parseval's formula (Theorem 2.12) for the windowed Fourier transform and for windows g in a subclass of $\mathbf{L}^2(\mathbb{R})$. It shows that $S_g : \mathbf{L}^2(\mathbb{R}) \to \mathbf{L}^2(\mathbb{R}^2)$. Like the original Parseval formula it is also extremely useful.

Theorem 4.1. *Let $f, h \in \mathbf{L}^2(\mathbb{R})$ and let g be a real symmetric function with $g \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ and $\|g\|_2 = 1$. Then:*

$$\langle f, h \rangle = \frac{1}{2\pi} \langle S_g f, S_g h \rangle_{\mathbf{L}^2(\mathbb{R}^2)}$$

Proof. We will need another fundamental result from real analysis, which is Young's inequality. Suppose that $f_1 \in \mathbf{L}^p(\mathbb{R})$, $f_2 \in \mathbf{L}^q(\mathbb{R})$, and

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$$

Then

$$\|f_1 * f_2\|_r \leq \|f_1\|_p \|f_2\|_q \tag{17}$$

Now define f_ξ as $f_\xi(u) = S_g f(u, \xi)$, so that we think of the windowed Fourier transform as a function in u with a parameter ξ. We first show that $f_\xi \in \mathbf{L}^2(\mathbb{R})$ and then compute its Fourier transform. Additionally, set $g_\xi(t) = g(t)e^{i\xi t}$; we can rewrite $f_\xi(u)$ as (using that g is symmetric):

$$\begin{aligned} f_\xi(u) &= \int_{\mathbb{R}} f(t) g(t-u) e^{-i\xi t}\, dt \\ &= e^{-iu\xi} \int_{\mathbb{R}} f(t) g(u-t) e^{i\xi(u-t)}\, dt \\ &= e^{-iu\xi} f * g_\xi(u) \end{aligned}$$

It thus follows, using Young's inequality, that

$$\|f_\xi\|_2 = \|f * g_\xi\|_2 \leq \|g\|_1 \|f\|_2$$

The Fourier transform of f_ξ is computed as:

$$\widehat{f_\xi}(\omega) = \widehat{f}(\omega + \xi)\widehat{g_\xi}(\omega + \xi) = \widehat{f}(\omega + \xi)\widehat{g}(\omega)$$

Let us now compute the inner product between $S_g f$ and $S_g h$. Since $f_\xi, h_\xi \in \mathbf{L}^2(\mathbb{R})$ we can use Parseval's formula and our computation for their Fourier transform to get:

$$\begin{aligned} \frac{1}{2\pi} \langle S_g f, S_g h \rangle &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} S_g f(u, \xi) S_g h^*(u, \xi)\, du\, d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f_\xi(u) h_\xi^*(u)\, du \right) d\xi \\ &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \widehat{f}(\omega + \xi) \widehat{h}^*(\omega + \xi) |\widehat{g}(\omega)|^2\, d\omega\, d\xi \end{aligned} \tag{18}$$

We would like to switch the order of integration using Fubini. To do so we need to bound:

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}} |\widehat{f}(\omega+\xi)\widehat{h}^*(\omega+\xi)|\widehat{g}(\omega)|^2\, d\omega\, d\xi \\
&\quad = \int_{\mathbb{R}} |\widehat{g}(\omega)|^2 \int_{\mathbb{R}} |\widehat{f}(\omega+\xi)\widehat{h}^*(\omega+\xi)|\, d\xi\, d\omega \\
&\quad \leq \int_{\mathbb{R}} |\widehat{g}(\omega)|^2 \left(\int_{\mathbb{R}} |\widehat{f}(\omega+\xi)|^2\, d\xi\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} |\widehat{h}(\omega+\xi)|^2\, d\xi\right)^{\frac{1}{2}} d\omega \\
&\quad \leq (2\pi)^2 \|g\|_2^2 \|f\|_2 \|h\|_2 < \infty
\end{aligned}$$

Thus we can apply Fubini and continuing from (18) we have:

$$\begin{aligned}
(18) &= \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(\omega)|^2 \left(\frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega+\xi)\widehat{h}^*(\omega+\xi)\, d\xi\right) d\omega \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(\omega)|^2 \left(\frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\xi)\widehat{h}^*(\xi)\, d\xi\right) d\omega \\
&= \langle f, h\rangle \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(\omega)|^2\, d\omega \\
&= \langle f, g\rangle
\end{aligned}$$

□

The windowed Fourier transform can be extended to any real, symmetric window $g \in \mathbf{L}^2(\mathbb{R})$ using a density argument. Using this extension, we can also extend Theorem 4.1 to any real symmetric window $g \in \mathbf{L}^2(\mathbb{R})$.

Corollary 4.2. *Let $f, h \in \mathbf{L}^2(\mathbb{R})$ and let g be a real symmetric function with $g \in \mathbf{L}^2(\mathbb{R})$ and $\|g\|_2 = 1$. Then:*

$$\langle f, h\rangle = \frac{1}{2\pi}\langle S_g f, S_g h\rangle_{\mathbf{L}^2(\mathbb{R}^2)}$$

It follows from Theorem 4.1 that $S_g : \mathbf{L}^2(\mathbb{R}) \to \mathbf{L}^2(\mathbb{R}^2)$ and that it preserves the norm, up to a factor of $\sqrt{2\pi}$. This is the analog of the Plancheral formula; we collect it in the next corollary.

Corollary 4.3. *Let $g \in \mathbf{L}^2(\mathbb{R})$. The windowed Fourier transform is a linear map $S_g : \mathbf{L}^2(\mathbb{R}) \to \mathbf{L}^2(\mathbb{R}^2)$, and it is also an isometry up to a factor of $\sqrt{2\pi}$:*

$$\|f\|_2 = \frac{1}{\sqrt{2\pi}}\|S_g f\|_{\mathbf{L}^2(\mathbb{R}^2)}$$

Exercise 26. Prove Corollary 4.2.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 08: Inverse Windowed Fourier and Audio Models

February 4, 2020

Lecturer: Matthew Hirn

4.2.2 Inversion for Windowed Fourier

The windowed Fourier transform is highly redundant and covers the entire time frequency plane. Intuitively, we should have more than enough information to invert this transform. Indeed, that is the case:

Theorem 4.4. *Let g be a real symmetric window with $g \in \mathbf{L}^2(\mathbb{R})$ and $\|g\|_2 = 1$. Then for all $f \in \mathbf{L}^2(\mathbb{R})$,*

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} S_g f(u,\xi) g(t-u) e^{i\xi t} \, du \, d\xi$$

We are going to prove the inversion theorem using techniques from functional analysis. We collect the main points first. Let $\mathcal{H}$ be a Hilbert space with norm $\|\cdot\|$, and let $\ell : \mathcal{H} \to \mathbb{C}$ be a linear functional. We say that ℓ is continuous if for $v, h \in \mathcal{H}$ we have

$$\lim_{\|h\| \to 0} \|\ell(v+h) - \ell(v)\| = 0$$

The linear functional ℓ is bounded if there exists a universal constant $C \geq 0$ such that

$$|\ell(v)| \leq C\|v\|, \quad \forall \, v \in \mathcal{H}$$

It is a well known fact that linear functionals are continuous if and only if they are bounded.

Now let $\ell : \mathcal{H} \to \mathbb{C}$ be a continuous linear functional. The *Riesz Representation Theorem* states that for each such ℓ, there exists a unique $h \in \mathcal{H}$ such that

$$\ell(v) = \langle v, h \rangle, \quad \forall \, v \in \mathcal{H}$$

Since it is clear that the mappings $v \mapsto \langle v, h \rangle$ are continuous linear functionals for very $h \in \mathcal{H}$, the Riesz representation theorem shows that there is a bijective correspondence between $\mathcal{H}$ and continuous linear functionals on $\mathcal{H}$.

Finally, suppose now that $F : \mathbb{R} \to \mathcal{H}$, so that for each $u \in \mathbb{R}$, $F(u)$ is an element of the Hilbert space $\mathcal{H}$. One can think of F as a "vector valued function." For example, if $\mathcal{H} = \mathbf{L}^2(\mathbb{R})$ then $F(u)(t)$ is a square integrable function in the variable $t \in \mathbb{R}$ for each $u \in \mathbb{R}$. Using F, one can define a linear functional

$$\ell_F(v) = \int_{\mathbb{R}} \langle v, F(u) \rangle \, du$$

If ℓ_F is bounded / continuous, then by the Riesz representation theorem there exists a unique element $\widetilde{f} \in \mathcal{H}$ such that $\ell_F(v) = \langle v, \widetilde{f} \rangle$. Thus

$$\langle v, \widetilde{f} \rangle = \int_{\mathbb{R}} \langle v, F(u) \rangle \, du, \quad \forall\, v \in \mathcal{H} \tag{19}$$

We write

$$\widetilde{f}(t) = \int_{\mathbb{R}} F(u)(t) \, du$$

which means that (19) holds; this is a type of weak equality. We are going to prove Theorem 4.4 in this sense.

Proof of Theorem 4.4. Define a linear functional $\ell : \mathbf{L}^2(\mathbb{R}) \to \mathbb{C}$ as

$$\ell(h) = \frac{1}{2\pi} \langle S_g h, S_g f \rangle_{\mathbf{L}^2(\mathbb{R}^2)}, \quad \forall\, h \in \mathbf{L}^2(\mathbb{R})$$

By the Parseval theorem for windowed Fourier transforms (Theorem 4.1), we have:

$$\ell(h) = \langle h, f \rangle \leq \|f\|_2 \|h\|_2$$

Thus ℓ is a bounded, and hence a continuous, linear functional. At this point we could apply the Riesz representation theorem, but it would just tell us what we already know which is that $\ell(h) = \langle h, f \rangle$. Instead we come up with a "vector valued function" $F(u, \xi) \in \mathbf{L}^2(\mathbb{R})$ and show that

$$\ell(h) = \int_{\mathbb{R}} \int_{\mathbb{R}} \langle h, F(u, \xi) \rangle \, du \, d\xi \tag{20}$$

To do so, recall that we defined

$$g_{u,\xi}(t) = g(t - u) e^{i\xi t}$$

and that we can write

$$Sf(u, \xi) = \langle f, g_{u,\xi} \rangle$$

We have:

$$\begin{aligned} \ell(h) = \frac{1}{2\pi} \langle S_g h, S_g f \rangle_{\mathbf{L}^2(\mathbb{R}^2)} &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} S_g h(u, \xi) S_g f^*(u, \xi) \, du \, d\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \langle h, g_{u,\xi} \rangle S_g f^*(u, \xi) \, du \, d\xi \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \langle h, (2\pi)^{-1} S_g f(u, \xi) g_{u,\xi} \rangle \, du \, d\xi \end{aligned}$$

Therefore we have verified (20) with

$$F(u, \xi)(t) = \frac{1}{2\pi} S_g f(u, \xi) g_{u,\xi}(t)$$

It follows that $\ell(h)$ can be written as

$$\ell(h) = \langle h, \widetilde{f} \rangle$$

with

$$\widetilde{f}(t) = \int_{\mathbb{R}} F(u,\xi)\, du\, d\xi = \frac{1}{2\pi} \int_{\mathbb{R}} S_g f(u,\xi) g_{u,\xi}(t)\, du\, d\xi$$

But then $\ell(h) = \langle h, f \rangle = \langle h, \widetilde{f} \rangle$ for all $h \in \mathbf{L}^2(\mathbb{R})$, and so $f = \widetilde{f}$ (in the weak sense) and the inversion formula is proved. □

Exercise 27. Read Section 4.2.1 of *A Wavelet Tour of Signal Processing.*

4.2.3 Choice of the Window

The time frequency localization of the window g can be modified with a scaling. Suppose that the Heisenberg boxes of the time frequency atoms $g_{u,\xi}$ have time width σ_t and frequency width σ_ω. Let

$$g_s(t) = s^{-1/2} g(s^{-1}t)$$

be a dilation of g by the time scale s. One can show that if we replace the window g with g_s, then the resulting Heisenberg box has time width $s\sigma_t$ and frequency width $s^{-1}\sigma_\omega$. While the area remains $\sigma_t\sigma_\omega$, the resolution in time is modified by s while the resolution in frequency is modified by s^{-1}. Depending on the signal type we may want better localization in time or frequency, or a balance of both; the parameter s allows us to adjust accordingly while keeping the time frequency area of each box constant.

In numerical applications, the localized waveforms $g_{u,\xi}(t)$ can only be sampled a finite number of times, which means the support of the window g must be compact or it must be restricted to a compact set (as in the case of a Gaussian window). If g has compact support, then $\widehat{g}$ must have an infinite support. Since g is symmetric and often $g(t) \geq 0$ for all t, $\widehat{g}(\omega)$ will be symmetric with a main "lobe" (bump) centered at $\omega = 0$, which decays to zero with oscillations; see Figure 10.

The frequency resolution of the windowed Fourier transform is determined by the spread of $\widehat{g}$ around $\omega = 0$. Previously we used σ_ω to measure this spread, however the following three parameters give a more fine grained measure:

- The bandwidth $\Delta\omega$, which is defined by:

$$\frac{|\widehat{g}(\Delta\omega/2)|^2}{|\widehat{g}(0)|^2} = \frac{1}{2}$$

  This measures the energy concentration of $\widehat{g}_{u,\xi}(\omega)$ around $\omega = \xi$.

- The maximum amplitude A of the first side lobes located at $\omega = \pm\omega_0$. The important thing is the side lobe amplitude relative to the amplitude of the central lobe at $\omega = 0$; this ratio can be measured in decibels:

$$A = 10 \log_{10} \frac{|\widehat{g}(\omega_0)|^2}{|\widehat{g}(0)|^2}$$

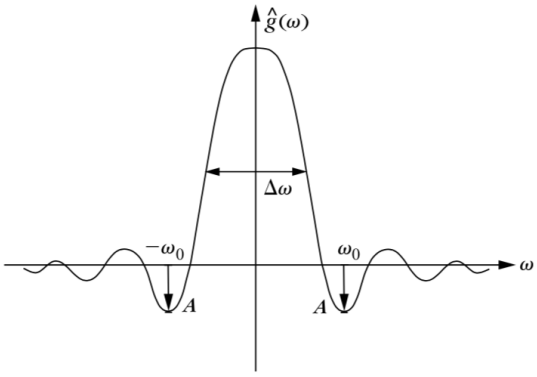


Figure 10: The energy spread of $\widehat{g}(\omega)$ is measured by its bandwidth and the maximum amplitude A of the first side lobes, located at $\pm\omega_0$.

Side lobes create echos of the response $Sf(u,\xi)$ at $Sf(u,\xi \pm \omega_0)$. If A is small (i.e., very negative), then the side lobe magnitude is small relative the main lobe amplitude and these echos will be negligible relative to the response at ξ.

- The polynomial exponent p, which gives the asymptotic decay of $|\widehat{g}(\omega)|$ for large frequencies,

$$|\widehat{g}(\omega)| = O(|\omega|^{-(p+1)})$$

This is important of several localized frequency phenomena occur close together in the time frequency plane. In this case it can be hard to "unmix" the various frequency tones unless p is large. We obtain a large p by using a smooth window.

Exercise 28. Read Sections 4.2.2 and 4.2.3 of *A Wavelet Tour of Signal Processing.*

4.3 Time Frequency Geometry of Instantaneous Frequencies

When listening to music we perceive several frequencies that change with time. This leads to the notion of an *instantaneous frequency*, which we define here at the outset.

4.3.1 Instantaneous Frequency

Section 4.4.1 of A Wavelet Tour of Signal Processing

If $f : \mathbb{R} \to \mathbb{C}$ is complex valued then $f(t)$ can be uniquely represented as

$$f(t) = a(t)e^{i\theta(t)}$$

where $a(t) = |f(t)|$ is the amplitude of $f(t)$ and $\theta(t) \in [0, 2\pi)$ is the phase of $f(t)$. In this case, we define the *instantaneous frequency* of $f(t)$ as $\theta'(t)$.

For real valued signals $f : \mathbb{R} \to \mathbb{R}$, we would like to decompose $f(t)$ as

$$f(t) = \alpha(t) \cos \vartheta(t)$$

However, this representation is not unique since it has two parameters $\alpha(t)$ and $\vartheta(t)$ for each real value $f(t)$. We settle on a particular representation by defining the analytic part of $f(t)$.

A function $h_a(t)$ is *analytic* if

$$\widehat{h}_a(\omega) = 0, \quad \forall\, \omega < 0$$

An analytic function is necessarily complex valued but is entirely characterized by its real part. Indeed, define $h(t) = \Re[h_a(t)]$ to be the real part of $h_a(t)$. Its Fourier transform is:

$$\widehat{h}(\omega) = \frac{\widehat{h}_a(\omega) + \widehat{h}_a^*(-\omega)}{2}$$

which in turn yields:

$$\widehat{h}_a(\omega) = \begin{cases} 2\widehat{h}(\omega) & \omega > 0 \\ \widehat{h}(\omega) & \omega = 0 \\ 0 & \omega < 0 \end{cases}$$

If we start with a real valued signal $f \in \mathbf{L}^1(\mathbb{R})$ then we define the analytic part $f_a(t)$ of $f(t)$ as the inverse Fourier transform of

$$\widehat{f}_a(\omega) = \begin{cases} 2\widehat{f}(\omega) & \omega > 0 \\ \widehat{f}(\omega) & \omega = 0 \\ 0 & \omega < 0 \end{cases}$$

Since the analytic part $f_a(t)$ of $f(t)$ is complex valued, it can be decomposed uniquely as

$$f_a(t) = a(t) e^{i\theta(t)}$$

Since $f(t) = \Re[f_a(t)]$ we have that

$$f(t) = a(t) \cos \theta(t)$$

This representation is uniquely defined because it is derived from the analytic part of f. We call $a(t)$ the analytic amplitude of $f(t)$ and $\theta'(t)$ its instantaneous frequency.

As a somewhat simple example we compute the analytic part of the real valued signal

$$f(t) = a \cos(\omega_0 t + \theta_0) = \frac{a}{2} \left(e^{i(\omega_0 t + \theta_0)} + e^{-i(\omega_0 t + \theta_0)} \right)$$

Its Fourier transform is:

$$\widehat{f}(\omega) = \pi a \left(e^{i\theta_0} \delta(\omega - \omega_0) + e^{-i\theta_0} \delta(\omega + \omega_0) \right)$$

If $\omega_0 > 0$, then the Fourier transform of the analytic part is:

$$\widehat{f_a}(\omega) = 2\widehat{f}(\omega) = 2\pi a e^{i\theta_0}\delta(\omega - \omega_0), \quad \omega \geq 0$$

and thus

$$f_a(t) = ae^{i(\omega_0 t + \theta_0)}$$

If we replace the constant a with an amplitude function $a(t)$, so that

$$f(t) = a(t)\cos(\omega_0 t + \theta_0)$$

then the Fourier transform of $f(t)$ is:

$$\widehat{f}(\omega) = \frac{1}{2}\left(e^{i\theta_0}\widehat{a}(\omega - \omega_0) + e^{-i\theta_0}\widehat{a}(\omega + \omega_0)\right)$$

If the variations of $a(t)$ are slow compared to the period $2\pi/\omega_0$, then it must be that $\operatorname{supp}\widehat{a} \subseteq [-\omega_0, \omega_0]$. In this case:

$$\widehat{f_a}(\omega) = 2\widehat{f}(\omega) = e^{i\theta_0}\widehat{a}(\omega - \omega_0), \quad \omega \geq 0$$

and

$$f_a(t) = a(t)e^{i(\omega_0 t + \theta_0)}$$

Thus the amplitude is $a(t)$ and the instantaneous frequency is ω_0.

Let us now consider a slightly more complicated example:

$$f(t) = a\cos(\omega_1 t) + a\cos(\omega_2 t)$$

In this case the analytic part of the signal is given by:

$$\begin{aligned} f_a(t) &= ae^{i\omega_1 t} + ae^{i\omega_2 t} \\ &= 2a\cos\left(\frac{(\omega_1 - \omega_2)t}{2}\right)e^{i(\omega_1+\omega_2)t/2} \end{aligned}$$

Thus the instantaneous frequency is

$$\theta'(t) = \frac{\omega_1 + \omega_2}{2}$$

and the amplitude is

$$a(t) = 2a\left|\cos\left(\frac{(\omega_1 - \omega_2)t}{2}\right)\right|$$

The result is unsatisfying because the instantaneous frequency is the average of the frequencies of the two cosine waves. We would have no indication (forgetting the amplitude) that the signal is not in fact one cosine with frequency $(\omega_1 + \omega_2)/2$, but rather two separate cosines.

More generally, one would like to be able to analyze signals of the form

$$f(t) = \sum_{k=1}^{K} a_k(t) \cos \theta_k(t) \tag{21}$$

where $a_k(t)$ and $\theta_k(t)$ vary slowly in time. Such decompositions can be used to model music and other auditory signals. We want to isolate the different amplitudes $a_k(t)$ and instantaneous frequencies $\theta_k'(t)$. A windowed Fourier transform can help with this.

Exercise 29. Read Section 4.4.1 of *A Wavelet Tour of Signal Processing.*

Exercise 30. The analytic part $x_a[n]$ of a real valued discrete signal $x \in \mathbb{R}^N$ is defined by

$$\widehat{x}_a[n] = \begin{cases} \widehat{x}[k] & k = 0, N/2 \\ 2\widehat{x}[k] & 0 < k < N/2 \\ 0 & N/2 < k < N \end{cases}$$

(a) Suppose that $y \in \mathbb{C}^N$ is a complex valued discrete signal and let $y_r[n] = \Re(y[n])$ be the real part of y. Prove that

$$\widehat{y}_r[k] = \frac{\widehat{y}[k] + \widehat{y}^*[-k]}{2}$$

(b) For $x \in \mathbb{R}^N$ prove that $\Re(x_a) = x$.

Exercise 31. Let $f(t) = e^{i\theta(t)}$ and let g be a real, symmetric window function with $\|g\|_2 = 1$.

(a) Prove that

$$\int_{\mathbb{R}} |Sf(u,\xi)|^2 \, d\xi = 2\pi, \quad \forall\, u \in \mathbb{R}$$

(b) Prove that

$$\int_{\mathbb{R}} \xi |Sf(u,\xi)|^2 \, d\xi = 2\pi \int_{\mathbb{R}} \theta'(t) |g(t-u)|^2 \, dt, \quad \forall\, u \in \mathbb{R}$$

and interpret this result.

Exercise 32. We are going to investigate further the windowed Fourier transform with a Gaussian window.

(a) Let $g_\sigma(t)$ be the Gaussian window:

$$g_\sigma(t) = \frac{1}{(2\pi\sigma^2)^{1/4}} e^{-t^2/4\sigma^2}$$

which is normalized so that $\|g\|_2 = 1$ and the time spread of $g(t)$ is $\sigma_t^2 = \sigma^2$. In practice, even though $g(t)$ has infinite support, we will have to sample it over a finite interval $[-N/2, N/2)$ of length N. Let $g_{\sigma,N}(t)$ be the restriction of $g_\sigma(t)$ to this interval:

$$g_{\sigma,N}(t) = \mathbf{1}_{[-N/2,N/2)}(t) g_\sigma(t)$$

Give an upper bound for the error $\|g_\sigma - g_{\sigma,N}\|_2$ in terms of σ and N. Recall our intuition that the "essential" width of $g_\sigma(t)$ is σ. If we take $N = \sigma$, how big is the bound on the error? Is this error acceptable?

(b) Implement a discrete version of the windowed Fourier transform. Assume that your signal is $x \in \mathbb{R}^N$ with N even. For each $0 \le k < N$, compute the discrete vectors:

$$g_{\sigma,k}[n] = \begin{cases} g_\sigma(n) \exp\left(\frac{2\pi i k n}{N}\right) & 0 \le n < N/2 \\ g_\sigma(n-N) \exp\left(\frac{2\pi i k n}{N}\right) & N/2 \le n < N \end{cases}$$

Define $S_\sigma x[n,k]$ as:

$$S_\sigma x[n,k] = \exp\left(-\frac{2\pi i k n}{N}\right) \cdot (x \circledast g_{\sigma,k})[n], \quad 0 \le n < N, \;\; 0 \le k < N$$

(You should convince yourself this definition is consistent with the definition above for signals f). Use your work on the earlier exercises to get a fast implementation with $O(N^2 \log N)$ run time. Then analyze the signal $f(t)$ defined as:

$$f(t) = \cos\left(\frac{\pi}{N} t^2\right)$$

with a sampling

$$x[n] = \begin{cases} f(n) & 0 \le n < N/2 \\ f(n-N) & N/2 \le n < N \end{cases}$$

Compute the power spectrum of x, which is $|\widehat{x}[k]|^2$, and the spectrogram $P_S x[n,k] = |S_\sigma x[n,k]|^2$. Plot them both and give an interpretation of your results.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 09: Windowed Fourier Ridges

February 6, 2020

Lecturer: Matthew Hirn

4.3.2 Windowed Fourier Ridges

Section 4.4.2 of A Wavelet Tour of Signal Processing.

We are going to use the windowed Fourier transform, and in particular the local maxima of the windowed Fourier transform, to isolate individual amplitudes $a_k(t)$ and instantaneous frequencies $\theta_k'(t)$ as in the signal model (21), repeated here:

$$f(t) = \sum_{k=1}^{K} a_k(t) \cos \theta_k(t)$$

We make some additional assumptions on the real symmetric window $g(t)$. We suppose that:

- $\operatorname{supp} g = [-1/2, 1/2]$
- $g(t) \geq 0$ so that $|\widehat{g}(\omega)| \leq \widehat{g}(0)$ for all $\omega \in \mathbb{R}$
- $\|g\|_2 = 1$ but also $\widehat{g}(0) = \int g(t)\, dt = \|g\|_1 \approx 1$

For a scale σ set

$$g_\sigma(t) = \sigma^{-1/2} g(\sigma^{-1} t)$$

Note that

$$\operatorname{supp} g_\sigma = [-\sigma/2, \sigma/2] \quad \text{and} \quad \|g_\sigma\|_2 = 1$$

We define the windowed Fourier transform with the scale parameter σ as:

$$S_\sigma f(u, \xi) = \int_{\mathbb{R}} f(t) g_\sigma(t - u) e^{-i\xi t}\, dt$$

The next theorem relates $S_\sigma f(u, \xi)$ to the instantaneous frequency of $f(t)$.

Theorem 4.5. *Let $f(t) = a(t) \cos \theta(t)$. If $\xi \geq 0$, then*

$$S_\sigma f(u, \xi) = \frac{\sqrt{\sigma}}{2} a(u) e^{i[\theta(u) - \xi u]} \Big(\widehat{g}(\sigma[\xi - \theta'(u)]) + \varepsilon(u, \xi) \Big)$$

where

$$|\varepsilon(u,\xi)| \leq \varepsilon_{a,1}(u,\xi) + \varepsilon_{a,2}(u,\xi) + \varepsilon_{\theta,2}(u,\xi) + \sup_{|\omega|\geq\sigma\theta'(u)} |\widehat{g}(\omega)|$$

with

$$\varepsilon_{a,1}(u,\xi) \leq \frac{\sigma|a'(u)|}{|a(u)|}$$

and

$$\varepsilon_{a,2}(u,\xi) \leq \sup_{|t-u|\leq\sigma/2} \frac{\sigma^2|a''(t)|}{|a(u)|}$$

Furthermore, if $\sigma|a'(u)||a(u)|^{-1} \leq 1$*, then*

$$\varepsilon_{\theta,2}(u,\xi) \leq \sup_{|t-u|\leq\sigma/2} \sigma^2|\theta''(t)|$$

And finally, if $\xi = \theta'(u)$*, then*

$$\varepsilon_{a,1}(u,\xi) = \frac{\sigma|a'(u)|}{|a(u)|}|\widehat{g}'(2\sigma\theta'(u))|$$

We omit the proof, which is given in pages 119–122 of *A Wavelet Tour of Signal Processing*. If we can neglect the error term $\varepsilon(u,\xi)$, then we will see that $S_\sigma f(u,\xi)$ enables us to measure $a(u)$ and $\theta'(u)$. This will be the case if $a(t)$ and $\theta(t)$ vary slowly. In particular, $\varepsilon_{a,1}$ is small if $a(t)$ varies slowly over the whole real line, while $\varepsilon_{a,2}$ and $\varepsilon_{\theta,2}$ only require the second derivatives of $a(t)$ and $\theta(t)$ to be small over an interval of length equal to the support of the window g. The fourth part of the error term is small if

$$\Delta\omega \leq \sigma\theta'(u) \tag{22}$$

where recall $\Delta\omega$ is the bandwidth of g.

Let us now suppose that the error term can disregarded, so that

$$S_\sigma f(u,\xi) \approx \frac{\sqrt{\sigma}}{2} a(u) e^{i[\theta(u)-\xi u]} \widehat{g}(\sigma[\xi - \theta'(u)])$$

Since the maximum of $|\widehat{g}(\omega)|$ is at $\omega = 0$, we see that for each u the spectrogram $P_S f(u,\xi) = |S_\sigma f(u,\xi)|^2$ is maximum at $\xi_u = \theta'(u)$. These time frequency points (u,ξ_u), which form curves in the time frequency plane, are called *ridges*. At ridge points we have:

$$S_\sigma f(u,\xi_u) = S_\sigma f(u,\theta'(u)) = \frac{\sqrt{\sigma}}{2} a(u) e^{i[\theta(u)-u\theta'(u)]} (\widehat{g}(0) + \varepsilon(u,\theta'(u)))$$

If the bandwidth satisfies (22), then Theorem 4.5 shows that the $\varepsilon_{a,1}(u,\xi)$ error term is negligible, since in this case $|\widehat{g}'(2\sigma\theta'(u))|$ will be negligible.

We can calculate the amplitude from the ridges as well:

$$a(u) \approx \frac{2|S_\sigma f(u,\theta'(u))|}{\sqrt{\sigma}|\widehat{g}(0)|}$$

if the error term $\varepsilon(u, \theta'(u))$ is small.

The spectrogram computes the instantaneous frequency by computing the magnitude of $S_\sigma(u, \xi)$ along the ridges. Another way to calculate the instantaneous frequency is to look at the phase of $S_\sigma f(u, \xi)$ along the ridges. Let $\Theta_S f(u, \xi)$ be the complex phase of $S_\sigma f(u, \xi)$, which again if the error term can be disregarded, is just:

$$\Theta_S f(u, \xi) \approx \theta(u) - \xi u$$

It follows that

$$\frac{\partial \Theta_S f(u, \xi)}{\partial u} = \theta'(u) - \xi$$

and thus the instantaneous frequency can be computed by estimating this partial derivative and solving for its zeros.

Consider now a signal model

$$f(t) = \sum_{k=1}^{K} a_k(t) \cos \theta_k(t)$$

where $a_k(t)$ and $\theta'_k(t)$ have small variations over intervals of size σ and $\sigma\theta'_k(t) \geq \Delta\omega$ (in other words, we can neglect the error term). Since the windowed Fourier transform is linear, we have:

$$S_\sigma f(u, \xi) \approx \frac{\sqrt{\sigma}}{2} \sum_{k=1}^{K} a_k(u) e^{i[\theta_k(u) - \xi u]} \widehat{g}(\sigma[\xi - \theta'_k(u)])$$

We can distinguish between the K different instantaneous frequencies if

$$\widehat{g}(\sigma[\theta'_k(u) - \theta'_l(u)]) \ll 1, \quad \forall\, u \in \mathbb{R},\ \ k \neq l \tag{23}$$

We can obtain this condition if the bandwidth of g satisfies

$$\Delta\omega \leq \sigma |\theta'_k(u) - \theta'_l(u)|, \quad \forall\, u \in \mathbb{R},\ \ k \neq l$$

In this case, when $\xi = \theta'_l(u)$, we have

$$\begin{aligned} S_\sigma f(u, \theta'_l(u)) &\approx \frac{\sqrt{\sigma}}{2} \left(a_l(u) e^{i[\theta_l(u) - u\theta'_l(u)]} \widehat{g}(0) + \underbrace{\sum_{k \neq l} a_k(u) e^{i[\theta_k(u) - u\theta'_l(u)]} \widehat{g}(\sigma[\theta'_l(u) - \theta'_k(u)])}_{\ll 1} \right) \\ &\approx \frac{\sqrt{\sigma}}{2} a_l(u) e^{i[\theta_l(u) - u\theta'_l(u)]} \widehat{g}(0) \end{aligned}$$

and thus we can estimate the instantaneous frequency $\theta'_l(u)$ and corresponding amplitude $a_l(u)$. Notice that the ridge points are distributed along the K time frequency curves $\{(u, \theta'_k(u)) : u \in \mathbb{R},\ 1 \leq k \leq K\}$. So long as these curves remain well separated (as

measured by (23)), we will be able to recover the instantaneous frequencies. However, if the curves get too close, or even worse intersect, then the windowed Fourier transform will have interference and the ridge pattern will be destroyed in that neighborhood.

We have already seen that along the ridge points the error term $\varepsilon_{a,1}(u,\xi)$ is negligible if the bandwidth $\Delta\omega$ is small enough. But we still need make sure the error terms $\varepsilon_{a,2}(u,\xi)$ and $\varepsilon_{\theta,2}(u,\xi)$ are small, which means from Theorem 4.5 we need:

$$\varepsilon_{a,2}(u,\xi) \leq \max_k \sup_{|t-u|\leq\sigma/2} \frac{\sigma^2|a_k''(t)|}{|a_k(u)|} \ll 1$$

and

$$\varepsilon_{\theta,2}(u,\xi) \leq \max_k \sup_{|t-u|\leq\sigma/2} \sigma^2|\theta_k''(t)| \ll 1$$

These place a condition on σ in which we would like to make σ small. However, recall that to make $\varepsilon_{a,1}(u,\xi)$ small at the ridge points and the fourth part of the error term small, we needed

$$\Delta\omega \leq \sigma\theta_k'(u)$$

which means we would like to make σ large. Since $\operatorname{supp} g_\sigma = [-\sigma/2, \sigma/2]$, this means we need to carefully select the window size. Notice how this leads to a tradeoff between localization in time and localization in frequency.

Let us now consider some examples. A linear chirp is of the form:

$$\widetilde{f}(t) = a\cos(bt^2 + ct)$$

It is called linear because its instantaneous frequency is $\theta'(t) = 2bt + c$. Suppose we have a signal consisting of two linear chirps:

$$f(t) = a_1\cos(bt^2 + ct) + a_2\cos(bt^2)$$

To distinguish these two linear chirps, we need our window g to have bandwidth $\Delta\omega$ satisfying

$$\Delta\omega \leq \sigma|\theta_1'(t) - \theta_2'(t)| = \sigma|c|$$

Since the amplitudes are constant, the error term $\varepsilon_{a,2}$ is zero. However, $\varepsilon_{\theta,2}(u,\xi)$ places an upper bound on the time support, which is:

$$\sigma^2|\theta_k''(u)| = 2b\sigma^2 \ll 1, \quad k = 1, 2$$

Combining the previous two inequalities we get:

$$\frac{\Delta\omega}{c} \leq \sigma \ll \frac{1}{\sqrt{b}} \Longrightarrow \Delta\omega \ll \frac{c}{\sqrt{b}}$$

Figure 11 illustrates an example.

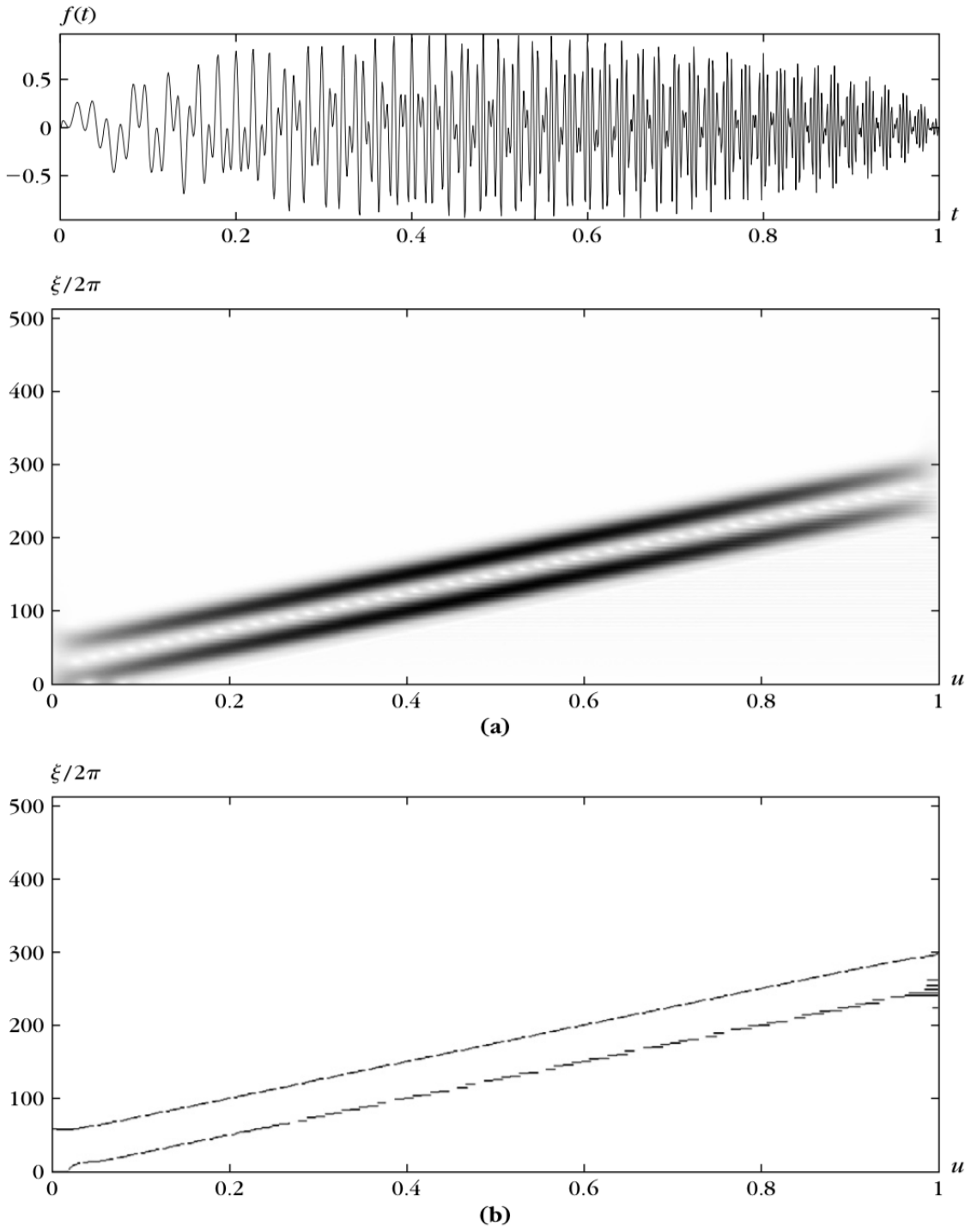


Figure 11: Top: Sum of two parallel linear chirps. Middle: Spectrogram. Bottom: Windowed Fourier ridges.

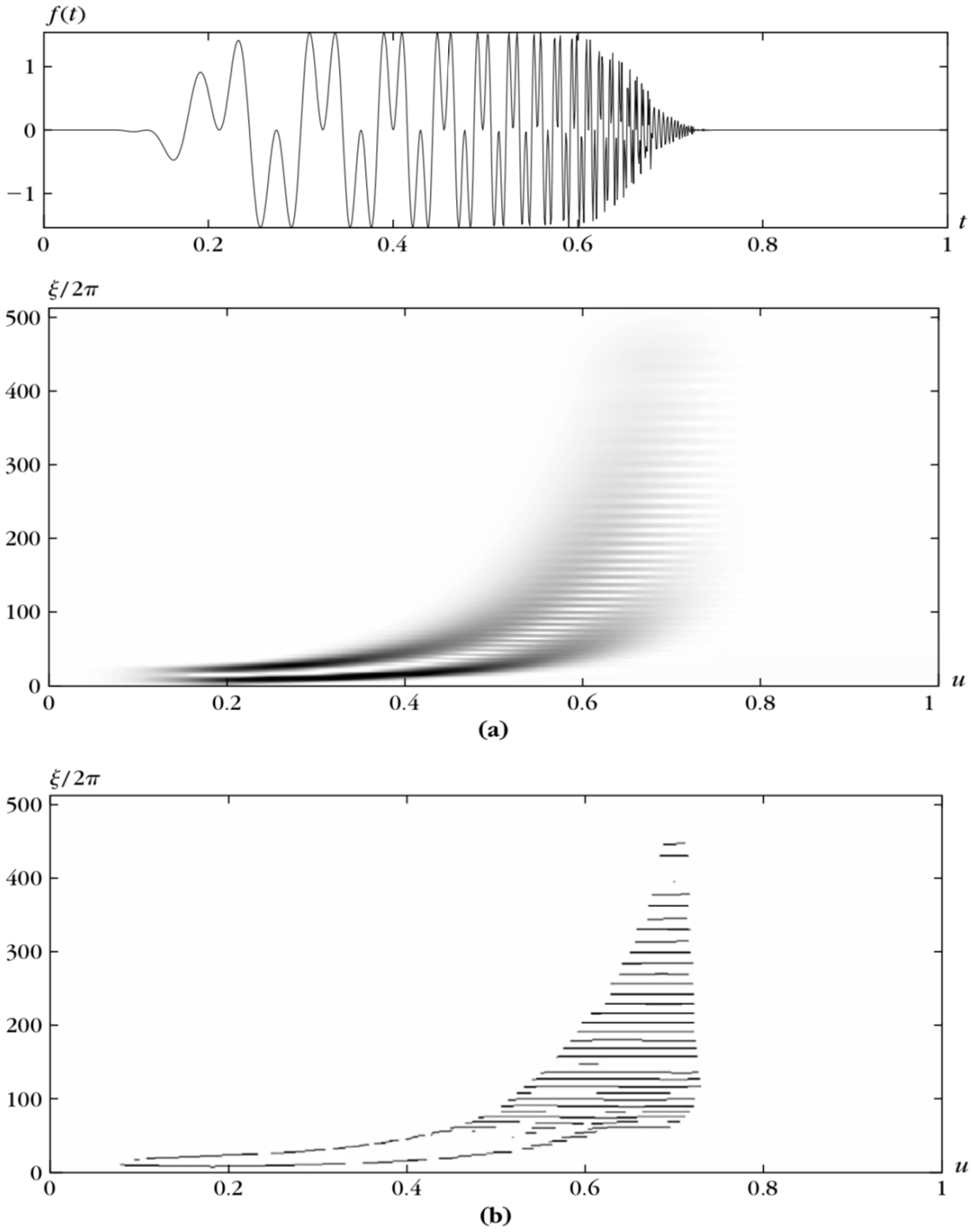


Figure 12: Top: Sum of two hyperbolic chirps. Middle: Spectrogram. Bottom: Windowed Fourier ridges.

Now consider a hyperbolic chirp, which has the form:

$$\widetilde{f}(t) = a\cos\left(\frac{\alpha}{\beta - t}\right)$$

When $t < \beta$, it has instantaneous frequency

$$\theta'(t) = \frac{\alpha}{(\beta - t)^2}$$

which varies quickly as t approaches β. Indeed, as $t \to \beta$ we have $\theta'(t) \to +\infty$, which means that the instantaneous frequency increases to $+\infty$ in a finite amount of time. This is a problem for the windowed Fourier transform because it has a fixed scale in time, and so cannot resolve the fast high frequency changes of the hyperbolic chirp. More precisely, for the error term $\varepsilon_{\theta,2}(u,\xi)$ in Theorem 4.5, we have $\varepsilon_{\theta,2}(u,\xi) \leq \sigma^2|\theta''(u)|$ but for the hyperbolic chirp,

$$\sigma^2|\theta''(u)| = \frac{\sigma^2\alpha}{|\beta - u|^3} > 1, \quad \forall\, |u - \beta| < (\sigma^2\alpha)^{1/3}$$

Therefore the error term is uncontrolled, which leads to a lot of interference in the time frequency response $S_\sigma f(u,\xi)$. Figure 12 illustrates the point for the sum of two hyperbolic chirps,

$$f(t) = a_1\cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2\cos\left(\frac{\alpha_2}{\beta_2 - t}\right)$$

Exercise 33. Read Section 4.4.2 of *A Wavelet Tour of Signal Processing*.

Exercise 34. Now we are going to use your windowed Fourier transform code to reproduce some results from the book.

(a) Read Example 4.5 (p. 94) of *A Wavelet Tour of Signal Processing* and determine what the signal is (write it out analytically). Then compute the windowed Fourier transform and corresponding spectrogram, and recreate something similar to Figure 4.3(a). Provide a plot of your spectrogram.

(b) Consider the signal

$$f(t) = a_1\cos(bt^2 + ct) + a_2\cos(bt^2)$$

which consists of two real valued linear chirps. Compute the windowed Fourier transform and spectrogram of $f(t)$. Can you find a window g and parameters a_1, a_2, b, c such that you can recreate something similar to Figure 4.13(a)? Provide a plot of your spectrogram. Note: Unlike Exercise 32(b) in which you sampled the single linear chirp on $[-N/2, N/2)$, here sample it on $[0, N)$ so the instantaneous frequency is monotonic.

(c) Consider the signal

$$f(t) = a_1\cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2\cos\left(\frac{\alpha_2}{\beta_2 - t}\right)$$

which consists of two hyperbolic chirps. Select parameters $a_1, a_2, \alpha_1, \alpha_2, \beta_1, \beta_2$ and compute the windowed Fourier transform and spectrogram of $f(t)$. Do you get something like Figure 4.14(a)? Provide a plot of your spectrogram.

Exercise 35. We are going to compute numerically windowed Fourier ridges.

(a) Take your windowed Fourier code and add in code to estimate the Fourier ridges by estimating the local maxima of $P_S f(u,\xi) = |S_\sigma f(u,\xi)|^2$.

(b) Now write code to estimate Fourier ridges using the alternate approach, which was to let $\Theta_S f(u,\xi)$ be the complex phase of $S_\sigma f(u,\xi)$, and to solve for ξ such that

$$\frac{\partial \Theta_S f}{\partial u}(u,\xi) = 0$$

(c) Test your code by computing the windowed Fourier ridges of the signals from Exercise 34. Do you get results similar to those from Figures 4.12, 4.13(b), 4.14(b) in *A Wavelet Tour of Signal Processing*? Turn in your plots of the ridges.

Exercise 36. (a) Let $f(t) = \cos(a\cos(bt))$. We want to compute precisely the instantaneous frequency of $f(t)$ from the ridges of its windowed Fourier transform. Find a sufficient condition on the window support size as a function of a and b.

(b) Now let $f(t) = \cos(a\cos(bt)) + \cos(a\cos(bt) + ct)$. In terms of the bandwidth $\Delta\omega$, find a condition on a, b and c in order to measure both instantaneous frequencies with the ridges of a windowed Fourier transform.

(c) Verify your calculations for (a) and (b) numerically using your windowed Fourier ridge code from the previous exercise. Turn in plots of the spectrogram for (a) and (b) and plots of the ridges for (a) and (b).

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 10: The Wavelet Transform

February 11, 2020

Lecturer: Matthew Hirn

4.4 Wavelet Transforms

Section 4.3 of A Wavelet Tour of Signal Processing.

The hyperbolic chirp example illustrates that some types of signals require a transform that can vary its scale to account for multiscale characteristics within the signal. Another example is given in Figure 13, in which we have a signal with multiple types of behavior, some over large scales (like the general increasing and decreasing nature of the signal) and others over smaller scales (like the singular points). As we shall see a bit later, a multiscale approach

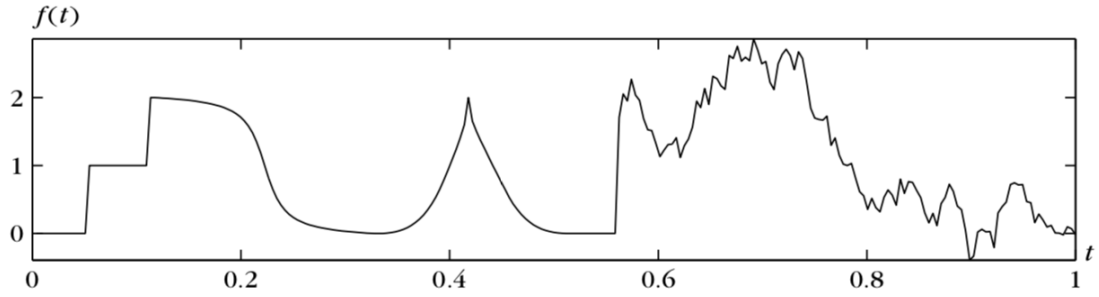


Figure 13: A signal with multiscale structure.

is also necessary for characterizing local singularities in a signal $f(t)$, as it is the ability to "zoom in," which will allow us to achieve such analysis. This is particular important in image processing, since natural images often have many different types of patterns.

A third motivation comes from Remark 3.5. There we defined the operator Δ_j which on the real line would map $f \in \mathbf{L}^2(\mathbb{R})$ to:

$$(\Delta_j f)(t) = \int_{\mathbb{R}} \widehat{f}(\omega) \underbrace{\mathbf{1}_{2^{-j} \leq |\omega| < 2^{-j+1}}(\omega)}_{\widehat{h}_j(\omega)} e^{it\omega} \, d\omega \, .$$

Since the frequency supports of $\widehat{h}_j(\omega)$ decrease as j increases, it must be that the time support of $h_j(t)$ increases as j increases.

All of these examples motivate the introduction of a new multiscale time frequency transform, which will be the wavelet transform. A wavelet $\psi \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ is a function with

zero average,

$$\int_{\mathbb{R}} \psi(t)\, dt = 0$$

which is well localized in time and frequency. We normalize ψ so that $\|\psi\|_2 = 1$ and such that it is centered at $t = 0$.

A dictionary of time frequency atoms is obtained by dilating ψ by s and translating it by u:

$$\mathcal{D} = \{\psi_{u,s}\}_{u\in\mathbb{R},\, s>0}, \quad \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$$

Note that $\|\psi_{u,s}\|_2 = 1$ for all (u, s). The wavelet transform of $f \in \mathbf{L}^2(\mathbb{R})$ computes:

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \int_{\mathbb{R}} f(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right)\, dt$$

The wavelet dictionary is a translation invariant dictionary, and hence can be written as a family of convolutions. Set

$$\overline{\psi}_s(t) = \frac{1}{\sqrt{s}}\psi^*\left(\frac{-t}{s}\right)$$

and observe that

$$Wf(u,s) = f * \overline{\psi}_s(u)$$

If we set $f_s(u) = Wf(u, s)$, then:

$$\widehat{f}_s(\omega) = \widehat{f}(\omega)\widehat{\overline{\psi}}_s(\omega), \quad \widehat{\overline{\psi}}_s(\omega) = \sqrt{s}\widehat{\psi}^*(s\omega)$$

Since ψ has zero average, the support $\widehat{\psi}(\omega)$ must be away from $\omega = 0$. It follows that the wavelet transform computes a *bandpass* filtering of f with a family dilated bandpass filters $\overline{\psi}_s$.

Wavelets can be either real valued or complex valued. Often in the latter case they are taken to be complex analytic or nearly complex analytic. Such wavelet transforms are good for analyzing instantaneous frequencies, which we studied previously with the windowed Fourier transform. Real valued wavelets, on the other hand, are good for detecting sharp transitions in a signal, such as singular points. In this case the wavelets are defined as multiscale derivative operators. We will start with complex analytic wavelets since they will parallel the windowed Fourier transform to a certain degree, and then study real valued wavelets.

Exercise 37. Read the first part of Section 4.3 of *A Wavelet Tour of Signal Processing*, up to (but not including) Section 4.3.1.

4.4.1 Analytic Wavelet Transform

Section 4.3.2 of A Wavelet Tour of Signal Processing

An analytic wavelet is a complex valued wavelet $\psi \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ such that

$$\widehat{\psi}(\omega) = 0, \quad \omega \leq 0$$

We can measure the Heisenberg boxes of the wavelet time frequency atom $\psi_{u,s}$. Suppose that ψ is centered at $t = 0$ so that its time variance is:

$$\sigma_t^2 = \int_{\mathbb{R}} t^2 |\psi(t)|^2 \, dt$$

Since $\psi(t)$ is centered at zero, $\psi_{u,s}(t)$ is centered at $t = u$ and its time variance is computed as:

$$\begin{aligned}
\sigma_t^2(u,s) &= \int_{\mathbb{R}} (t-u)^2 |\psi_{u,s}(t)|^2 \, dt \\
&= \int_{\mathbb{R}} (t-u)^2 \frac{1}{s} \left| \psi\left(\frac{t-u}{s}\right) \right|^2 \, dt, \quad v = \frac{t-u}{s} \\
&= s^2 \int_{\mathbb{R}} v^2 |\psi(v)|^2 \, dv \\
&= s^2 \sigma_t^2
\end{aligned}$$

Since ψ is analytic, the center frequency η of ψ is

$$\eta = \frac{1}{2\pi} \int_0^{+\infty} \omega |\widehat{\psi}(\omega)|^2 \, d\omega$$

The Fourier transform of $\psi_{u,s}$ is

$$\sqrt{s}\widehat{\psi}(s\omega) e^{-iu\omega}$$

and thus its center frequency is

$$\xi_{u,s} = \eta / s$$

The frequency variance of ψ is

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_0^{+\infty} (\omega - \eta)^2 |\widehat{\psi}(\omega)|^2 \, d\omega$$

and using a similar change of variables as before, we obtain:

$$\sigma_\omega^2(u,s) = \frac{\sigma_\omega^2}{s^2}$$

Thus the time frequency Heisenberg box of $\psi_{u,s}$ is centered at $(u, \eta/s)$ and has length $s\sigma_t$ along the time axis, and length σ_ω / s along the frequency axis. Figure 14 illustrates the

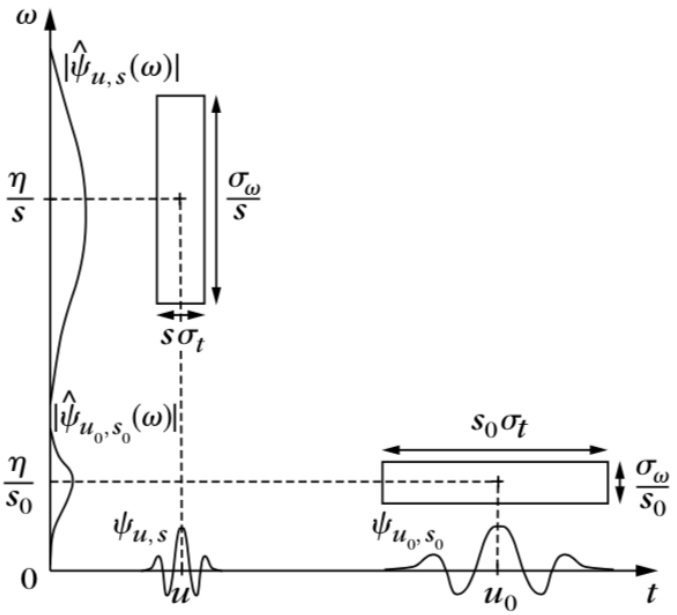


Figure 14: Heisenberg boxes of the analytic wavelet time frequency atoms.

idea. A wavelet transform is thus multiscale in time, as it tests the signal f against localized oscillating waveforms at different scales s. It is additionally, though, multiresolution in frequency, with better frequency localization in the low frequencies, and worse frequency resolution in the high frequencies.

Analogous to the spectrogram for the windowed Fourier transform, an analytic wavelet transform defines a local time frequency energy density $P_W f$, which measures the energy of f in the Heisenberg box $\psi_{u,s}$. This density is called the *scalogram*, and it is defined as:

$$P_W f(u,s) = |Wf(u,s)|^2$$

An analytic wavelet can be constructed in a similar fashion as the windowed Fourier transform. Let g once gain be a real symmetric window, and set

$$\psi(t) = g(t)e^{i\eta t}$$

Recall that Fourier transform of ψ is:

$$\widehat{\psi}(\omega) = \widehat{g}(\omega - \eta)$$

Thus if

$$\widehat{g}(\omega) = 0, \quad |\omega| > \eta$$

then ψ is analytic; see Figure 15. Recall as well that since g is real and symmetric, $\widehat{g}(\omega)$ is real and symmetric as well and thus the center of frequency of ψ is η.

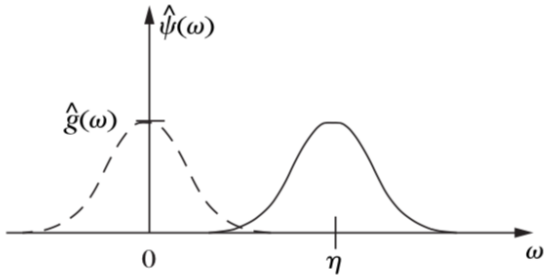


Figure 15: Fourier transform $\widehat{\psi}(\omega)$ of a wavelet $\psi(t) = g(t)e^{i\eta t}$.

A *Gabor wavelet* $\psi(t) = g_\sigma(t)e^{i\eta t}$ is obtained with a Gaussian window

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

where here g_σ is normalized so that $\|\psi\|_1 = 1$. The Fourier transform of the Gaussian window is

$$\widehat{g}_\sigma(\omega) = e^{-\sigma^2\omega^2/2}$$

Thus if $\sigma^2\eta^2 \gg 1$, then $\widehat{g}_\sigma(\omega) \approx 0$ for $|\omega| > \eta$ and the Gabor wavelet has nearly zero average and is nearly analytic. It is thus not a wavelet in the strict sense of the term.

A *Morlet wavelet* modifies the Gabor wavelet to have precisely zero average; it is defined as:

$$\psi(t) = g(t)\left(e^{i\eta t} - C_{\sigma,\eta}\right), \quad C_{\sigma,\eta} = e^{-\sigma^2\eta^2/2}$$

A Morlet wavelet is thus a wavelet since $\int \psi = 0$. It is nearly analytic, but has a small negative response in the negative frequencies. Both 2D Gabor and 2D Morlet wavelets are often used in 2D computer vision and image processing tasks.

The next theorem shows that the analytic wavelet transform is invertible so long as the wavelet ψ satisfies a weak admissibility condition given by (25). Recall that $f_a(t)$ is the analytic part of a real valued signal $f(t)$.

Theorem 4.6. *Let $f \in \mathbf{L}^2(\mathbb{R})$ be real valued and let $\psi \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ be an analytic wavelet. Then*

$$Wf(u,s) = \frac{1}{2} Wf_a(u,s) \tag{24}$$

Furthermore, if

$$C_\psi = \int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|^2}{\omega}\, d\omega < \infty \tag{25}$$

then

$$f(t) = \frac{2}{C_\psi} \Re\left[\int_0^{+\infty} \int_{\mathbb{R}} Wf(u,s)\psi_s(t-u)\, du\, \frac{ds}{s^2}\right] \tag{26}$$

and

$$\|f\|_2^2 = \frac{2}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} |Wf(u,s)|^2 \, du \, \frac{ds}{s^2} \tag{27}$$

Proof. First note that the $\mathbf{L}^1(\mathbb{R})$ norm of ψ_s is:

$$\|\psi_s\|_1 = \sqrt{s}\|\psi\|_1$$

which is obtained through a simple change of variables.

We now prove (24) first. Recall that $\overline{\psi}_s(t) = \psi^*(-t)$ and consider

$$f_s(u) = Wf(u,s) = f * \overline{\psi}_s(u)$$

Young's inequality (17) proves that $f_s \in \mathbf{L}^2(\mathbb{R})$:

$$\|f_s\|_2 = \|f * \overline{\psi}_s\|_2 \leq \|f\|_2 \|\overline{\psi}_s\|_1 = \sqrt{s}\|\psi\|_1 \|f\|_2 < \infty$$

The Fourier transform of f_s is:

$$\widehat{f}_s(\omega) = \widehat{f}(\omega)\widehat{\overline{\psi}}_s(\omega) = \widehat{f}(\omega)\sqrt{s}\widehat{\psi}^*(s\omega)$$

Since $\widehat{\psi}(\omega) = 0$ for $\omega \leq 0$, and $\widehat{f}_a(\omega) = 2\widehat{f}(\omega)$ for $\omega > 0$, we derive that

$$\widehat{f}_s(\omega) = \frac{1}{2}\widehat{f}_a(\omega)\sqrt{s}\widehat{\psi}^*(\omega)$$

which is the Fourier transform of $(1/2) f_a * \overline{\psi}_s(u)$.

Now let us prove (27). We use the Plancheral formula and Tonelli:

$$\begin{aligned}
\frac{2}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} |Wf(u,s)|^2 \, du \, \frac{ds}{s^2} &= \frac{1}{2C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} |f_a * \overline{\psi}_s(u)|^2 \, du \, \frac{ds}{s^2} \\
&= \frac{1}{C_\psi 4\pi} \int_0^{+\infty} \int_{\mathbb{R}} |\widehat{f}_a(\omega)|^2 |\widehat{\psi}(s\omega)|^2 \, d\omega \, \frac{ds}{s} \\
&= \frac{1}{4\pi} \int_0^{+\infty} |\widehat{f}_a(\omega)|^2 \frac{1}{C_\psi} \int_0^{+\infty} |\widehat{\psi}(s\omega)|^2 \frac{ds}{s} \, d\omega
\end{aligned} \tag{28}$$

Now make the change of variables $\xi = s\omega$, which induces the change of measure $ds = (s/\xi)d\xi$. We have:

$$\begin{aligned}
(28) &= \frac{1}{4\pi} \int_0^{+\infty} |\widehat{f}_a(\omega)|^2 \frac{1}{C_\psi} \int_0^{+\infty} \frac{|\widehat{\psi}(\xi)|^2}{\xi} \, d\xi \, d\omega \\
&= \frac{1}{4\pi} \int_0^{+\infty} |\widehat{f}_a(\omega)|^2 \, d\omega \\
&= \frac{1}{2}\|f_a\|_2^2 = \|f\|_2^2
\end{aligned}$$

Finally we prove (26). To do so we prove that

$$f_a(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} W f_a(u,s) \psi_s(t-u)\, du\, \frac{ds}{s^2} \tag{29}$$

The result will then follow since $W f_a(u,s) = 2Wf(u,s)$ and $\Re[f_a(t)] = f(t)$.

To prove (29) we write the right hand side of (29) as:

$$h(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} W f_a(u,s) \psi_s(t-u)\, du\, \frac{ds}{s^2} = \frac{1}{C_\psi} \int_0^{+\infty} f_a * \overline{\psi}_s * \psi_s(t)\, \frac{ds}{s^2}$$

We compute the Fourier transform of $h(t)$, which gives:

$$\begin{aligned}
\widehat{h}(\omega) &= \frac{1}{C_\psi} \int_0^{+\infty} \widehat{f}_a(\omega) \widehat{\overline{\psi}}_s(\omega) \widehat{\psi}_s(\omega)\, \frac{ds}{s^2} \\
&= \frac{\widehat{f}_a(\omega)}{C_\psi} \int_0^{+\infty} \sqrt{s}\widehat{\psi}^*(s\omega) \sqrt{s}\widehat{\psi}(s\omega)\, \frac{ds}{s^2} \\
&= \frac{\widehat{f}_a(\omega)}{C_\psi} \int_0^{+\infty} |\widehat{\psi}(s\omega)|^2 \frac{ds}{s}
\end{aligned}$$

But this is exactly the same integral we encountered in (28), where we proved that it equals C_ψ. Thus we obtain $\widehat{h}(\omega) = \widehat{f}_a(\omega)$. But then we must have $h(t) = f_a(t)$ and the theorem is proved. □

The condition (25) is the *admissibility condition* of the wavelet ψ. Note that to be finite, we must have $\widehat{\psi}(0) = 0$, or equivalently $\int \psi = 0$. This condition is nearly sufficient. If $\widehat{\psi} \in \mathbf{C}^1(\mathbb{R})$ as well, then one can verify that the admissibility condition is satisfied. By Theorem 2.15, if ψ is sufficiently localized in time such that

$$|\psi(t)| \leq \frac{C}{1 + |t|^{n+1+\epsilon}}$$

then we must have $\widehat{\psi} \in \mathbf{C}^1(\mathbb{R})$.

Exercise 38. Read Section 4.3.2 of *A Wavelet Tour of Signal Processing.*

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 11: Wavelet Ridges

February 13, 2020

Lecturer: Matthew Hirn

4.4.2 Wavelet Ridges

Section 4.4.3 of A Wavelet Tour of Signal Processing.

Motivated by the hyperbolic chirp example and the poor performance of the windowed Fourier ridges for this example, we define and study wavelet ridges. We utilize an approximately analytic "wavelet" $\psi(t)$ of the form:

$$\psi(t) = g(t)e^{i\eta t}$$

where the window function $g(t)$ satisfies the same assumptions as in the windowed Fourier case; namely:

- $\operatorname{supp} g = [-1/2, 1/2]$
- $g(t) \geq 0$ so that $|\widehat{g}(\omega)| \leq \widehat{g}(0)$ for all $\omega \in \mathbb{R}$
- $\|g\|_2 = 1$ but also $\widehat{g}(0) = \int g(t)\,dt = \|g\|_1 \approx 1$

Let $\Delta\omega$ be the bandwidth of $\widehat{g}$. If $\eta > \Delta\omega$ then

$$\widehat{\psi}(\omega) = \widehat{g}(\omega - \eta) \ll 1, \quad \forall\, \omega \leq 0$$

Thus $\psi(t)$ is not strictly a wavelet nor is it strictly analytic, but it nearly satisfies both conditions.

Notice that dilated and translated wavelets can be written as:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) = g_{s,u,\eta/s}(t)e^{-i(\eta/s)u}$$

where

$$g_{s,u,\xi}(t) = \frac{1}{\sqrt{s}}g\left(\frac{t-u}{s}\right)e^{i\xi t}$$

The resulting wavelet transform use time frequency atoms similar to those of the windowed Fourier transform. However, in this case the scale s varies over $(0, +\infty)$ and $\xi = \eta/s$:

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \langle f, g_{u,s,\eta/s}\rangle e^{i(\eta/s)u}$$

Theorem 4.5 computes $\langle f, g_{u,s,\xi}\rangle$ when $f(t) = a(t)\cos\theta(t)$. Applying this theorem to the wavelet transform gives:

$$Wf(u,s) = \langle f, g_{u,s,\eta/s}\rangle e^{i(\eta/s)u} = \frac{\sqrt{s}}{2} a(u) e^{i\theta(u)} \left[\widehat{g}(s[\eta/s - \theta'(u)]) + \varepsilon(u,\eta/s)\right]$$

A normalized scalogram computes

$$\widetilde{P}_W f(u,\eta/s) = \frac{|Wf(u,s)|^2}{s} = \frac{1}{4} a(u)^2 \left|\widehat{g}(s[\eta/s - \theta'(u)]) + \varepsilon(u,\eta/s)\right|^2$$

If the error term $\varepsilon(u,\eta/s)$ is negligible, $\widetilde{P}_W f(u,\eta/s)$ obtains its maxima at $(u,\eta/s_u)$ where

$$\frac{\eta}{s_u} = \theta'(u) \Longrightarrow s_u = \frac{\eta}{\theta'(u)}$$

The corresponding points $(u,\eta/s_u)$ are called *wavelet ridges*.

Recall the error term $\varepsilon(u,\eta/s)$ is broken into four components:

$$|\varepsilon(u,\eta/s)| \le \varepsilon_{a,1}(u,\eta/s) + \varepsilon_{a,2}(u,\eta/s) + \varepsilon_{\theta,2}(u,\eta/s) + \underbrace{\sup_{\omega \ge s\theta'(u)} |\widehat{g}(\omega)|}_{\text{(iv)}}$$

At the ridge points $(u,\eta/s_u)$ the first error term $\varepsilon_{a,1}$ and the fourth error term can be made negligible if the bandwidth $\Delta\omega$ satisfies

$$\Delta\omega \le s_u\theta'(u) \Longrightarrow \Delta\omega \le \eta$$

but this was assumed from the start in order to make ψ an approximately analytic wavelet, so these two error terms are guaranteed to be small by the choice of the wavelet. Using Theorem 4.5, the second order terms at the ridge points are bounded as:

$$\varepsilon_{a,2}(u,\eta/s_u) \le \sup_{|t-u|\le s_u/2} \frac{s_u^2|a''(t)|}{|a(u)|} = \sup_{|t-u|\le \eta/2\theta'(u)} \frac{\eta^2}{\theta'(u)^2}\frac{|a''(t)|}{|a(u)|}$$

and

$$\varepsilon_{\theta,2}(u,\eta/s_u) \le \sup_{|t-u|\le s_u/2} s_u^2|\theta''(t)| = \sup_{|t-u|\le \eta/2\theta'(u)} \frac{\eta^2}{\theta'(u)^2}|\theta''(t)|$$

Thus since $\theta'(u)$ is in the denominator, we see that if the instantaneous frequency is small, $a'(u)$ and $\theta'(u)$ must have slow variations (i.e., $a''(u)$ and $\theta''(u)$ need to be small), but $a'(u)$ and $\theta'(u)$ are allowed to vary much more quickly when the instantaneous frequency is large.

Now turn to our more general signal model:

$$f(t) = \sum_{k=1}^{K} a_k(t)\cos\theta_k(t)$$

Recall that to separate the K instantaneous frequencies we require that

$$\widehat{g}(s_u^k[\theta_k'(u) - \theta_l'(u)]) \ll 1, \quad \forall\, k \neq l, \quad \frac{\eta}{s_u^k} = \theta_k'(u)$$

which can be obtained if

$$\Delta\omega \leq s_u^k|\theta_k'(u) - \theta_l'(u)| = \frac{\eta|\theta_k'(u) - \theta_l'(u)|}{\theta_k'(u)}, \quad k \neq l$$

Under the assumptions on the window g, the primary free parameter one has is the frequency η. There is a tension between on the one hand wanting to make η large relative to the bandwidth, so that the wavelet is nearly analytic, the error terms $\varepsilon_{a,1}$ and (iv) are small, and so multiple instantaneous frequencies are separated; however, the second order error terms $\varepsilon_{a,2}$ and $\varepsilon_{\theta,2}$ may blow up if η is made too large.

Let us now return to the examples of the linear and hyperbolic chirps. We start with the sum of two hyperbolic chirps, which the windowed Fourier transform had trouble analyzing:

$$f(t) = a_1 \cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2 \cos\left(\frac{\alpha_2}{\beta_2 - t}\right)$$

In this case $\theta_k(t) = \alpha_k/(\beta_k - t)$ and $\theta_k'(t) = \alpha_k/(\beta_k - t)^2$. Since the amplitudes a_1 and a_2 are constant, the second order term $\varepsilon_{a,2}(u, s_u) = 0$. The other second order error term is bounded as:

$$\begin{aligned}
\varepsilon_{\theta,2}(u, s_u) &\leq \max_{k=1,2} \sup_{|t-u| \leq \eta/2\theta_k'(u)} \eta^2 \frac{|\theta_k''(t)|}{\theta'(u)^2} \\
&\leq \max_{k=1,2} \sup_{|t-u| \leq \eta(\beta_k - u)^2/2\alpha_k} \eta^2 \frac{2\alpha_k}{(\beta_k - t)^3} \frac{(\beta_k - t)^4}{\alpha_k^2} \\
&\leq \max_{k=1,2} \sup_{|t-u| \leq \eta(\beta_k - u)^2/2\alpha_k} \eta^2 \frac{2(\beta_k - t)}{\alpha_k}
\end{aligned}$$

This error term will be small if

$$\eta^2 \ll \frac{\alpha_k}{\beta_k - t}$$

This will be the case if, for example, $t \in [0, \beta_k)$ and $\eta \ll \sqrt{\alpha_k/\beta_k}$. Figure 16 illustrates how the wavelet ridges successfully follow the instantaneous frequencies of the two hyperbolic chirps.

Now let us go back to the two linear chirps signal

$$f(t) = a_1 \cos(bt^2 + ct) + a_2 \cos(bt^2)$$

which has frequencies $\theta_1(t) = bt^2 + ct$ and $\theta_2(t) = bt^2$. We thus have

$$\frac{|\theta_1'(u) - \theta_2'(u)|}{\theta_1'(u)} = \frac{|c|}{2bt} \to 0 \text{ as } t \to +\infty$$

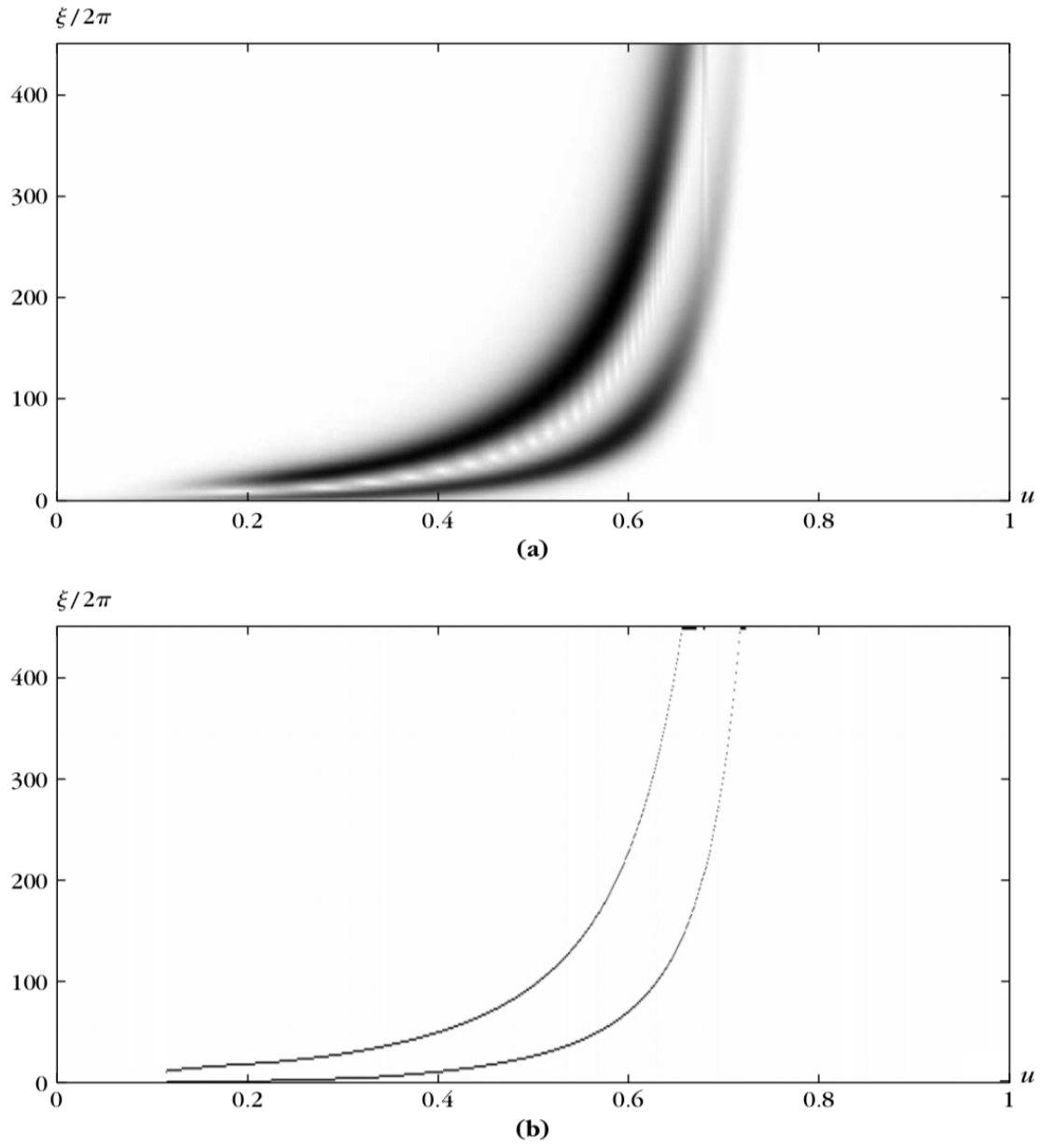


Figure 16: Analysis of a signal consisting of two hyperbolic chirps. (a) Normalized scalogram $\widetilde{P}_W f(u, \eta/s)$; (b) Wavelet ridges.

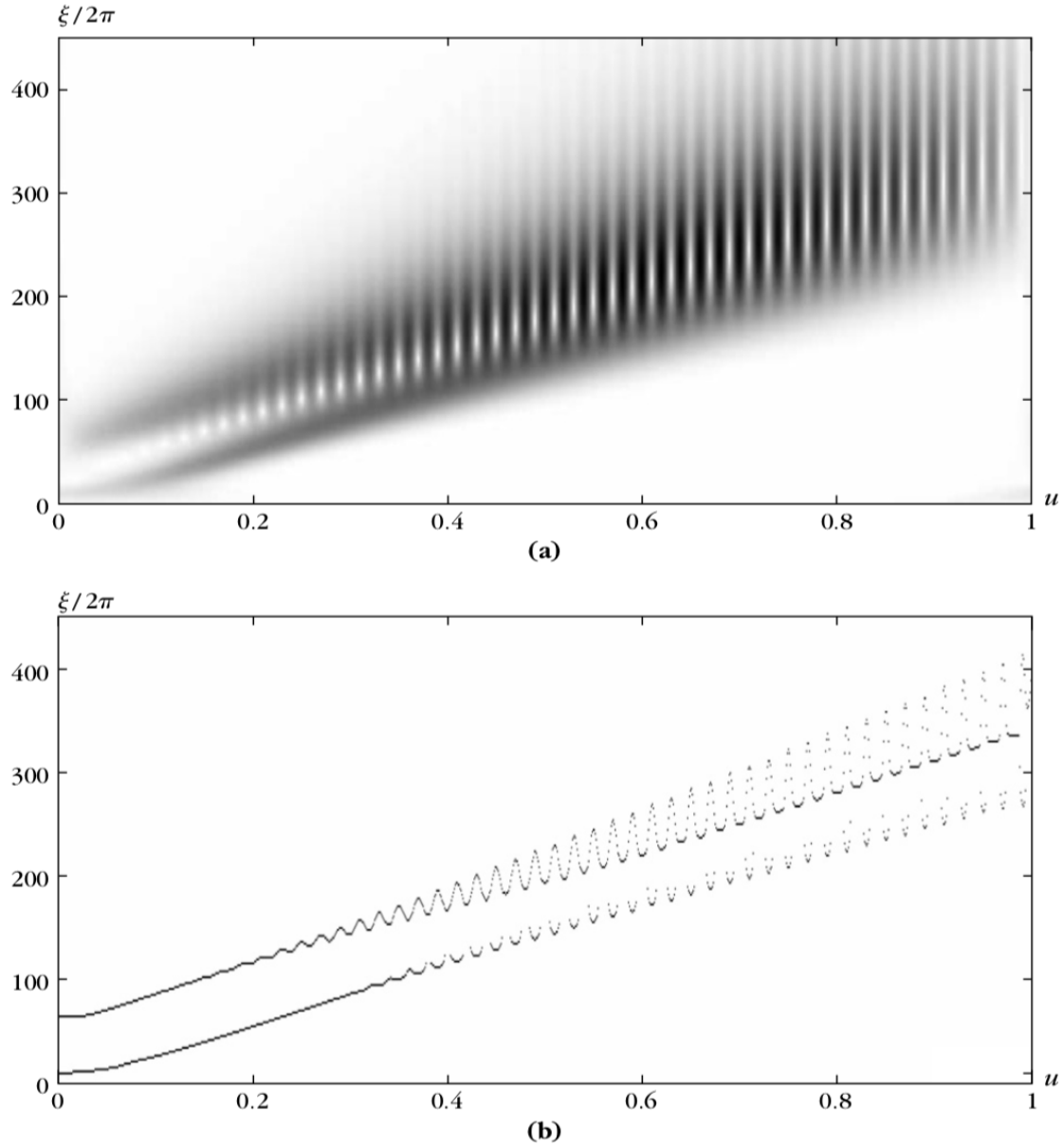


Figure 17: Analysis of a signal consisting of two linear chirps. (a) Normalized scalogram $\widetilde{P}_W f(u, \eta/s)$; (b) Wavelet ridges.

Thus for some finite t we will not be able to separate the instantaneous frequencies because of interferences. Figure 17 illustrates this phenomenon, as for large t the ridges follow the interference patterns rather than the instantaneous frequencies.

The take home message is that better is more sparse. This is true of course for compression, where sparser representations require less memory to store. But the linear and hyperbolic chirp examples show that sparsity also means we have found a time frequency transform that has a resolution adapted to the time frequency properties of the signal, in which case the number of ridge points is small. Conversely, if signal structures do not match our dictionary of time frequency atoms, then their energy will diffuse over many such atoms which produces more ridge points.

Exercise 39. Read Section 4.4.3 of *A Wavelet Tour of Signal Processing.*

Exercise 40. Adapt your windowed Fourier ridge code from Exercise 35 to compute the normalized scalogram $\widetilde{P}_W f(u, \eta/s)$ and corresponding wavelet ridges. Test your code on the sum of two linear chirps and the sum of two hyperbolic chirps. Turn in plots of your wavelet ridges. Do you get something similar to the plots in Figures 16 and 17?

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 12: Real Wavelets

February 18, 2020

Lecturer: Matthew Hirn

4.4.3 Real Wavelets

Section 4.3.1 of A Wavelet Tour of Signal Processing.

We now shift our focus to real wavelets. As we shall see, real valued wavelets are good for measuring sharp signal transitions, and in particular measuring the the regularity of $f(t)$ at a specific point $t = u$. Indeed, since $\int \psi = 0$, the wavelet transform $Wf(u,s) = \langle f, \psi_{u,s} \rangle$ measures the variation of f in a neighborhood of u proportional s. "Zooming in" on these variations will allow us to measure the regularity of f at u.

For now we show that like the analytic wavelet transform, the real wavelet transform is invertible and preserves the energy of f. We collect these results in the next theorem.

Theorem 4.7. *Let $\psi \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ be a real function such that*

$$C_\psi = \int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|^2}{\omega} \, d\omega < +\infty$$

Then, for any $f \in \mathbf{L}^2(\mathbb{R})$,

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} Wf(u,s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) du \, \frac{ds}{s^2}$$

and

$$\|f\|_2^2 = \frac{1}{C_\psi} \int_0^{+\infty} \int_{\mathbb{R}} |Wf(u,s)|^2 \, du \, \frac{ds}{s^2}$$

The proof is nearly identical to the proof of Theorem 4.6, and so is omitted. An example of a real valued wavelet, which we shall use later and which also satisfies the conditions of Theorem 4.7, is the so called "Mexican hat wavelet." The Mexican hat wavelet is the second derivative of a suitably normalized Gaussian function $g_\sigma(t)$ with mean zero and standard deviation σ:

$$\psi(t) = g_\sigma''(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}} \left(\frac{t^2}{\sigma^2} - 1\right) e^{-t^2/2\sigma^2} \tag{30}$$

The Fourier transform of $\psi(t)$ is

$$\widehat{\psi}(\omega) = -\omega^2 \widehat{g}_\sigma(\omega) = -\frac{\sqrt{8}\sigma^{5/2}\pi^{1/4}}{\sqrt{3}} \omega^2 e^{-\sigma^2\omega^2/2}$$

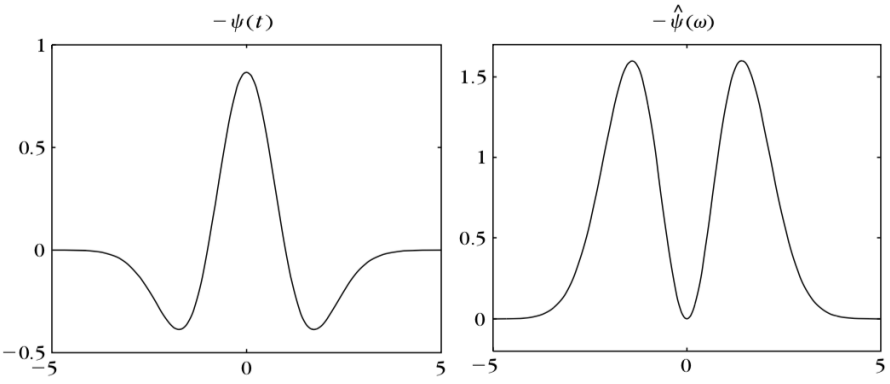


Figure 18: Mexican hat wavelet for $\sigma = 1$ and its Fourier transform.

Figure 18 plots the Mexican hat wavelet and its Fourier transform.

Figure 19 computes the wavelet transform of the signal from Figure 13 using the Mexican hat wavelet. The maximum scale is smaller than one because the support of $f(t)$ is limited to $t \in [0, 1]$. In all numerical calculations, including this one, the minimum scale is limited by the sampling interval of the discretized signal due to aliasing effects. When the scale decreases, the wavelet transform $Wf(u, s)$ has a rapid decay to zero in the regions where the signal is regular. Isolated singularities create cones of large amplitude wavelet coefficients that converge to the locations of the isolated singularities, as on the left hand side of the signal. The right hand side of the signal is singular almost everywhere. If this part can be modeled as a random process or multifractal, then under certain assumptions the distribution of singularities can be estimated from $Wf(u, s)$, which can characterize the underlying signal generation process.

In numerical applications, both the minimum and maximum scale are limited. We now examine the wavelet transform when we compute $Wf(u, s)$ only for $s < s_0$. In this case we lose the low frequency components of $f(t)$, since the supports of $\overline{\psi}_s(\omega) = \sqrt{s}\widehat{\psi}^*(s\omega)$ as $s \to +\infty$ collapse in around $\omega = 0$; see Figure 20 for an illustration.

In order to recover this lost low frequency information, we introduce a single *scaling function* ϕ that is an aggregation of all wavelets at scales larger than one. The modulus of its Fourier transform is defined as:

$$|\widehat{\phi}(\omega)|^2 = \int_1^{+\infty} |\widehat{\psi}(s\omega)|^2 \frac{ds}{s} = \int_\omega^{+\infty} \frac{|\widehat{\psi}(\xi)|^2}{\xi} d\xi \tag{31}$$

The complex phase of $\widehat{\phi}(\omega)$ can be arbitrarily chosen; in particular we can set it to zero so that ϕ is real valued. One can verify that $\|\phi\|_2 = 1$. The definition of $\widehat{\phi}(\omega)$ and the admissibility condition yields:

$$|\widehat{\phi}(\omega)|^2 \leq |\widehat{\phi}(0)|^2 = \lim_{\xi \to 0} |\widehat{\phi}(\xi)|^2 = C_\psi, \quad \forall\, \omega \in \mathbb{R}$$

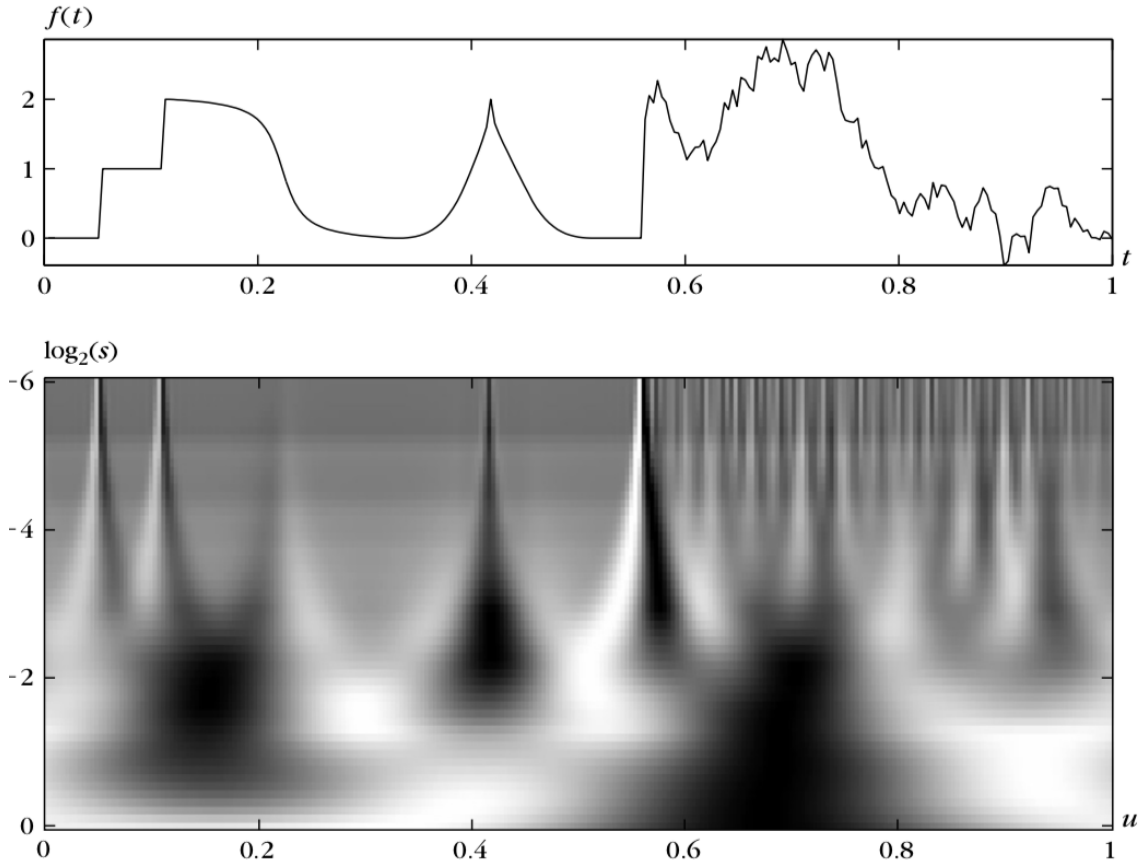


Figure 19: Real wavelet transform $Wf(u,s)$ computed with a Mexcian hat wavelet. The vertical axis represents $\log_2 s$, Black, gray and white points correspond, respectively, to positive, zero, and negative wavelet coefficients.

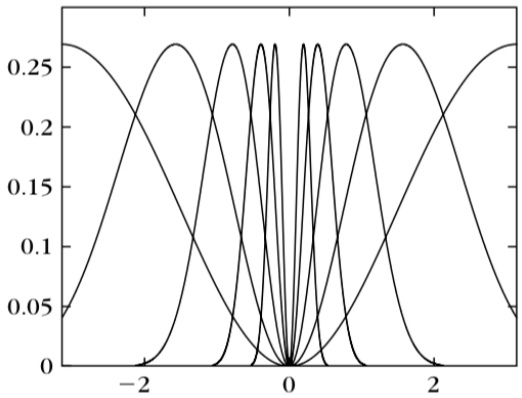


Figure 20: Scaled Fourier transform $|\widehat{\psi}(2^j\omega)|^2$ for $1 \leq j \leq 5$ and $\omega \in [-\pi, \pi]$. Notice the gap around $\omega = 0$ due to the limitation of the largest scale $s = 2^5$.

Denote by $\phi_s(t)$ the scaling of $\phi(t)$ by s:

$$\phi_s(t) = \frac{1}{\sqrt{s}}\phi\left(\frac{t}{s}\right), \quad \overline{\phi}_s(t) = \phi_s^*(-t)$$

The low frequency approximation of f at scale s is

$$Lf(u,s) = f * \overline{\phi}_s(u)$$

In this case, the wavelet inversion formula becomes:

$$f(t) = \frac{1}{C_\psi}\int_0^{s_0} Wf(\cdot,s) * \psi_s(t)\,\frac{ds}{s^2} + \frac{1}{C_\psi s_0} Lf(\cdot,s_0) * \phi_{s_0}(t) \tag{32}$$

For the Mexican hat wavelet defined in (30), the Fourier transform of the scaling function is:

$$\widehat{\phi}(\omega) = \frac{2\sigma^{5/2}\pi^{1/4}}{\sqrt{3}}\sqrt{\omega^2 + \frac{1}{\sigma^2}}e^{-\sigma^2\omega^2/2}$$

See Figure 21 for a plot.

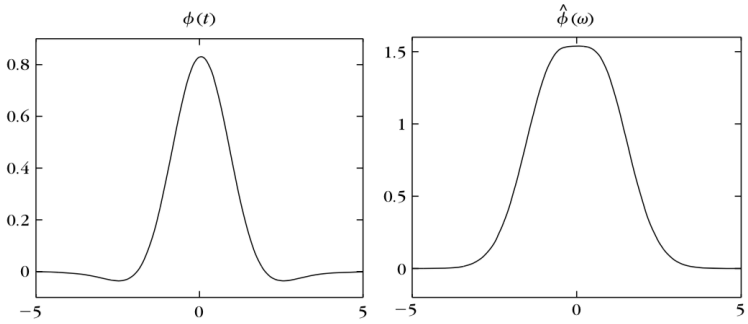


Figure 21: Scaling function associated to a Mexican hat wavelet and its Fourier transform.

Exercise 41. Read Section 4.5 of *A Wavelet Tour of Signal Processing*.

Exercise 42. Let ϕ be the scaling function defined by (31). Prove that $\|\phi\|_2 = 1$.

Exercise 43. Prove the reconstruction formula given in (32), which can be rewritten as:

$$f(t) = \frac{1}{C_\psi}\int_0^{s_0} f * \overline{\psi}_s * \psi_s(t)\,\frac{ds}{s^2} + \frac{1}{C_\psi s_0} f * \overline{\phi}_{s_0} * \phi_{s_0}(t)$$

Exercise 44. [20 points] Implement the real wavelet transform $Wf(u,s)$ using the Mexican hat wavelet. Write down a signal $f(t)$ similar to the one from Figure 13 (does not have to be exactly the same!) and compute $Wf(u,s)$ numerically. Turn in a plot of your signal and a plot of $Wf(u,s)$, as in Figure 19.

5 Wavelet Zoom

Chapter 6 of A Wavelet Tour of Signal Processing [1].

Wavelet transforms can zoom in and characterize the regularity of a function $f(t)$ at individual points $t = v$. We show how in this section.

5.1 Lipschitz Regularity

Section 6.1 of A Wavelet Tour of Signal Processing

5.1.1 Lipschitz Definition and Fourier Analysis

Section 6.1.1 of A Wavelet Tour of Signal Processing

We begin by defining the spaces $\dot{\mathbf{C}}^\alpha(\mathbb{R})$ and $\mathbf{C}^\alpha(\mathbb{R})$ for any $\alpha > 0$, $\alpha \notin \mathbb{Z}$. Let $0 < \alpha < 1$. Define the modulus of continuity of $f(t)$, $\omega_f(h)$, as:

$$\omega_f(h) = \sup\{|f(t) - f(u)| : |t - u| \leq h\}\,.$$

The space $\dot{\mathbf{C}}^\alpha(\mathbb{R})$, for $0 < \alpha < 1$, consists of those functions for which

$$\forall\, h > 0\,, \quad \omega_f(h) \leq K_f h^\alpha$$

Another way of writing this condition is:

$$\forall\, (t, u) \in \mathbb{R}^2\,, \quad |f(t) - f(u)| \leq K_f |t - u|^\alpha$$

We thus see that $\dot{\mathbf{C}}^\alpha(\mathbb{R})$ consists of those functions that satisfy a type of global Lipschitz condition; they are called α-Hölder functions. The space $\mathbf{C}^\alpha(\mathbb{R})$ contains all those functions in $\dot{\mathbf{C}}^\alpha(\mathbb{R})$ that are also bounded, i.e.,

$$\mathbf{C}^\alpha(\mathbb{R}) = \dot{\mathbf{C}}^\alpha(\mathbb{R}) \cap \mathbf{L}^\infty(\mathbb{R})$$

These definitions are extended to arbitrary $\alpha > 0$, $\alpha \notin \mathbb{Z}$, in the following way. Let $n < \alpha < n + 1$ for some $n \in \mathbb{N}$. Then

$$f \in \dot{\mathbf{C}}^\alpha(\mathbb{R}) \quad \Longleftrightarrow \quad f \in \mathbf{C}^n(\mathbb{R}) \text{ and } f^{(n)} \in \dot{\mathbf{C}}^{\alpha - n}(\mathbb{R})$$

where $\dot{\mathbf{C}}^{\alpha-n}(\mathbb{R})$ is defined as above since $0 < \alpha - n < 1$. Similarly,

$$f \in \mathbf{C}^\alpha(\mathbb{R}) \quad \Longleftrightarrow \quad f \in \mathbf{C}^n(\mathbb{R}) \text{ and } f^{(k)} \in \mathbf{C}^{\alpha-n}(\mathbb{R}) \;\; \forall\, k \leq n$$

We can link these definitions to Taylor's theorem. Suppose that $f \in \dot{\mathbf{C}}^\alpha(\mathbb{R})$ for $n < \alpha < n+1$. Let $v \in \mathbb{R}$ and let $J_v f(t)$ be the *jet* of f at v, which is the n-degree Taylor polynomial of f around v:

$$J_v f(t) = \sum_{k=0}^{n} \frac{f^{(k)}(v)}{k!} (t - v)^k$$

Denote the residual by $R_v f(t)$, i.e.,

$$R_v f(t) = f(t) - J_v f(t)$$

Then using Taylor's theorem and the fact that $f \in \dot{\mathbf{C}}^\alpha(\mathbb{R})$,

$$|R_v f(t)| \leq K|t-v|^\alpha$$

Conversely, suppose that $f(t)$ is a continuous function for which there exists a universal constant K and such that for each $v \in \mathbb{R}$, there exists a polynomial $p_v(t)$ of degree at most n such that

$$|f(t) - p_v(t)| \leq K|t-v|^\alpha$$

Then $f \in \dot{\mathbf{C}}^\alpha(\mathbb{R})$.

We can use this link with Taylor's theorem to define notions of local regularity rather than global regularity. In particular, a function $f(t)$ is pointwise Lipschitz α at $v \in \mathbb{R}$, for $n < \alpha \leq n+1$, if there exists $K_v > 0$ and a polynomial $p_v(t)$ of degree n such that

$$\forall\, t \in \mathbb{R}\,, \quad |f(t) - p_v(t)| \leq K_v|t-v|^\alpha \tag{33}$$

Furthermore, a function f is uniformly Lipschitz α over an interval $[a,b]$ if it satisfies (33) for all $v \in [a,b]$ with a constant K that is independent of v.

Remark 5.1. At each $v \in \mathbb{R}$ the polynomial $p_v(t)$ is unique. Additionally, if f is n times continuously differentiable in a neighborhood of v, then $p_v(t) = J_v f(t)$.

Remark 5.2. A function that is bounded but discontinuous at v is said to Lipschitz $\alpha = 0$ at v. If $\alpha < 1$ at v, then f is not differentiable at v and α characterizes the type of singularity.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 13: Wavelet Vanishing Moments

February 20, 2020

Lecturer: Matthew Hirn

The next theorem generalizes Theorem 2.15 by relating the decay of the Fourier transform of $f(t)$ to the α regularity of f.

Theorem 5.3. *Suppose that $f \in \mathbf{L}^1(\mathbb{R})$. If*

$$\int_{\mathbb{R}} |\widehat{f}(\omega)|(1+|\omega|^\alpha)\,d\omega < +\infty \tag{34}$$

then $f \in \mathbf{C}^\alpha(\mathbb{R})$.

Proof. Equation (34) implies that $\widehat{f} \in \mathbf{L}^1(\mathbb{R})$, and so the Fourier inversion formula (2) holds. We use it to prove $f \in \mathbf{L}^\infty(\mathbb{R})$:

$$\begin{aligned} |f(t)| \leq \frac{1}{2\pi}\left|\int_{\mathbb{R}} \widehat{f}(\omega)e^{i\omega t}\,d\omega\right| &\leq \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{f}(\omega)|\,d\omega \\ &\leq \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{f}(\omega)|(1+|\omega|^\alpha)\,d\omega < \infty \end{aligned}$$

Now suppose that $0 < \alpha < 1$ and show that $f \in \dot{\mathbf{C}}^\alpha(\mathbb{R})$. To do so we need to show there exists $K > 0$ such that

$$|f(t) - f(v)| \leq K|t-v|^\alpha, \quad \forall\, t, v \in \mathbb{R}$$

By the Fourier inversion formula (2) we have that

$$f(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(\omega)e^{i\omega t}\,d\omega$$

It follows that

$$\frac{|f(t)-f(v)|}{|t-v|^\alpha} \leq \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{f}(\omega)|\frac{|e^{i\omega t}-e^{i\omega v}|}{|t-v|^\alpha}\,d\omega$$

For $|\omega| \geq |t-v|^{-1}$,

$$\frac{|e^{i\omega t}-e^{i\omega v}|}{|t-v|^\alpha} \leq \frac{2}{|t-v|^\alpha} \leq 2|\omega|^\alpha \tag{35}$$

On the other hand, for $|\omega| \leq |t-v|^{-1}$, we note that if a function $h \in \mathbf{C}^1(\mathbb{R})$ with bounded derivative then

$$|h(t)-h(v)| \leq K|t-v|, \quad K = \sup_{u\in\mathbb{R}} |h'(u)|$$

Note that $e_\omega \in \mathbf{C}^1(\mathbb{R})$, where $e_\omega(t) = e^{i\omega t}$, and $|e'_\omega(t)| = |\omega|$. Therefore,

$$\frac{|e^{i\omega t} - e^{i\omega v}|}{|t-v|^\alpha} \leq \frac{|\omega||t-v|}{|t-v|^\alpha} = |\omega||t-v|^{1-\alpha} \leq |\omega||\omega|^{\alpha-1} = |\omega|^\alpha \tag{36}$$

Combining (35) and (36), we obtain

$$\frac{|f(t)-f(v)|}{|t-v|^\alpha} \leq \frac{1}{2\pi}\int_{\mathbb{R}} 2|\widehat{f}(\omega)||\omega|^\alpha \, d\omega = K$$

Equation (34) ensures that $K < \infty$, and so $f \in \mathbf{C}^\alpha(\mathbb{R})$.

We now extend the result to $\alpha > 1$, $\alpha \notin \mathbb{Z}$. Let $n = \lfloor \alpha \rfloor$. Theorem 2.15 proves that $f \in \mathbf{C}^n(\mathbb{R})$. Recall that $\widehat{f^{(k)}}(\omega) = (i\omega)^k \widehat{f}(\omega)$. Equation (34) gives:

$$\int_{\mathbb{R}} |\widehat{f^{(k)}}(\omega)|(1+|\omega|^{\alpha-n})\, d\omega = \int_{\mathbb{R}} |\widehat{f}(\omega)|(|\omega|^k + |\omega|^{\alpha-n+k})\, d\omega < \infty$$

Thus by our work above, we have that $f^{(k)} \in \mathbf{C}^{\alpha-n}(\mathbb{R})$ for $k \leq n$, which proves that $f \in \mathbf{C}^\alpha(\mathbb{R})$. □

As we have discussed previously for $\mathbf{C}^n$-smooth functions, the decay of the Fourier transform can only indicate the minimum regularity of $f(t)$. Wavelet transforms characterize both the global and pointwise regularity of functions.

Exercise 45. Read Section 6.1.1 of *A Wavelet Tour of Signal Processing.*

Exercise 46. Consider the function

$$f(t) = t \sin\left(\frac{1}{t}\right)$$

(a) Prove that $f(t)$ is pointwise Lipschitz 1 for all $t \in (-1, 1)$.

(b) Prove that $f \in \mathbf{C}^\alpha(-1,1)$ only for $\alpha \leq 1/2$ (*Hint:* Consider the points $t_n = (n + 1/2)^{-1}\pi^{-1}$).

5.1.2 Wavelet Vanishing Moments

Section 6.1.2 of A Wavelet Tour of Signal Processing.

We assume throughout that $\psi(t)$ is a real valued wavelet. A wavelet ψ has n vanishing moments if

$$\int_{\mathbb{R}} t^k \psi(t)\, dt = 0, \quad \forall\, 0 \leq k < n$$

A wavelet ψ with n vanishing moments is orthogonal to polynomials of degree $n-1$.

Suppose now that f is Lipschitz $\alpha \leq n$ at v, so that

$$f(t) = p_v(t) + \varepsilon_v(t)$$

with $p_v(t)$ a polynomial of degree $n-1$ and

$$|\varepsilon_v(t)| \leq K|t-v|^\alpha$$

We have that

$$Wp_v(u,s) = \int_{\mathbb{R}} p_v(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt = \sqrt{s} \int_{\mathbb{R}} p_v(st'+u)\psi(t')\, dt' = 0$$

Therefore,

$$Wf(u,s) = Wp_v(u,s) + W\varepsilon_v(u,s) = W\varepsilon_v(u,s)$$

Thus a wavelet transform with n vanishing moments analyzes $f(t)$ around $t = v$ by ignoring the polynomial approximation of $f(t)$ and focusing on the residual $\varepsilon_v(t)$.

A wavelet ψ has fast decay if

$$\forall\, m \in \mathbb{N},\ \ \exists\, C_m \text{ such that } |\psi(t)| \leq \frac{C_m}{1+|t|^m}, \quad \forall\, t \in \mathbb{R}$$

The following theorem shows that a wavelet ψ with fast decay and n vanishing moments is the n^{th} derivative of a function $\theta(t)$. The resulting wavelet transform is thus a multiscale differential operator.

Theorem 5.4. *A wavelet $\psi(t)$ with fast decay has n vanishing moments if and only if there exists $\theta(t)$ with a fast decay such that*

$$\psi(t) = (-1)^n \theta^{(n)}(t)$$

Consequently,

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f * \overline{\theta}_s)(u)$$

where $\overline{\theta}_s(t) = s^{-1/2}\theta(-t/s)$. Furthermore, ψ has no more than n vanishing moments if and only if

$$\int_{\mathbb{R}} \theta(t)\, dt \neq 0$$

Proof. Suppose that ψ has fast decay and n vanishing moments. Since ψ has fast decay we must have that $\widehat{\psi} \in \mathbf{C}^\infty(\mathbb{R})$; this follows from Theorem 2.15 by setting $f = \widehat{\psi}$. Thus we can differentiate $\widehat{\psi}(\omega)$ as many times as we like.

Recall that the Fourier transform of $h(t) = (-it)^k \psi(t)$ is $\widehat{h}(\omega) = \widehat{\psi}^{(k)}(\omega)$. It follows that

$$\widehat{\psi}^{(k)}(0) = \int_{\mathbb{R}} (-it)^k \psi(t)\, dt = (-i)^k \int_{\mathbb{R}} t^k \psi(t)\, dt = 0, \quad \forall\, 0 \leq k < n$$

We can therefore write $\widehat{\psi}$ as

$$\widehat{\psi}(\omega) = (-i\omega)^n \widehat{\theta}(\omega)$$

where $\widehat{\theta} \in \mathbf{L}^\infty(\mathbb{R})$ since $\widehat{\psi} \in \mathbf{L}^\infty(\mathbb{R})$. It follows that

$$\psi(t) = (-1)^n \theta^{(n)}(t)$$

The fast decay of $\theta(t)$ is proved with an induction on n. For $n = 1$,

$$\widehat{\psi}(\omega) = -i\omega\widehat{\theta}(\omega) \Longrightarrow \psi(t) = -\theta'(t)$$

It follows that

$$\theta(t) = -\int_{-\infty}^{t} \psi(u)\, du$$

Thus, using the fast decay of $\psi(t)$,

$$|\theta(t)| \leq \int_{-\infty}^{t} |\psi(u)|\, du \leq \int_{-\infty}^{t} \frac{C_m}{1+|u|^m}\, du \leq \frac{C'_{m-1}}{1+|t|^{m-1}}, \quad \forall\, m \geq 2$$

Now make the inductive hypothesis that if $\Psi(t)$ is any wavelet with fast decay and

$$\widehat{\Psi}(\omega) = (-i\omega)^k \widehat{\Theta}(\omega), \quad 1 \leq k \leq n$$

then $\Theta(t)$ has fast decay. Consider now a wavelet ψ with fast decay that has $n+1$ vanishing moments, so that $\widehat{\psi}(\omega) = (-i\omega)^{n+1}\widehat{\theta}(\omega)$. Define

$$\widehat{\Theta}(\omega) = -i\omega\widehat{\theta}(\omega) \Longrightarrow \widehat{\psi}(\omega) = (-i\omega)^n \widehat{\Theta}(\omega)$$

By the inductive hypothesis, $\Theta(t)$ has fast decay. But then since $\widehat{\Theta}(\omega) = -i\omega\widehat{\theta}(\omega)$, we can apply the inductive hypothesis again to conclude that $\theta(t)$ has fast decay.

Conversely, suppose that $\psi(t) = (-1)^n \theta^{(n)}(t)$ and $\theta(t)$ has fast decay. Because of the fast decay,

$$|\widehat{\theta}(\omega)| \leq \int_{\mathbb{R}} |\theta(t)|\, dt \leq \int_{\mathbb{R}} \frac{C_m}{1+|t|^m}\, dt < +\infty, \quad m \geq 2$$

Thus $\widehat{\theta} \in \mathbf{L}^\infty(\mathbb{R})$. The Fourier transform of $\psi(t)$ is

$$\widehat{\psi}(\omega) = (-i\omega)^n \widehat{\theta}(\omega)$$

It follows that $\widehat{\psi}^{(k)}(0) = 0$ for $k < n$. But then

$$\int_{\mathbb{R}} t^k \psi(t)\, dt = i^k \widehat{\psi}^{(k)}(0) = 0, \quad 0 \leq k < n$$

Thus $\psi(t)$ has n vanishing moments.

To test whether $\psi(t)$ has more than n vanishing moments, we compute:

$$\int_{\mathbb{R}} t^n \psi(t)\, dt = i^n \widehat{\psi}^{(n)}(0) = (-i)^n n! \widehat{\theta}(0)$$

Clearly then ψ has no more than n vanishing moments if and only if

$$\widehat{\theta}(0) = \int_{\mathbb{R}} \theta(t)\, dt \neq 0$$

Recall the wavelet transform can be written as

$$Wf(u,s) = f * \overline{\psi}_s(u)$$

where

$$\overline{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{-t}{s}\right) = \frac{(-1)^n}{\sqrt{s}} \theta^{(n)}\left(-\frac{t}{s}\right) = (-1)^n \overline{\theta}_s^{(n)}(t)$$

A simple calculation also shows that

$$\frac{d^n}{dt^n}\overline{\theta}_s(t) = \frac{1}{s^n}\frac{(-1)^n}{\sqrt{s}} \theta^{(n)}\left(-\frac{t}{s}\right) = \frac{(-1)^n}{s^n}\overline{\theta}_s^{(n)}(t) = \frac{\overline{\psi}_s(t)}{s^n}$$

Therefore $\overline{\psi}_s(t) = s^n (d^n/dt^n)\overline{\theta}_s(t)$. We then have:

$$Wf(u,s) = f * \overline{\psi}_s(u) = s^n f * \theta_s^{(n)}(u) = s^n \frac{d^n}{du^n}(f * \theta)(u)$$

□

If $K = \widehat{\theta}(0) \neq 0$, then the convolution $f * \overline{\theta}_s(t)$ can be interpreted as a weighted average of f with a kernel dilated by s. Theorem 5.4 proves that $Wf(u,s)$ is an n^{th} order derivative of an averaging of f over a domain proportional to s and centered at u. Figure plots $Wf(u,s)$ calculated with $\psi(t) = -\theta'(t)$, where $\theta(t)$ is a Gaussian. Notice how the sign and magnitude of the wavelet coefficients corresponds to the derivative of f averaged over a window of size proportional to s. Compare to Figure 19, which computed $Wf(u,s)$ with the Mexican hat wavelet $\psi(t) = \theta''(t)$ (θ again a Gaussian).

Since $\theta(t)$ has fast decay, once can verify that for any f that is continuous at u,

$$\lim_{s \to 0} f * \frac{1}{\sqrt{s}} \overline{\theta}_s(u) = K f(u)$$

In the sense of distributions, we write

$$\lim_{s \to 0} \frac{1}{\sqrt{s}} \overline{\theta}_s(t) = K\delta(t)$$

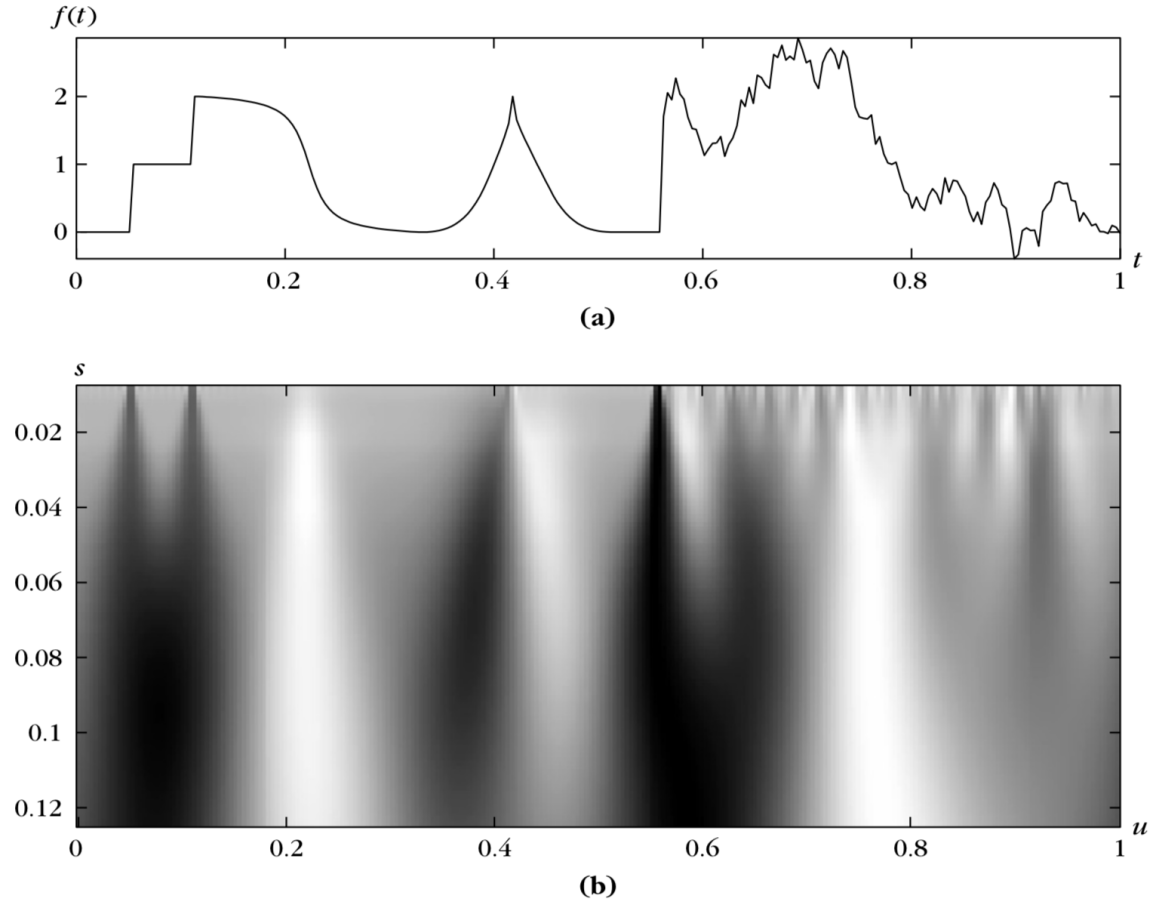


Figure 22: Wavelet transform $Wf(u,s)$ calculated with $\psi = -\theta'$, where θ is a Gaussian, for the signal $f(t)$ shown in (a). Position parameter u and scale s vary, respectively, along the horizontal and vertical axes. (b) Black, gray, and white points correspond to positive, zero, and negative wavelet coefficients. Singularities create large-amplitude coefficients in their cone of influence.

If f is n times continuously differentiable in the neighborhood of u, then using Theorem 5.4,

$$\lim_{s\to 0} \frac{Wf(u,s)}{s^{n+1/2}} = \lim_{s\to 0} \frac{1}{\sqrt{s}} \frac{d^n}{dt^n}(f * \overline{\theta}_s)(u) = \lim_{s\to 0} f^{(n)} * \frac{1}{\sqrt{s}}\overline{\theta}_s(u) = Kf^{(n)}(u) \tag{37}$$

In particular, if $f \in \mathbf{C}^n(\mathbb{R})$, then $|Wf(u,s)| = O(s^{n+1/2})$. This gives us a first relation between the decay of $|Wf(u,s)|$ as $s \to 0$ and the uniform regularity of f. Next we push harder and obtain finer relations.

Exercise 47. Read Section 6.1.2 of *A Wavelet Tour of Signal Processing.*

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 14 & 15: Wavelet Characterization of Pointwise Regularity

February 25 & 27, 2020

Lecturer: Matthew Hirn

5.1.3 Regularity Measurements with Wavelets

Section 6.1.3 of A Wavelet Tour of Signal Processing.

We now prove that "zooming in" on the wavelet coefficients of a signal f characterizes the pointwise regularity of f. We utilize a real valued wavelet $\psi \in \mathbf{C}^n(\mathbb{R})$ with n vanishing moments and with derivatives that have fast decay. The latter point means that for any $0 \leq k \leq n$ and $m \in \mathbb{N}$ there exists $C_m \geq 0$ such that

$$|\psi^{(k)}(t)| \leq \frac{C_m}{1+|t|^m}, \quad \forall\, t \in \mathbb{R}$$

Let $n-1 < \alpha \leq n$ for some $n \in \mathbb{N}$. Recall that $f(t)$ is Lipschitz $\alpha > 0$ at $v \in \mathbb{R}$ if there exists a degree $n-1$ polynomial $p_v(t)$ and a constant $K_v \geq 0$ such that

$$\forall\, t \in \mathbb{R}\,, \quad |f(t) - p_v(t)| \leq K_v |t-v|^\alpha$$

Additionally, f is uniformly Lipschitz α on an interval $[a,b]$ if it satisfies the above condition for all $v \in [a,b]$ with a constant K that is independent of v.

Theorem 5.5. *If $f \in \mathbf{L}^2(\mathbb{R})$ is Lipschitz $\alpha \leq n$ at $v \in \mathbb{R}$, then there exits $A > 0$ such that*

$$|Wf(u,s)| \leq A s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^\alpha\right), \quad \forall\, (u,s) \in \mathbb{R} \times (0,\infty) \tag{38}$$

Conversely if $\alpha < n$ is not an integer and there exist A and $\alpha' < \alpha$ such that

$$\forall\, (u,s) \in \mathbb{R} \times (0,\infty)\,, \quad |Wf(u,s)| \leq A s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha'}\right) \tag{39}$$

then $f(t)$ is Lipschitz α at $t = v$.

Proof. We first prove (38). Since f is Lipschitz α at v, there exists a polynomial $p_v(t)$ of degree $n-1$ and K such that

$$|f(t) - p_v(t)| \leq K|t-v|^\alpha$$

Recall that since ψ has n vanishing moments, $Wp_v(u,s) = 0$. Therefore:

$$\begin{aligned}|Wf(u,s)| &= \left| \int_{\mathbb{R}} [f(t) - p_v(t)] \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt \right| \\ &\leq \int_{\mathbb{R}} K|t-v|^\alpha \frac{1}{\sqrt{s}} \left| \psi\left(\frac{t-u}{s}\right) \right| dt\end{aligned}$$

Now make the change of variables $x = (t-u)/s$, which induces the change of measure $dt = sdx$,

$$\begin{aligned}|Wf(u,s)| &\leq K\sqrt{s} \int_{\mathbb{R}} |sx + u - v|^\alpha |\psi(x)|\, dx \\ &\leq K\sqrt{s} \int_{\mathbb{R}} 2^\alpha (|sx|^\alpha + |u-v|^\alpha) |\psi(x)|\, dx \\ &= K 2^\alpha s^{\alpha+1/2} \left(\int_{\mathbb{R}} |x|^\alpha |\psi(x)|\, dx + \left|\frac{u-v}{s}\right|^\alpha \int_{\mathbb{R}} |\psi(x)|\, dx \right) \\ &\leq A s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^\alpha \right)\end{aligned}$$

where

$$A = K 2^\alpha \max\left(\int_{\mathbb{R}} |x|^\alpha |\psi(x)|\, dx, \|\psi\|_1 \right)$$

and where we used the fact that $|a+b|^\alpha \leq 2^\alpha (|a|^\alpha + |b|^\alpha)$.

Now we prove (39). This is a difficult proof that adapts the Littlewood-Paley approach referenced earlier in the course. Recall the real wavelet inverse formula from Theorem 4.7,

$$f(t) = \frac{1}{C_\psi} \int_{\mathbb{R}} \int_0^{+\infty} Wf(u,s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \frac{ds}{s^2}\, du$$

We are going to break up the scale integral into dyadic intervals. Define:

$$\Delta_j(t) = \frac{1}{C_\psi} \int_{\mathbb{R}} \int_{2^j}^{2^{j+1}} Wf(u,s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \frac{ds}{s^2}\, du, \quad j \in \mathbb{Z}$$

Note that we have the following Littlewood-Paley type sum:

$$f(t) = \sum_{j \in \mathbb{Z}} \Delta_j(t) \tag{40}$$

Let $\Delta_j^{(k)}(t)$ be the k^{th} order derivative of $\Delta_j(t)$. To prove that f is Lipschitz α at v we need a polynomial $p_v(t)$ of degree $\lfloor \alpha \rfloor$ and a constant K such that $|f(t) - p_v(t)| \leq K|t-v|^\alpha$. We propose

$$p_v(t) = \sum_{k=0}^{\lfloor \alpha \rfloor} \left(\sum_{j \in \mathbb{Z}} \Delta_j^{(k)}(v) \right) \frac{(t-v)^k}{k!} \tag{41}$$

as a candidate. The remainder of the proof is in showing that (41) does the job. Notice that if f is $n-1$ times differentiable at v, then using (40) we have $\sum_j \Delta_j^{(k)}(v) = f^{(k)}(v)$, and we get the jet $J_v f(t) = p_v(t)$. However, we can compute $\Delta_j^{(k)}(t)$ for each $j \in \mathbb{Z}$ even when f is not n times differentiable at v. But then we need to show that $p_v(t)$ is well defined even when f is not n times differentiable at v, and in particular we need to show that $\sum_{j\in\mathbb{Z}} \Delta_j^{(k)}(v)$ is finite. Once we do that, we can think of (41) as a generalization of jets, and in particular the sums $\sum_j \Delta_j^{(k)}(v)$ as generalizations of pointwise derivatives at v.

We first prove that $\sum_j \Delta_j^{(k)}(v)$ is finite by getting appropriate upper bounds on $|\Delta_j^{(k)}(t)|$ for $k \le \lfloor \alpha \rfloor + 1 \le n$. To simplify notation, **we let K be a generic constant that may change from line to line, but does not depend on j and t.** Equation (39) plus the fast decay of the wavelet derivatives yield:

$$\begin{aligned}
|\Delta_j^{(k)}(t)| &= \frac{1}{C_\psi} \left| \int_{\mathbb{R}} \int_{2^j}^{2^{j+1}} Wf(u,s) \frac{1}{\sqrt{s}} \frac{d^k}{dt^k} \psi\left(\frac{t-u}{s}\right) \frac{ds}{s^2}\, du \right| \\
&\le \frac{1}{C_\psi} \int_{\mathbb{R}} \int_{2^j}^{2^{j+1}} |Wf(u,s)| \frac{1}{\sqrt{s}} \frac{1}{s^k} \left| \psi^{(k)}\left(\frac{t-u}{s}\right) \right| \frac{ds}{s^2}\, du \\
&\le K \int_{\mathbb{R}} \int_{2^j}^{2^{j+1}} s^{\alpha-k} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha'}\right) \frac{C_m}{1+|(t-u)/s|^m} \frac{ds}{s^2}\, du \qquad (42)
\end{aligned}$$

Now observe that on the interval $[2^j, 2^{j+1}]$, we have

$$\sup_{s\in[2^j,2^{j+1}]} s^{\alpha-k} = 2^{\alpha-k} 2^{j(\alpha-k)} = K 2^{j(\alpha-k)}$$

It follows that

$$\begin{aligned}
(42) &\le K \int_{\mathbb{R}} 2^{j(\alpha-k)} \left(1 + \left|\frac{u-v}{2^j}\right|^{\alpha'}\right) \frac{1}{1+|(t-u)/2^j|^m} \left(\int_{2^j}^{2^{j+1}} \frac{ds}{s^2}\right) du \\
&\le K \int_{\mathbb{R}} 2^{j(\alpha-k)} \left(1 + \left|\frac{u-v}{2^j}\right|^{\alpha'}\right) \frac{1}{1+|(t-u)/2^j|^m} \frac{du}{2^j} \qquad (43)
\end{aligned}$$

Now make the change of variables $u' = 2^{-j}(u-t)$ and once again use the inequality $|a+b|^{\alpha'} \le 2^{\alpha'}(|a|^{\alpha'} + |b|^{\alpha'})$ to arrive at:

$$\begin{aligned}
(43) &= K 2^{j(\alpha-k)} \int_{\mathbb{R}} \left(1 + \left|u' + \frac{t-v}{2^j}\right|^{\alpha'}\right) \frac{1}{1+|u'|^m}\, du' \\
&\le K 2^{j(\alpha-k)} \int_{\mathbb{R}} \frac{1 + 2^{\alpha'}|u|^{\alpha'} + 2^{\alpha'}|(t-v)/2^j|^{\alpha'}}{1+|u'|^m}\, du' \\
&\le K 2^{j(\alpha-k)} \left[\int_{\mathbb{R}} \frac{1+|u'|^{\alpha'}}{1+|u'|^m}\, du' + \left|\frac{t-v}{2^j}\right|^{\alpha'} \int_{\mathbb{R}} \frac{1}{1+|u'|^m}\, du' \right]
\end{aligned}$$

Choosing $m = \alpha' + 2$ yields:

$$|\Delta_j^{(k)}(t)| \leq K 2^{j(\alpha-k)} \left(1 + \left|\frac{t-v}{2^j}\right|^{\alpha'}\right), \quad \forall\, k \leq \lfloor\alpha\rfloor + 1 \tag{44}$$

At $t = v$, we obtain

$$|\Delta_j^{(k)}(v)| \leq K 2^{j(\alpha-k)}$$

which guarantees a fast decay of $|\Delta_j^{(k)}(v)|$ as $j \to -\infty$ for $k \leq \lfloor\alpha\rfloor$ (i.e., in the small scale regime), because α is not an integer and so $\alpha > \lfloor\alpha\rfloor$.

At large scales, since

$$|Wf(u,s)| = |f * \overline{\psi}_s(u)| \leq \|f\|_2 \|\psi\|_2$$

with the change variables $u' = (t-u)/s$ we have

$$\begin{aligned}
|\Delta_j^{(k)}(t)| &\leq \frac{1}{C_\psi} \int_{\mathbb{R}} \int_{2^j}^{2^{j+1}} |Wf(u,s)| \frac{1}{\sqrt{s}} \left|\frac{d^k}{dt^k}\psi\left(\frac{t-u}{s}\right)\right| \frac{ds}{s^2}\, du \\
&\leq \frac{\|f\|_2\|\psi\|_2}{C_\psi} \int_{\mathbb{R}} \int_{2^j}^{2^{j+1}} |\psi^{(k)}(u')| \frac{ds}{s^{3/2+k}}\, du' \\
&\leq \frac{K\|f\|_2\|\psi\|_2\|\psi^{(k)}\|_1}{C_\psi} 2^{-j(k+1/2)}
\end{aligned}$$

and therefore

$$|\Delta_j^{(k)}(v)| \leq K 2^{-j(k+1/2)}$$

Thus we can bound $\sum_j \Delta_j^{(k)}(v)$ since

$$\begin{aligned}
\forall\, k \leq \lfloor\alpha\rfloor, \quad \left|\sum_{j\in\mathbb{Z}} \Delta_j^{(k)}(v)\right| &\leq \sum_{j\in\mathbb{Z}} |\Delta_j^{(k)}(v)| \\
&= \sum_{j=-\infty}^{0} |\Delta_j^{(k)}(v)| + \sum_{j=1}^{+\infty} |\Delta_j^{(k)}(v)| \\
&\leq K \sum_{j=-\infty}^{0} 2^{j(\alpha-k)} + K \sum_{j=1}^{+\infty} 2^{-j(k+1/2)} \\
&< +\infty
\end{aligned}$$

With the Littlewood-Paley sum (40) we compute:

$$\begin{aligned}
|f(t) - p_v(t)| &= \left|\sum_{j\in\mathbb{Z}} \left(\Delta_j(t) - \sum_{k=0}^{\lfloor\alpha\rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!}\right)\right| \\
&\leq \sum_{j\in\mathbb{Z}} \left|\Delta_j(t) - \sum_{k=0}^{\lfloor\alpha\rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!}\right|
\end{aligned} \tag{45}$$

The sum over the scales $j \in \mathbb{Z}$ is divided in two at J such that

$$2^{J-1} \leq |t-v| \leq 2^J$$

We define:

$$I_J = \sum_{j \geq J} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor \alpha \rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right|$$
$$II_J = \sum_{j < J} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor \alpha \rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right|$$

Note this means our constant K cannot depend on J, since J depends on t. The summands of (45) are $\lfloor \alpha \rfloor$ Taylor approximations of $\Delta_j(t)$ around v. For the large scales corresponding to $j \geq J$, we can use the classical Taylor's theorem to get a bound:

$$\begin{aligned} I_J &= \sum_{j \geq J} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor \alpha \rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right| \\ &\leq \sum_{j \geq J} \frac{|t-v|^{\lfloor \alpha \rfloor + 1}}{(\lfloor \alpha \rfloor + 1)!} \sup_{h \in [t,v]} |\Delta_j^{\lfloor \alpha \rfloor + 1}(h)| \\ &\leq K |t-v|^{\lfloor \alpha \rfloor + 1} \sum_{j \geq J} \sup_{h \in [t,v]} |\Delta_j^{\lfloor \alpha \rfloor + 1}(h)| \end{aligned}$$

Inserting the bound (44) yields:

$$I_J \leq K |t-v|^{\lfloor \alpha \rfloor + 1} \sum_{j \geq J} 2^{-j(\lfloor \alpha \rfloor + 1 - \alpha)} \left(1 + \left| \frac{t-v}{2^j} \right|^{\alpha'} \right)$$

Since $|t-v| \leq 2^J$ we have $|(t-v)/2^j| \leq 1$ for $j \geq J$. Therefore we have arrived at:

$$I_J \leq K |t-v|^{\lfloor \alpha \rfloor + 1} \sum_{j \geq J} 2^{-j(\lfloor \alpha \rfloor + 1 - \alpha)}$$

We need to bound the series. We will do so with the following proposition.

Proposition 5.6. *For any $0 \leq r < 1$ and for any $J \in \mathbb{Z}$,*

$$\sum_{j \geq J} r^j = \frac{r^J}{1-r}$$

Proof of the Proposition 5.6. The proof will rely on the well known fact:

$$\forall\, R \neq 1\,, \quad \sum_{j=0}^{K-1} R^j = \frac{1-R^K}{1-R}$$

which gives the result for $J = 0$ by taking $R = r$ and $K = +\infty$. For $J > 0$ we have

$$\sum_{j \geq J} r^j = \sum_{j \geq 0} r^j - \sum_{j=0}^{J-1} r^j = \frac{1}{1-r} - \frac{1-r^J}{1-r} = \frac{r^J}{1-r}$$

For $J < 0$ we have:

$$\begin{aligned}
\sum_{j \geq J} r^j &= \sum_{j \geq 0} r^j + \sum_{j=J}^{-1} r^j = \frac{1}{1-r} + \sum_{j=1}^{-J} r^{-j} = \frac{1}{1-r} + \sum_{j=1}^{-J} (r^{-1})^j \\
&= \frac{1}{1-r} - 1 + \sum_{j=0}^{-J} (r^{-1})^j = \frac{r}{1-r} + \frac{1-(r^{-1})^{-J+1}}{1-r^{-1}} \\
&= \frac{r}{1-r} + \frac{1-r^{J-1}}{1-r^{-1}} = \frac{r}{1-r} + \frac{r-r^J}{r-1} = \frac{r^J}{1-r}
\end{aligned}$$

□

Now apply Proposition 5.6 with $r_\alpha = 2^{-(\lfloor \alpha \rfloor + 1 - \alpha)}$, which is less than one since $\lfloor \alpha \rfloor + 1 > \alpha$. We obtain:

$$I_J \leq K|t-v|^{\lfloor \alpha \rfloor + 1} \sum_{j \geq J} 2^{-j(\lfloor \alpha \rfloor + 1 - \alpha)} = K|t-v|^{\lfloor \alpha \rfloor + 1} \sum_{j \geq J} r_\alpha^j = K|t-v|^{\lfloor \alpha \rfloor + 1} \frac{r_\alpha^J}{1-r_\alpha}$$

Now $1/(1-r_\alpha)$ only depends on α and can be absorbed into K. Furthermore, since $2^{-J} \leq |t-v|^{-1}$ we also have:

$$r_\alpha^J = 2^{-J(\lfloor \alpha \rfloor + 1 - \alpha)} \leq |t-v|^{-(\lfloor \alpha \rfloor + 1 - \alpha)}$$

Thus we obtain:

$$I_J \leq K|t-v|^{\lfloor \alpha \rfloor + 1} r_\alpha^J \leq K|t-v|^{\lfloor \alpha \rfloor + 1} |t-v|^{-(\lfloor \alpha \rfloor + 1 - \alpha)} = K|t-v|^\alpha$$

Now consider the sum over $j < J$ and use (44),

$$
\begin{aligned}
II_J &= \sum_{j<J} \left| \Delta_j(t) - \sum_{k=0}^{\lfloor \alpha \rfloor} \Delta_j^{(k)}(v) \frac{(t-v)^k}{k!} \right| \\
&\leq K \sum_{j<J} \left[|\Delta_j(t)| + \sum_{k=0}^{\lfloor \alpha \rfloor} |\Delta_j(v)| \frac{|t-v|^k}{k!} \right] \\
&\leq K \sum_{j<J} \left[2^{j\alpha} \left(1 + \left| \frac{t-v}{2^j} \right|^{\alpha'} \right) + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{|t-v|^k}{k!} 2^{j(\alpha-k)} \right] \\
&= K \sum_{j<J} \left[2^{j\alpha} + 2^{j(\alpha-\alpha')} |t-v|^{\alpha'} + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{|t-v|^k}{k!} 2^{j(\alpha-k)} \right] \\
&= K \left[\sum_{j<J} 2^{j\alpha} + |t-v|^{\alpha'} \sum_{j<J} 2^{j(\alpha-\alpha')} + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{|t-v|^k}{k!} \sum_{j<J} 2^{j(\alpha-k)} \right] \\
&\leq K \left[\sum_{j\geq -J} (2^{-\alpha})^j + |t-v|^{\alpha'} \sum_{j\geq -J} (2^{\alpha'-\alpha})^j + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{|t-v|^k}{k!} \sum_{j\geq -J} (2^{k-\alpha})^j \right] \qquad (46)
\end{aligned}
$$

Apply Proposition 5.6 to each of the series in j with $r_\alpha = 2^{-\alpha} < 1$, $r_\alpha = 2^{\alpha'-\alpha} < 1$, and $r_\alpha = 2^{k-\alpha} < 1$, respectively, to obtain:

$$
(46) \leq K \left[2^{J\alpha} + 2^{J(\alpha-\alpha')} |t-v|^{\alpha'} + \sum_{k=0}^{\lfloor \alpha \rfloor} 2^{J(\alpha-k)} \frac{|t-v|^k}{k!} \right] \qquad (47)
$$

Now, since $2^{J-1} \leq |t-v| \leq 2^J$ we have

$$
2^J = 2 \cdot 2^{J-1} \leq 2|t-v|
$$

Plugging this inequality into (47), we have:

$$
\begin{aligned}
(47) &\leq K \left[2^\alpha |t-v|^\alpha + 2^{(\alpha-\alpha')} |t-v|^{\alpha-\alpha'} |t-v|^{\alpha'} + \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{|t-v|^k}{k!} 2^{\alpha-k} |t-v|^{\alpha-k} \right] \\
&\leq K|t-v|^\alpha + |t-v|^\alpha \sum_{k=0}^{\lfloor \alpha \rfloor} \frac{2^{\alpha-k}}{k!} \\
&\leq K|t-v|^\alpha
\end{aligned}
$$

Therefore we have $II_J \leq K|t-v|^\alpha$ for any J. As a result,

$$
\forall\, t \in \mathbb{R}\,, \quad |f(t) - p_v(t)| \leq I_J + II_J \leq K|t-v|^\alpha
$$

which proves that f is Lipschitz α at v. □

Recall that we are utilizing a real valued wavelet $\psi \in \mathbf{C}^n(\mathbb{R})$ with n vanishing moments and with derivatives that have fast decay. We first remark that the proof of Theorem 5.5 can be adapted to give the following theorem, which measures the uniform Lipschitz regularity of f over arbitrary intervals $[a, b]$.

Theorem 5.7. *If $f \in \mathbf{L}^2(\mathbb{R})$ is uniformly Lipschitz $\alpha \leq n$ over $[a, b]$, then there exists $A > 0$ such that*

$$|Wf(u, s)| \leq As^{\alpha+1/2}, \quad \forall\, (u, s) \in [a, b] \times (0, \infty) \tag{48}$$

Conversely, suppose that f is bounded and that $Wf(u, s)$ satisfies (48) for $\alpha < n$ that is not an integer. Then f is uniformly Lipschitz α on $[a + \epsilon, b - \epsilon]$ for any $\epsilon > 0$.

Proof. The proof relies on Theorem 5.5 and modifications of its proof. See pages 211–212 of *A Wavelet Tour of Signal Processing* for the details. □

We now make a few remarks. First, the condition (48) is only meaningful when $s \to 0$, since in general we have

$$|Wf(u, s)| = |\langle f, \psi_{u,s}\rangle| \leq \|f\|_2 \|\psi\|_2$$

which will supersede (48) for large s. Thus the localized regularity of f is measured by zooming in on the points $u \in [a, b]$.

Second, if ψ has exactly n vanishing moments but f is uniformly Lipschitz $\alpha > n$ on $[a, b]$, then $f \in \mathbf{C}^n(a, b)$ and we showed already in (37) that $\lim_{s\to 0} s^{-(n+1/2)} Wf(u, s) = Kf^{(n)}(u)$ with $K \neq 0$. Thus the wavelet coefficients will not decay as $O(s^{\alpha+1/2})$ despite the higher regularity of f.

Finally, for the converse of Theorems 5.5 and 5.7, there is the requirement that $\alpha \notin \mathbb{Z}$. Indeed, the wavelet decay conditions are not sufficient to conclude α-Lipschitz regularity when $\alpha = n \in \mathbb{Z}$. In the case of $[a, b] = \mathbb{R}$ and $\alpha = 1$, the decay (48) is only sufficient to conclude that f is in the Zygmund class, which consists of all bounded, continuous functions for which there exists a constant K such that

$$|f(t + v) + f(t - v) - 2f(t)| \leq K|v|, \quad \forall\, t, v \in \mathbb{R}$$

For more details, see [4, Chapter 6].

Exercise 48. Read Section 6.1.3 of *A Wavelet Tour of Signal Processing.*

Exercise 49. Show that f may be pointwise Lipschitz $\alpha > 1$ at v, while f' is not pointwise Lipschitz $\alpha - 1$ at v. Consider $f(t) = t^2 \cos(1/t)$ at $t = 0$.

Exercise 50. Let $f(t) = |t|^\alpha$. Show that $Wf(u, s) = s^{\alpha+1/2} Wf(u/s, 1)$. Prove that it is not sufficient to measure the decay of $|Wf(u, s)|$ when $s \to 0$ at $u = 0$ in order to compute the Lipschitz regularity of f at $t = 0$.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 16: Wavelet Modulus Maxima

March 10, 2020

Lecturer: Matthew Hirn

5.2 Detection of Singularities via Wavelet Transform Modulus Maxima

Section 6.2.1 of A Wavelet Tour of Signal Processing.

Theorems 5.5 and 5.7 prove that the local regularity of f at v depends on the decay of $|Wf(u,s)|$ as $s \to 0$ in the neighborhood of v. However, it is not necessary to measure this decay in the entire time-scale plane $(u,s) \in \mathbb{R} \times (0,\infty)$. Rather, $|Wf(u,s)|$ can be controlled from its local maxima values.

A wavelet *modulus maximum* is defined as a point (u_0, s_0) such that $|Wf(u,s_0)|$ is locally maximum at $u = u_0$. This implies that

$$\frac{\partial Wf(u_0,s_0)}{\partial u} = 0$$

This local maximum should be a strict local maximum in either the left or right neighborhood of u_0 to avoid having any local maxima when $|Wf(u,s_0)|$ is constant. We call any connected curve $(u, s(u))$ in the time-scale plane along which all points are modulus maxima a maxima line. See Figure 23, which computes the wavelet modulus maxima of the signal from Figure 13 using a wavelet $\psi = -\theta'$.

Compare to the wavelet ridges used to analyze instantaneous frequencies, which were defined as the local maxima of the scalogram $P_W f(u,s) = |Wf(u,s)|^2$. In that case the ridge points followed the instantaneous frequencies of the signal over time; here the wavelet modulus maxima trace back to the isolated singularities at a fixed time.

Recall from Theorem 5.4 that if ψ has exactly n vanishing moments and a fast decay, then there exists θ with fast decay such that

$$\psi = (-1)^n \theta^{(n)}, \quad \widehat{\theta}(0) \neq 0$$

in which case the wavelet transform can be rewritten as

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f * \overline{\theta}_s)(u)$$

If the wavelet ψ has only one vanishing moment, then (with the second equality only if f is differentiable)

$$Wf(u,s) = s(f * \overline{\theta}_s)'(u) = sf' * \overline{\theta}_s(u)$$

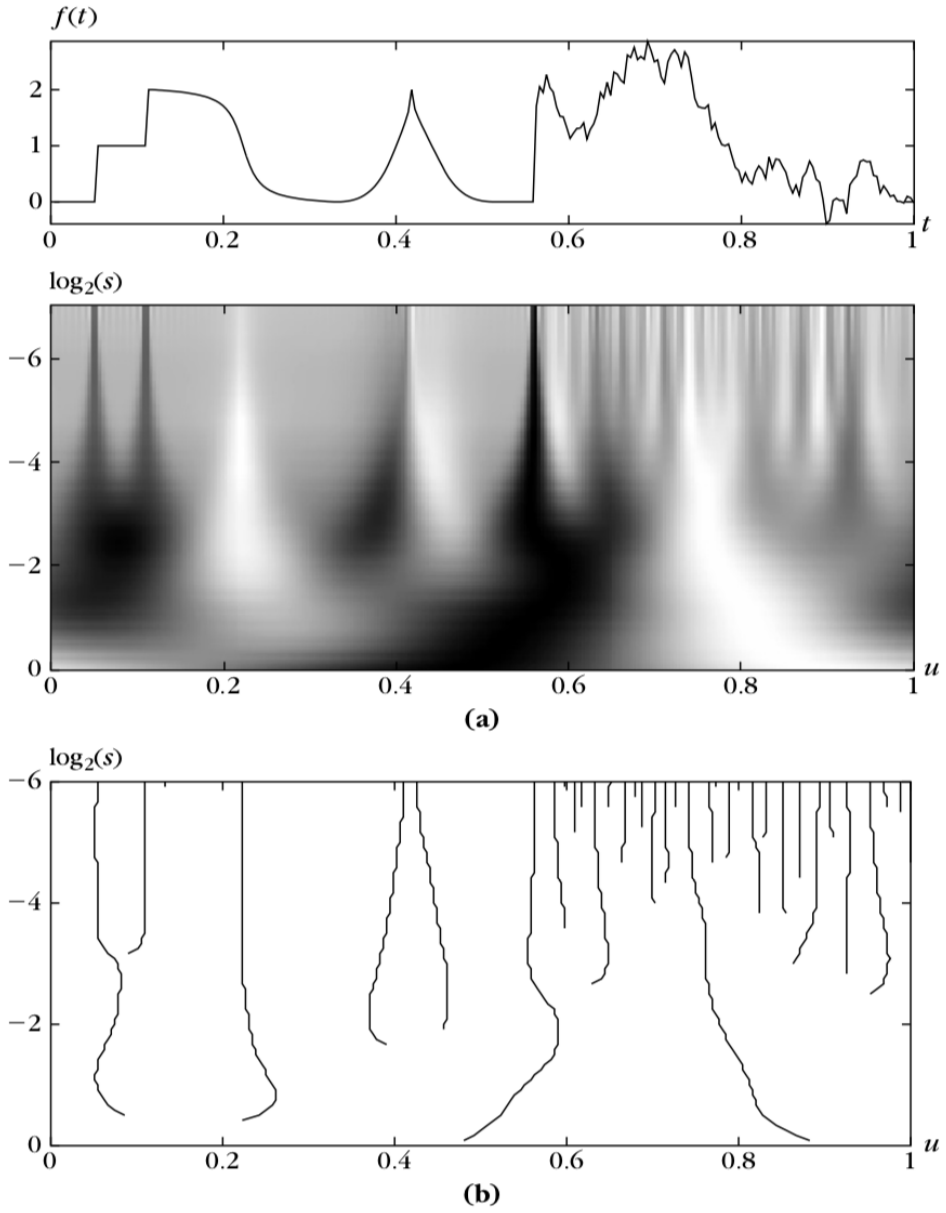


Figure 23: (a) Wavelet transform $Wf(u,s)$; the horizontal and vertical axes give u and $\log_2 s$, respectively. (b) Modulus maxima of $Wf(u,s)$.

and so the wavelet modulus maxima are the maxima of the derivative of f after being smoothed by $\overline{\theta}_s$. These maxima locate discontinuities and edges in images. If the wavelet has two vanishing moments, then the wavelet modulus maxima correspond to high curvatures. See Figure 24 for an illustration. The next theorem proves that if $Wf(u,s)$ has no modulus maxima at fine scales, then f is locally regular.

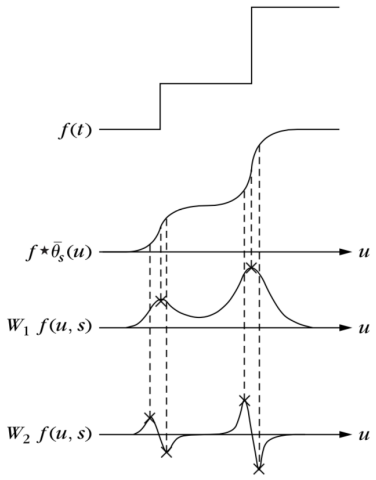


Figure 24: The convolution $f * \overline{\theta}_s(u)$ averages f over a domain proportional to s. If $\psi = -\theta'$, then $W_1 f(u,s) = s\frac{d}{du}(f * \overline{\theta}_s)(u)$ has modulus maxima at sharp variation points of $f * \overline{\theta}_s(u)$. If $\psi = \theta''$, then the modulus maxima of $W_2 f(u,s) = s^2 \frac{d^2}{du^2}(f * \overline{\theta}_s)(u)$ correspond to locally maximum curvatures.

Theorem 5.8. *Suppose that $\psi \in \mathbf{C}^n(\mathbb{R})$ with compact support, and $\psi = (-1)^n \theta^{(n)}$ with $\widehat{\theta}(0) \neq 0$. Let $f \in \mathbf{L}^1[a,b]$. If there exists $s_0 > 0$ such that $|Wf(u,s)|$ has no local maximum for $u \in [a,b]$ and $s < s_0$, then f is uniformly Lipschitz n on $[a+\epsilon, b-\epsilon]$ for any $\epsilon > 0$.*

This theorem implies that f can be singular at a point v (i.e., f is Lipschitz $\alpha < 1$ at v) only if there is a sequence of wavelet maxima points $(u_p, s_p)_{p\in\mathbb{N}}$ that converges to v at fine scales:

$$\lim_{p\to\infty} u_p = v \quad \text{and} \quad \lim_{p\to\infty} s_p = 0$$

These modulus maxima points may or may not be along the same maxima line, however the theorem does guarantee that all singularities are detected by following the wavelet modulus maxima at fine scales. The maxima lines in Figure 23 illustrates the result.

Not all wavelets $\psi = (-1)^n \theta^{(n)}$ guarantee that a modulus maxima located at (u_0, s_0) belongs to a maxima line that propagates toward finer scales. When s decreases, $Wf(u,s)$ may have no more maxima in the neighborhood of $u = u_0$. This leads to spurious wavelet modulus maxima that make it harder to detect isolated singularities. This phenomena cannot occur, though, if θ is a Gaussian. Interestingly, in this case the wavelet transform can be

written as the solution of the heat equation, where s is proportional to the diffusion time, and the maximum principle proves that the maxima may not disappear.

We first recall the heat equation on the interval $[a, b]$ (which physically can be thought of as a rod). Let $g(x, \tau)$ be a function of space $x \in [a, b]$ and time $\tau \in [0, T]$, which measures the temperature of the rod at $x \in [a, b]$ at time τ. We wish to find a g that satisfies the *heat equation*,

$$\begin{aligned} \frac{\partial g}{\partial \tau} &= \frac{\partial^2 g}{\partial x^2} \\ g(x, 0) &= h(x) \end{aligned}$$

for some initial temperature distribution given by $h(x)$. To make this problem well defined we need to specify the boundary condition of g at $x = a, b$, which can be either Dirichlet, meaning that we enforce $g(a, \tau) = g(b, \tau) = 0$, or Neumann, meaning that $\partial_x g(a, \tau) = \partial_x g(b, \tau) = 0$. Since the theorem we are interested in, the maximum principle, holds regardless of boundary condition, we do not impose a specific one here.

For simplicity, suppose that $[a, b] = [0, 1]$, although this is not necessary. Define R to be the space time rectangle

$$R = [0, 1] \times [0, T]$$

and define B to be the boundary of R, not including the "top,"

$$B = \{(x, \tau) \in R : \tau = 0 \text{ or } x = 0 \text{ or } x = 1\}$$

The maximum principle proves that the maximum of $g(x, \tau)$ on R must be attained somewhere on B.

Theorem 5.9. *If $g(x, \tau)$ satisfies the heat equation for $x \in [0, 1]$ and $\tau \in [0, T]$, then the maximum value of g occurs at $\tau = 0$ (at the initial condition) or for $x = 0$ or $x = 1$ (at the ends of the rod). More precisely,*

$$\sup_{(x,\tau) \in R} g(x, \tau) = \sup_{(x,\tau) \in B} g(x, \tau)$$

Proof. Suppose that (x_0, τ_0) is an interior maximum point of g. In this case we must have

$$\partial_\tau g(x_0, \tau_0) = 0$$

Additionally, since maxima occur where the function is concave down, we must have that

$$\partial_{xx} g(x_0, \tau_0) \leq 0$$

The inequality cannot be strict, since g is a solution of the heat equation and therefore

$$\partial_\tau g(x, \tau) = \partial_{xx} g(x, \tau), \quad \forall\, (x, \tau) \in R$$

In particular, it must be that $0 = \partial_\tau(x_0, \tau_0) = \partial_{xx} g(x_0, \tau_0)$.

Now define a new function

$$g_\varepsilon(x,\tau) = g(x,\tau) + \varepsilon x^2, \quad \varepsilon > 0$$

Since g satisfies the heat equation it is easy to see that

$$\partial_\tau g_\varepsilon(x,\tau) - \partial_{xx} g_\varepsilon(x,\tau) = -2\varepsilon < 0 \tag{49}$$

We first show that g_ε cannot achieve its maximum at an interior point (x_0,τ_0). Suppose that it does. By the same reasoning as above, $\partial_\tau g_\varepsilon(x_0,\tau_0) = 0$ and $\partial_{xx} g_\varepsilon(x_0,\tau_0) \leq 0$. But then

$$\partial_\tau g_\varepsilon(x_0,\tau_0) - \partial_{xx} g_\varepsilon(x_0,\tau_0) \geq 0$$

which contradicts (49).

Thus the maximum of g_ε must occur on ∂R, the boundary of R. We now show that it cannot occur when $\tau = T$. Suppose that it does occur at a point (x_0,T). Then again it must be that $\partial_{xx} g_\varepsilon(x_0,T) \leq 0$ and in this case, $\partial_\tau g_\varepsilon(x_0,T) \geq 0$ since

$$\partial_\tau g_\varepsilon(x_0,T) = \lim_{h\to 0^+} \frac{g_\varepsilon(x_0,T) - g_\varepsilon(x_0,T-h)}{h} \geq 0$$

where the inequality follows from the assumption that (x_0,T) is the location of the maximum, which implies that $g_\varepsilon(x_0,T) \geq g_\varepsilon(x_0,T-h)$. But this would again imply that $\partial_\tau g_\varepsilon(x_0,T) - \partial_{xx} g_\varepsilon(x_0,T) \geq 0$, contradicting (49).

Therefore we must have

$$\begin{aligned}
\sup_{(x,\tau)\in R} g_\varepsilon(x,\tau) = \sup_{(x,\tau)\in B} g_\varepsilon(x,\tau) &= \sup_{(x,\tau)\in B} [g(x,\tau) + \varepsilon x^2] \\
&\leq \sup_{(x,\tau)\in B} g(x,\tau) + \sup_{(x,\tau)\in B} \varepsilon x^2 \\
&= \sup_{(x,\tau)\in B} g(x,\tau) + \varepsilon
\end{aligned}$$

On the other hand we also have:

$$\sup_{(x,\tau)\in R} g_\varepsilon(x,\tau) = \sup_{(x,\tau)\in R} [g(x,\tau) + \varepsilon x^2] \geq \sup_{(x,\tau)\in R} g(x,\tau)$$

It follows that

$$\sup_{(x,\tau)\in R} g(x,\tau) \leq \sup_{(x,\tau)\in B} g(x,\tau) + \varepsilon, \quad \forall\, \varepsilon > 0$$

Letting $\varepsilon \to 0$, we obtain

$$\sup_{(x,\tau)\in R} g(x,\tau) \leq \sup_{(x,\tau)\in B} g(x,\tau)$$

But since g is continuous and $B \subset R$, we must have equality. □

Now we can prove the following theorem:

Theorem 5.10. *Let $\psi = (-1)^n \theta^{(n)}$, where θ is a Gaussian. For any $f \in \mathbf{L}^2(\mathbb{R})$, the modulus maxima of $Wf(u,s)$ belong to connected curves that are never interrupted as the scale decreases.*

Proof. To simplify the proof we use a normalized Gaussian

$$\theta(t) = \frac{1}{2\sqrt{\pi}} e^{-t^2/4} \Longrightarrow \widehat{\theta}(\omega) = e^{-\omega^2}$$

Note as well that $\overline{\theta}_s = \theta_s$. Let $f^{(n)}$ be the n^{th} derivative of f, which is defined in the sense of distributions if $f^{(n)}$ is not defined otherwise. Theorem 5.4 proves that

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f * \theta_s)(u) = s^n f^{(n)} * \theta_s(u)$$

Consider the heat equation

$$\partial_\tau g(u,\tau) = \partial_{xx} g(u,\tau), \quad g(u,0) = h(u)$$

We can compute the solution by taking the Fourier transform of both sides with respect to u:

$$\partial_\tau \widehat{g}(\omega,\tau) = -\omega^2 \widehat{g}(\omega,\tau)$$

It follows that

$$\widehat{g}(\omega,\tau) = \widehat{h}(\omega) e^{-\tau\omega^2}$$

from which we derive:

$$g(u,\tau) = \frac{1}{\sqrt{\tau}} h * \theta_{\sqrt{\tau}}(u)$$

If we set $\tau = s^2$ and $h = f^{(n)}$, then we obtain

$$s^{n+1} g(u,s^2) = s^n f^{(n)} * \theta_s(u) = Wf(u,s) \tag{50}$$

Thus the wavelet transform is proportional to the solution of the heat equation with initial temperature distribution given by $f^{(n)}$.

Recall that a modulus maxima of $Wf(u,s)$ at (u_0,s_0) is a local maxima of $|Wf(u,s)|$ for a fixed s and varying u, which due to (50) corresponds to a local maxima of $|g(u,s^2)|$ for a fixed s and varying u. Suppose that a curve of wavelet modulus maxima is interrupted at (u_1,s_1) with $s_1 > 0$. Then one can show there exists $\varepsilon > 0$ such that on the domain $[u_1-\varepsilon, u_1+\varepsilon] \times [s_1-\varepsilon, s_1]$, the global maximum of $|g(u,s^2)|$ is at (u_1,s_1). However, this contradicts the maximum principle, and so the curve cannot be interrupted. □

Exercise 51. Let $f(t) = \cos(\omega_0 t)$ and $\psi(t)$ a wavelet that is even.

(a) Verify that $Wf(u,s) = \sqrt{s}\widehat{\psi}(s\omega_0)\cos(\omega_0 u)$.

(b) Find the equations of the curves of wavelet modulus maxima in the time-scale plane (u, s). Relate the decay of $|Wf(u, s)|$ along these curves to the number n of vanishing moments of ψ.

Exercise 52. Prove that if $f(t) = \mathbf{1}_{[0,+\infty)}(t)$ then the number of modulus maxima of $Wf(u, s)$ at each scale s is larger than or equal to the number of vanishing moments of ψ.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 17: Estimating Pointwise Regularity

March 12, 2020

Lecturer: Matthew Hirn

A wavelet transform, even with $\psi = (-1)^n \theta^{(n)}$ for θ 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 α 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 (38) and (39) of Theorem 5.5, suppose that for $s < s_0$ all modulus maxima that converge to v are included in a cone $\mathcal{C}_v$ defined as:

$$\mathcal{C}_v = \{(u, s) \in \mathbb{R} \times (0, \infty) : |u - v| \leq Cs\}$$

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 $v = 0$).

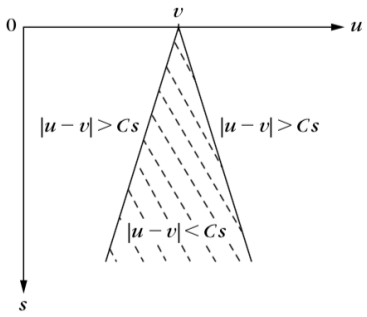


Figure 25: The cone of influence $\mathcal{C}_v$ of an abscissa v consists of the time-scale points (u, s)

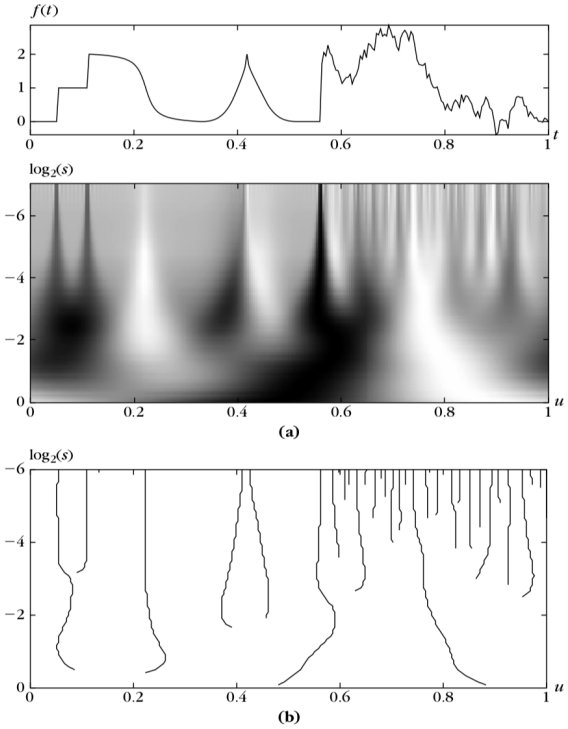
Within the cone $\mathcal{C}_v$ we have $|u - v|/s \leq C$, and so the conditions (38) and (39) of Theorem 5.5 can be written for these points as:

$$|Wf(u, s)| \leq A' s^{\alpha + 1/2}, \quad \forall (u, s) \in \mathcal{C}_v$$

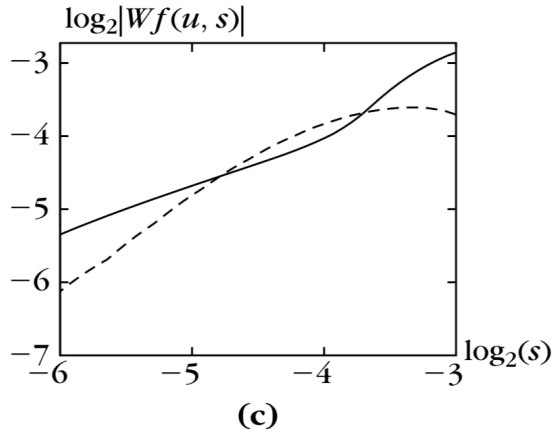
This is equivalent to:

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

Thus the Lipschitz regularity at v can be estimated by computing 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.



(a) Figure 23 revisited



(b) $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ along two maxima lines.

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 $1/2$.

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 of rapid transition will technically be $\mathbf{C}^\infty$. 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

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

where

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

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

Theorem 5.11. *Let $\psi = (-1)^n \theta^{(n)}$ with*

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

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

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

Proof. Using Theorem 5.4 we write the wavelet transform as:

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

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

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

Therefore we can rewrite the wavelet transform as

$$\begin{aligned} Wf(u,s) &= s^n \sqrt{\frac{s}{s_0}} \frac{d^n}{du^n}(f_0 * \theta_{s_0})(u) \\ &= \left(\frac{s}{s_0}\right)^{n+1/2} s_0^n \frac{d^n}{du^n}(f_0 * \theta_{s_0})(u) \\ &= \left(\frac{s}{s_0}\right)^{n+1/2} Wf_0(u, s_0) \end{aligned}$$

Since f_0 is uniformly Lipschitz α on $[v-\varepsilon, v+\varepsilon]$, Theorem 5.7 proves that there exists $A > 0$ such that

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

Therefore,

$$\begin{aligned}
|Wf(u,s)| &\leq \left(\frac{s}{s_0}\right)^{n+1/2} |Wf_0(u,s_0)| \\
&\leq \left(\frac{s}{s_0}\right)^{n+1/2} A s_0^{\alpha+1/2} \\
&= A s^{n+1/2} s_0^{-(n-\alpha)} \\
&= A s^{n+1/2} \left(s^2 + \frac{\sigma^2}{\beta^2}\right)^{-(n-\alpha)/2} \\
&= A s^{\alpha+1/2} \left(1 + \frac{\sigma^2}{\beta^2 s^2}\right)^{-(n-\alpha)/2}
\end{aligned}$$

□

This theorem relates the wavelet transform decay expected by the Lipschitz α 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 $\mathbf{C}^\infty$, and thus the decay at fine scales will necessarily be limited by the finite number of vanishing moments. Figure 27 gives an example.

Exercise 53. Read Section 6.2.1 of *A Wavelet Tour of Signal Processing.*

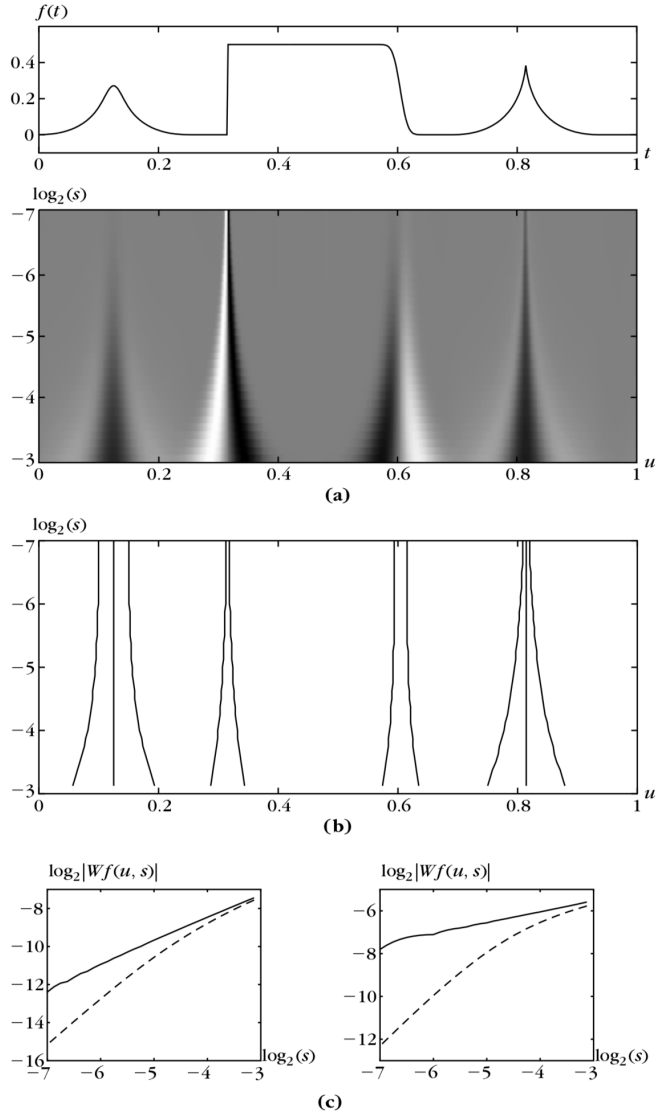


Figure 27: Top: Signal with two types of singularities, a jump discontinuity at $t = 0.35$ and a cusp at $t = 0.81$. Blurred versions of the same singularities are located at $t = 0.60$ and $t = 0.12$, respectively. (a) The wavelet transform $Wf(u, s)$ using a wavelet $\psi = \theta''$, where θ is a 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 $s \leq \sigma = 2^{-5}$.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 18 & 19: Stochastic Processes

March 19 & 24, 2020

Lecturer: Matthew Hirn

5.3 Stochastic Processes

References for this section are:

1. *Wavelet Tour of Signal Processing* [1, Section 6.4]: This covers multi-fractals, of which fractional Brownian motion is an example. We will cover fractional Brownian motion, but not general multi-fractals.

2. *Stochastic Calculus for Finance II* [6]: Chapter 1 and 2 are good references for the basics of measure theoretic probability. Chapter 3 is a good reference for constructing the Wiener process and understanding its properties.

3. *Introduction to random fields and scale invariance* [7]: Some additional good information. I am getting the plots from these notes. Even though the focus is on random fields (that is, stochastic processes in which the index variable $t \in \mathbb{R}^d$), there is good info on stochastic processes as well.

The right hand side of the signal in Figure 23 can be modeled as a stochastic process. Many phenomena of interest can be modeled as stochastic processes that are singular almost everywhere, e.g., financial instruments (stocks), heart records, and textures. Knowing the distribution of singularities is important for analyzing the properties of such processes. However, pointwise measurements are not possible because the singularities are not isolated. If the stochastic process is also self-similar, though, wavelet transforms and in particular wavelet zoom through the layers of self-similarity can extract information about the distribution of singularities. We illustrate this concept on fractional Brownian motions, which are statistically singular almost everywhere with the same type of singularity, specified by its Hurst parameter H.

Recall a probability space consists of three things: (i) the set of all outcomes Ω; (ii) the set of all events $\mathcal{F}$, which is a set of sets, and in which each set $A \in \mathcal{F}$ is a subset $A \subseteq \Omega$. We will require that $\mathcal{F}$ be a σ-algebra, meaning that (a) $\emptyset \in \mathcal{F}$; (b) if $A \in \mathcal{F}$, then the complement of A, denoted A^c, is also in $\mathcal{F}$; and (c) if $A_1, A_2, \ldots \in \mathcal{F}$ then $\cup_{i \geq 1} A_i \in \mathcal{F}$. And finally (iii) a probability measure $\mathbb{P} : \mathcal{F} \rightarrow [0,1]$ that assigns each $A \in \mathcal{F}$ a probability $\mathbb{P}(A)$. The probability measure must satisfy (a) $\mathbb{P}(\Omega) = 1$; and (b) if $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then

$$\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(A_i)$$

The Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest σ-algebra on $\mathbb{R}$ that contains all intervals. Now we can define a random variable.

Definition 5.12. We say X is a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$X : \Omega \to \mathbb{R}$$

and

$$\forall\, B \in \mathcal{B}\,, \quad \{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

Definition 5.13. The distribution of a random variable X is the probability measure $\mu_X : \mathcal{B} \to [0, 1]$ defined as

$$\mu_X(B) = \mathbb{P}(X \in B)$$

Definition 5.14. A real valued stochastic process $X = (X(t))_{t\in\mathbb{R}}$ is a family of random variables

$$X(t) : \Omega \to \mathbb{R}$$

defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

We will assume that all of our stochastic processes are real valued.

Definition 5.15. The distribution of a stochastic process X is given by all its finite dimensional distributions, that is, the distribution of all real random vectors

$$(X(t_1), \ldots, X(t_d))\,, \quad \forall\, d \geq 1\,, \ \ \forall\, t_1, t_2, \ldots, t_d$$

Definition 5.16. A stochastic process X is a second order process if $\mathbb{E}[X(t)^2] < +\infty$ for all $t \in \mathbb{R}$. In this case we may define its:

- Mean function

  $$m_X(t) = \mathbb{E}[X(t)] = \int_\Omega X(t)(\omega)\, d\mathbb{P}(\omega)$$

  The process is centered if $m_X(t) = 0$ for all $t \in \mathbb{R}$. Also note that, unfortunately, the standard notation in harmonic analysis for the frequency variable is ω, but the standard notation in probability for an outcome is also ω. Hopefully the context will always be clear and things will not be too confusing.

- Covariance function

  $$\mathrm{Cov}_X(s, t) = \mathrm{Cov}(X(s), X(t)) = \mathbb{E}[(X(s) - \mathbb{E}[X(s)])(X(t) - \mathbb{E}[X(t)])]$$

  Note that the variance is given by

  $$\mathrm{Var}_X(t) = \mathrm{Var}(X(t)) = \mathrm{Cov}(X(t), X(t)) = \mathbb{E}[(X(t) - [X(t)])^2]$$

Definition 5.17. A stochastic process X is Gaussian if for all $t_1, t_2, \ldots, t_d$ the probability distribution of the random vector $(X(t_1), \ldots, X(t_d)) \in \mathbb{R}^d$ is normally distributed.

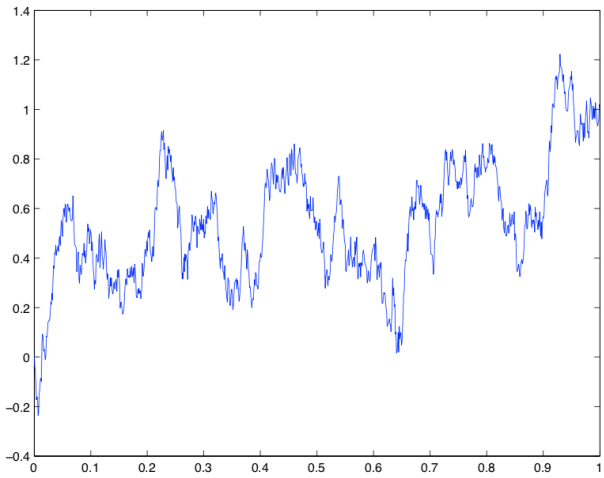


Figure 28: A sample path (realization) of the Wiener process (Brownian motion).

From the properties of the normal distribution it follows that the probability distribution of a Gaussian process is entirely determined by the mean function $m_X(t)$ and the covariance function $\mathrm{Cov}_X(s,t)$. A very useful and famous example of a Gaussian process is the Wiener process (also referred to as Brownian motion), which has wide use in physics and finance and other fields. Let us denote it by $W(t)$. The Wiener process satisfies the following conditions:

- W is a Gaussian process with $W(0) = 0$
- $W(t)$ is continous in t
- $m_W(t) = \mathbb{E}[W(t)] = 0$ for all $t \in \mathbb{R}$
- $\mathrm{Cov}_W(s,t) = \frac{1}{2}(|s| + |t| - |t-s|) = \min(|s|, |t|)$ for all $s, t \in \mathbb{R}$

Figure 28 plots a sample path (that is, a realization) of the Wiener process.

Let us describe now how to construct the Wiener process. It will give us some intuition about stochastic processes in general. The main idea is that we are going to construct a random walk out of an infinite sequence of coin flips. We will then let these coin flips happen with increasing frequency, until in the limit there is no time between the flips. Let us now be more precise.

First consider the experiment of flipping a coin once. There are two possible outcomes, heads or tails, and our probability space is the following:

$$\begin{aligned} \Omega_1 &= \{H, T\} \\ \mathcal{F}_1 &= \{\emptyset, \{H\}, \{T\}, \Omega_2\} \\ \mathbb{P}_1(H) &= p \\ \mathbb{P}_1(T) &= q = 1 - p \end{aligned}$$

Note that $\mathbb{P}_1(\emptyset) = 0$ and $\mathbb{P}_1(\Omega_1) = 1$ by the properties of probability measures. We define a random variable on $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ that takes the value $+1$ for the outcome heads, and -1 for the outcome tails:

$$Z(\omega) = \begin{cases} +1 & \omega = H \\ -1 & \omega = T \end{cases} \tag{51}$$

Now let us consider the experiment of flipping a coin infinitely many times, in which all the flips are independent. In this case our set of outcomes is:

$$\Omega_\infty = \text{ all infinite sequences of heads (H) and tails (T)}$$

This outcome space is uncountably infinite, so more care is needed in defining its σ-algebra $\mathcal{F}_\infty$ and its probability measure $\mathbb{P}_\infty$. We will do so by specifying the probability of all events that are based on a finite number of coin tosses. Note that an outcome $\omega \in \Omega_\infty$ can be written as:

$$\omega = \omega_1 \omega_2 \omega_3 \dots$$

where each $\omega_i \in \{H, T\}$. Now let us build up $\mathcal{F}_\infty$. We know we have to put $\emptyset, \Omega_\infty \in \mathcal{F}_\infty$ with $\mathbb{P}_\infty(\emptyset) = 0$ and $\mathbb{P}_\infty(\Omega_\infty) = 1$. Now let us also add in the two events:

$$A_H = \{\omega \in \Omega_\infty : \omega_1 = H\} = \text{ first coin is a heads}$$
$$A_T = \{\omega \in \Omega_\infty : \omega_1 = T\} = \text{ first coin is a tails}$$

Based on the single coin toss probability space, we set

$$\mathbb{P}_\infty(A_H) = p \quad \text{and} \quad \mathbb{P}_\infty(A_T) = q$$

Remember that if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, but in this case $A_H^c = A_T$ so we are okay. Also note the union is $A_T \cup A_H = \Omega_\infty$. Now we add in events based on the first two coin tosses, where the definitions of these events should be clear:

$$A_{HH}\ ,\ A_{HT}\ ,\ A_{TH}\ ,\ A_{TT}$$

We set the probabilities accordingly:

$$\mathbb{P}_\infty(A_{HH}) = p^2$$
$$\mathbb{P}_\infty(A_{HT}) = pq$$
$$\mathbb{P}_\infty(A_{TH}) = qp$$
$$\mathbb{P}_\infty(A_{TT}) = q^2$$

Now we have to take these four new events, and also consider their compliments and unions, and also add those events into $\mathcal{F}_\infty$, and specify their probabilities. This can be done. Then we continue by considering events based on the first three coin tosses, then the first four coins tosses, and so on, adding everything into $\mathcal{F}_\infty$ along with unions and compliments, and specifying probabilities. We do this for all events which are based on the first k coin tosses,

for all $k \in \mathbb{N}$. Then we complete $\mathcal{F}_\infty$ by adding in the minimal number of all other events required to have a σ-algebra.

Now we define the random walk, which is a discrete stochastic process defined over the probability space $(\Omega_\infty, \mathcal{F}_\infty, \mathbb{P}_\infty)$. We set $p = q = 1/2$, so the coin if fair, which will make the walk unbiased. Let $M = (M(n))_{n\in\mathbb{N}_0}$ be the random walk, where $\mathbb{N}_0 = \{0, 1, 2, \dots, \}$, and where each $M(n)$ is a random variable on $(\Omega_\infty, \mathcal{F}_\infty, \mathbb{P}_\infty)$. Define

$$M(0)(\omega) = 0\,, \quad \forall\, \omega \in \Omega_\infty$$

That is to say, our random walk will always start at zero. For the remaining steps, recall that $\omega = \omega_1\omega_2\omega_3\dots$ is an outcome in Ω_∞. Let $Z_i(\omega_i)$ be defined as in (51) for each coin flip ω_i. Define $M(n)$ for every $n \in \mathbb{N}$ as

$$M(n)(\omega) = \sum_{i=1}^{n} Z_i(\omega_i)\,, \quad \omega = \omega_1\omega_2\omega_3\dots$$

The Wiener process on (Brownian motion) $[0, \infty)$ is obtained by rescaling the random walk M. Define $W^{(m)} = (W^{(m)}(t))_{t\in[0,\infty)}$ as

$$W^{(m)}(t) = \begin{cases} (1/\sqrt{m})M(mt) & mt \in \mathbb{N}_0 \\ (1/\sqrt{m})[(\lceil mt\rceil - mt)M(\lfloor mt\rfloor) + (mt - \lfloor mt\rfloor)M(\lceil mt\rceil)] & mt \notin \mathbb{N}_0 \end{cases}$$

We then obtain $W = (W(t))_{t\in[0,\infty)}$ by taking $m \to \infty$, that is

$$W(t) = \lim_{m\to\infty} W^{(m)}(t)$$

To obtain a Wiener process on $\mathbb{R}$, we take two independent Wiener processes $W_1 = (W_1(t))_{t\in[0,\infty)}$ and $W_2 = (W_2(t))_{t\in[0,\infty)}$ and we create one on $\mathbb{R}$ by setting:

$$W(t) = \begin{cases} W_1(t) & t \geq 0 \\ W_2(-t) & t < 0 \end{cases}$$

The Wiener process inherits the properties of the random walk. In particular, $W(0) = 0$ and for all $t_0 < t_1 < t_2 < \cdots t_k$ the increments

$$W(t_1) - W(t_0)\,,\, W(t_2) - W(t_1)\,, \dots,\, W(t_k) - W(t_{k-1})$$

are independent and each increment is normally distributed with

$$\begin{aligned} \mathbb{E}[W(t_{i+1}) - W(t_i)] &= 0 \\ \mathrm{Var}(W(t_{i+1}) - W(t_i)) &= t_{i+1} - t_i \end{aligned}$$

The other properties in our original definition also follow from this construction.

Let us now consider another important class of stochastic processes.

Definition 5.18. A stochastic process X is stationary if, for all $u \in \mathbb{R}$, the stochastic process $(X(t+u))_{t\in\mathbb{R}}$ has the same distribution as $X = (X(t))_{t\in\mathbb{R}}$.

The Wiener process is not stationary, but its increments are. We will come back to this point later. For now, we remark that stationary processes are translation invariant since their distribution does not change with a temporal translation by u. Their statistics inherit this invariance, as the following proposition illustrates.

Proposition 5.19. *If a second order stochastic process X is stationary, then*

- *Its mean function is constant, that is $m_X(t) = m_X$ for some constant value m_X. We will sometimes write $m_X = \mathbb{E}[X]$.*
- *Its covariance function only depends on $t - s$, that is*

$$\mathrm{Cov}_X(s,t) = R_X(t-s)$$

for some even function $R_X : \mathbb{R} \to \mathbb{R}$. The function R_X also satisfies $R_X(0) \geq 0$ and $|R_X(\tau)| \leq R_X(0)$ for all $\tau \in \mathbb{R}$.

Proof. By the stationarity of X we have $X(t) \overset{d}{=} X(0)$ (that is, $X(t)$ and $X(0)$ have the same distribution) and so $m_X(t) = \mathbb{E}[X(t)] = \mathbb{E}[X(0)] = m_X(0)$ for all $t \in \mathbb{R}$. For the covariance set $R_X(\tau) = \mathrm{Cov}_X(0,\tau)$. For any $s \in \mathbb{R}$ we have, again by the stationarity of X, that $(X(s), X(\tau+s)) \overset{d}{=} (X(0), X(\tau))$ and so $\mathrm{Cov}_X(s, \tau+s) = \mathrm{Cov}(0,\tau) = R_X(\tau)$. Hence for $\tau = t - s$ we have $\mathrm{Cov}_X(s,t) = R_X(t-s)$. Since $(X(0), X(\tau)) \overset{d}{=} (X(-\tau), X(0))$ we have

$$R_X(\tau) = \mathrm{Cov}_X(0,\tau) = \mathrm{Cov}_X(-\tau,0) = \mathrm{Cov}_X(0,-\tau) = R_X(-\tau)$$

and so R_X is even. We also have

$$R_X(0) = \mathrm{Cov}_X(0,0) = \mathrm{Var}(X(0)) \geq 0$$

Finally, using the Cauchy-Schwarz inequality and the stationarity of X:

$$\begin{aligned} |R_X(\tau)| = |\mathrm{Cov}(X(0), X(\tau))| &\leq \sqrt{\mathrm{Var}(X(0))\mathrm{Var}(X(\tau))} \\ &= \sqrt{\mathrm{Var}(X(0))\mathrm{Var}(X(0))} \\ &= \mathrm{Var}(X(0)) \\ &= R_X(0) \end{aligned}$$

□

Examples of stationary Gaussian processes are given by Ornstein Uhlenbeck processes, which are defined for any $\theta > 0$ as:

$$X(t) = e^{-\theta t} W(e^{2\theta t})$$

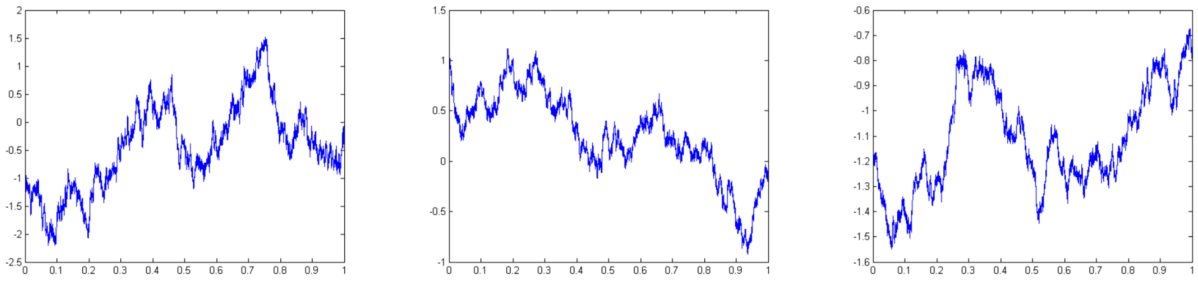


Figure 29: Sample paths of the Ornstein Uhlenbeck process for different values of θ.

where W is the Wiener process. It is clear $\mathbb{E}[X(t)] = 0$ for all $t \in \mathbb{R}$. Also a short calculation shows its covariance is

$$\mathrm{Cov}_X(s,t) = e^{-\theta|t-s|}$$

and thus only depends on $t - s$. Figure 29 plots Ornstein Uhlenbeck processes for different values of θ. Using the previous proposition, we can define the power spectral density (power spectrum) of a stationary process X.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 20: Time-Frequency Analysis of Stationary Processes

March 26, 2020

Lecturer: Matthew Hirn

Definition 5.20. The power spectral density of a second order stationary process X is the Fourier transform of $R_X(\tau)$, that is, $\widehat{R}_X(\omega)$.

For a stationary process X, the function $R_X(\tau) = \mathrm{Cov}_X(0, \tau)$ measures the variability of random fluctuations of X over time. The power spectral density organizes the total variability of X over all times into different frequency components. A time frequency transforms of a stationary process allow us to measure the variability of X within time-frequency Heisenberg boxes. For example, the wavelet coefficients of X define a family of new stochastic processes $X * \psi_s$, indexed by the scale parameter $s > 0$, which are defined as

$$WX(u, s) = X * \psi_s(u) = \int_{\mathbb{R}} X(t)\psi_s(u - t)\, dt$$

We assume $\psi_s(t)$ is continuous, real valued, and compactly supported. Note the integral of a stochastic process with continous sample paths times a continuous deterministic function $f(t)$, over a finite integral, is a random variable defined using the Riemann integral:

$$\int_a^b X(t)f(t)\, dt = \lim_{n\to\infty} \sum_{i=0}^{n-1} X(t_i^n)f(t_i^n)(t_{i+1}^n - t_i^n)$$

where $a = t_0^n < t_1^n < \cdots < t_{n-1}^n < t_n^n = b$ for all n, and $\delta_n = \max_{0\le i\le n-1} |t_{i+1} - t_i| \to 0$ as $n \to \infty$. The new stochastic process $X * \psi_s = (X * \psi_s(u))_{u\in\mathbb{R}}$ retains only the fluctuations of X at the scale s, for each time u; smaller and larger scale fluctuations are eliminated because the wavelet ψ_s has a frequency support essentially supported in a frequency band determined by the scale s. The next proposition encodes this statement more precisely.

Theorem 5.21. *Let X be a second order stationary process with continuous sample paths and with mean zero, i.e., $\mathbb{E}[X] = 0$, and let ψ be a continuous real valued wavelet with compact support. Then $X * \psi_s$ is a stationary process for each $s > 0$ and:*

$$\widehat{R}_{X*\psi_s}(\omega) = \widehat{R}_X(\omega)|\widehat{\psi}_s(\omega)|^2 = s|\widehat{\psi}(s\omega)|^2\widehat{R}_X(\omega) \tag{52}$$

Proof. The fact that $X*\psi_s$ is stationary is straightforward. We also note that since $\mathbb{E}[X] = 0$, we also have $\mathbb{E}[X * \psi_s] = 0$ for each $s > 0$. Thirdly, if $R_X \in \mathbf{L}^1(\mathbb{R})$ then $R_{X*\psi_s} \in \mathbf{L}^1(\mathbb{R})$;

indeed:

$$\begin{aligned}
\int_{\mathbb{R}} |R_{X*\psi_s}(\tau)|\,d\tau &= \int_{\mathbb{R}} |\mathbb{E}[X*\psi_s(0)X*\psi_s(\tau)]|\,d\tau \\
&= \int_{\mathbb{R}} \left|\mathbb{E}\left[\int_{\mathbb{R}} X(u)\psi_s(-u)\,du \cdot \int_{\mathbb{R}} X(v)\psi_s(\tau-v)\,dv\right]\right| d\tau \\
&= \int_{\mathbb{R}} \left|\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{E}[X(u)X(v)]\psi_s(-u)\psi_s(\tau-v)\,du\,dv\right| d\tau \\
&= \int_{\mathbb{R}} \left|\int_{\mathbb{R}}\int_{\mathbb{R}} R_X(u-v)\psi_s(-u)\psi_s(\tau-v)\,du\,dv\right| d\tau \qquad (\text{CoV: } t=u-v) \\
&= \int_{\mathbb{R}} \left|\int_{\mathbb{R}}\int_{\mathbb{R}} R_X(t)\psi_s(-(t+v))\psi_s(\tau-v)\,dt\,dv\right| d\tau \\
&= \int_{\mathbb{R}} \left|\int_{\mathbb{R}} \psi_s(\tau-v)\int_{\mathbb{R}} R_X(t)\psi_s(-v-t)\,dt\,dv\right| d\tau \\
&= \int_{\mathbb{R}} \left|\int_{\mathbb{R}} \psi_s(\tau-v)R_X*\psi_s(-v)\,dv\right| d\tau \\
&\leq \int_{\mathbb{R}}\int_{\mathbb{R}} |\psi_s(\tau-v)R_X*\psi_s(-v)|\,dv\,d\tau \\
&= \int_{\mathbb{R}} |R_X*\psi_s(-v)| \int_{\mathbb{R}} |\psi_s(\tau-v)|\,d\tau\,dv \\
&= \|\psi_s\|_1 \|R_X*\psi_s\|_1 \\
&\leq \|R_X\|_1 \|\psi_s\|_1^2
\end{aligned}$$

Since ψ is continuous and compactly supported, it is in $\mathbf{L}^1(\mathbb{R})$ and so the bound is finite, and $R_{X*\psi_s} \in \mathbf{L}^1(\mathbb{R})$.

Now let us prove (52). Many of the steps are the same as above.

$$\begin{aligned}
\widehat{R}_{X*\psi_s}(\omega) &= \int_{\mathbb{R}} R_{X*\psi_s}(\tau) e^{-i\omega\tau}\, d\tau \\
&= \int_{\mathbb{R}} \mathbb{E}[X*\psi_s(0) X*\psi_s(\tau)] e^{-i\omega\tau}\, d\tau \\
&= \int_{\mathbb{R}} \mathbb{E}\left[\int_{\mathbb{R}} X(u)\psi_s(-u)\, du \cdot \int_{\mathbb{R}} X(v)\psi_s(\tau - v)\, dv\right] e^{-i\omega\tau}\, d\tau \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{E}[X(u)X(v)]\psi_s(-u)\psi_s(\tau - v)\, du\, dv\right] e^{-i\omega\tau}\, d\tau \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}}\int_{\mathbb{R}} R_X(u - v)\psi_s(-u)\psi_s(\tau - v)\, du\, dv\right] e^{-i\omega\tau}\, d\tau \qquad (\text{CoV: } t = u - v) \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}}\int_{\mathbb{R}} R_X(t)\psi_s(-(t + v))\psi_s(\tau - v)\, dt\, dv\right] e^{-i\omega\tau}\, d\tau \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \psi_s(\tau - v) \int_{\mathbb{R}} R_X(t)\psi_s(-v - t)\, dt\, dv\right] e^{-i\omega\tau}\, d\tau \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \psi_s(\tau - v) R_X * \psi_s(-v)\, dv\right] e^{-i\omega\tau}\, d\tau \\
&= \int_{\mathbb{R}} R_X * \psi_s(-v) \int_{\mathbb{R}} \psi_s(\tau - v) e^{-i\omega\tau}\, d\tau\, dv \\
&= \widehat{\psi}_s(\omega) \int_{\mathbb{R}} R_X * \psi_s(-v) e^{-i\omega v}\, dv \\
&= \widehat{\psi}_s(\omega)\widehat{R}_x^*(\omega)\widehat{\psi}_s^*(\omega) \\
&= \widehat{R}_X(\omega)|\widehat{\psi}_s(\omega)|^2
\end{aligned}$$

where the last equality follows from recalling that $R_X(\tau)$ is an even function, and hence its Fourier transform is real valued. □

Stationarity is a pretty strict assumption, and as mentioned, does not include the Wiener process. The notion of a stochastic process with stationary increments relaxes this requirement and includes a much larger number of stochastic processes.

Definition 5.22. A stochastic process X has stationary increments if, for all $u \in \mathbb{R}$, the stochastic process $(X(t + u) - X(u))_{t\in\mathbb{R}}$ has the same distribution as $(X(t) - X(0))_{t\in\mathbb{R}}$.

Stochastic processes with stationary increments include many more processes than just stationary processes, which allows us to model a wider variety of phenomena. Note, in particular, if X has stationary increments then the mean and variance of an increment depends only on the length of the increment, not where it started. That is for any $u \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[X(t + u) - X(u)] &= \mathbb{E}[X(t) - X(0)] \\
\text{Var}(X(t + u) - X(u)) &= \text{Var}(X(t) - X(0))
\end{aligned}$$

An example of a stochastic process with stationary increments is the Wiener process.

Theorem 5.23. *The Wiener process, W, has stationary increments.*

Proof. Define the stochastic process $(\widetilde{W}(t))_{t\in\mathbb{R}}$ as

$$\widetilde{W}(t) = W(t+u) - W(u)$$

where $u \in \mathbb{R}$ is fixed but arbitrary. Our goal is to show distribution of $\widetilde{W}$ does not depend on u, which would mean that W has stationary increments. We first note the Wiener process, W, is a Gaussian process, and thus so is $\widetilde{W}$. Therefore, if we can show the mean function and the covariance function of $\widetilde{W}$ do not depend on u then we are finished. For the mean function we have

$$m_{\widetilde{W}}(t) = \mathbb{E}[\widetilde{W}(t)] = \mathbb{E}[W(t+u)] - \mathbb{E}[W(u)] = 0 - 0 = 0$$

which is obviously independent of u. For the covariance function we have:

$$\begin{aligned}
&2\mathrm{Cov}_{\widetilde{W}}(s,t) \\
&\quad = 2\mathbb{E}[\widetilde{W}(s)\widetilde{W}(t)] \\
&\quad = 2\mathbb{E}[(W(s+u) - W(u))(W(t+u) - W(u))] \\
&\quad = 2\mathbb{E}[W(s+u)W(t+u)] + 2\mathbb{E}[W(u)^2] - 2\mathbb{E}[W(u)W(s+u)] - 2\mathbb{E}[W(u)W(t+u)] \\
&\quad = |s+u| + |t+u| - |t-s| + |u| + |u| - |u| - |u+s| + |s| - |u| - |t-u| + |t| \\
&\quad = |t| + |s| - |t-s|
\end{aligned}$$

which is also independent of u. □

Fractional Brownian motion [7, 8] is a generalization of Brownian motion (i.e., the Wiener process). It depends on a parameter H, which is called the Hurst parameter.

Definition 5.24. A stochastic process $B_H = (B_H(t))_{t\in\mathbb{R}}$ is called a fractional Brownian motion (fBm) with Hurst parameter $H \in (0,1)$ if it satisfies the following:

- B_H is a Gaussian process with $B_H(0) = 0$
- $B_H(t)$ is continuous in t
- $m_{B_H}(t) = \mathbb{E}[B_H(t)] = 0$ for all $t \in \mathbb{R}$
- $\mathrm{Cov}_{B_H}(s,t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |t-s|^{2H})$ for all $s,t \in \mathbb{R}$.

Notice that when $H = 1/2$ we obtain regular Brownian motion, i.e., the Wiener process. Figure 30 plots three sample paths of fBm for $H = 0.75$, while Figure 31 plots sample paths of fBm for $H = 0.15, 0.55, 0.95$.

First note, that like the Wiener process, fractional Brownian motion has stationary increments for any $H \in (0,1)$. Indeed, the proof is essentially identical.

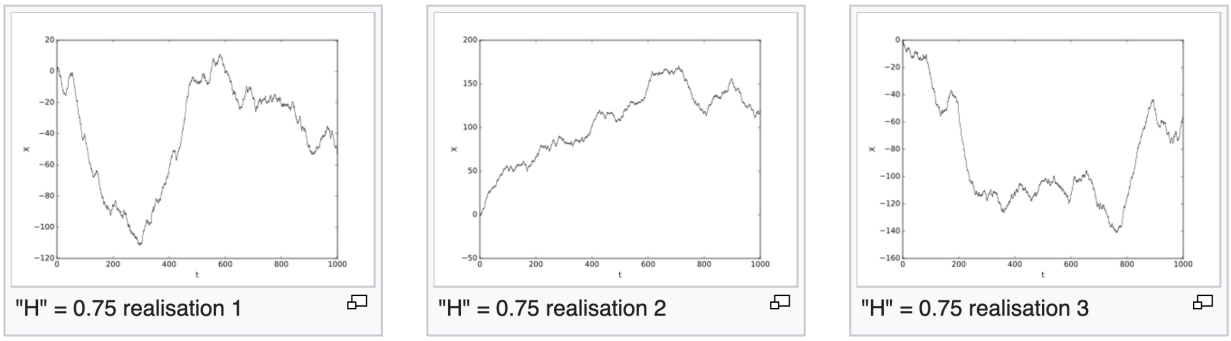


Figure 30: Three sample paths of fractional Brownian motion with Hurst parameter $H = 0.75$. Figure taken from https://en.wikipedia.org/wiki/Fractional_Brownian_motion.

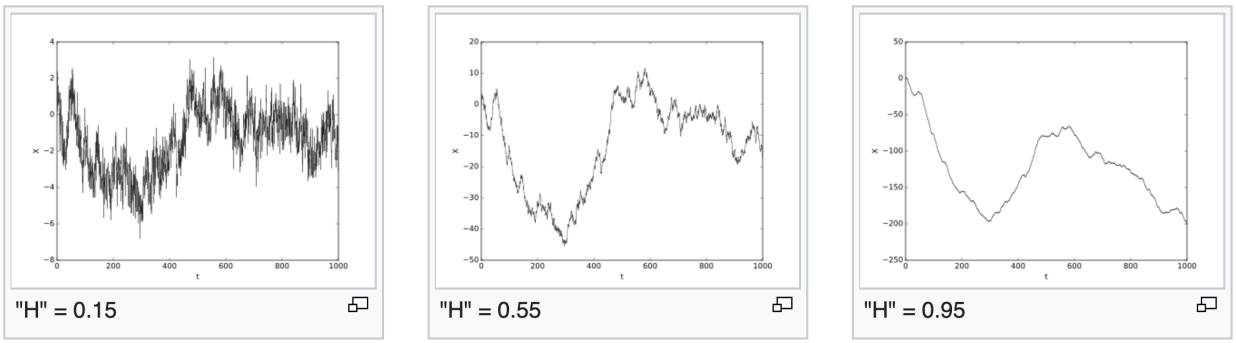


Figure 31: Sample paths of fractional Brownian motion with Hurst parameter $H = 0.15$ (left), $H = 0.55$ (middle), and $H = 0.95$ (right). Figure taken from https://en.wikipedia.org/wiki/Fractional_Brownian_motion.

We also remark that fBm, and hence the Wiener process too, are self similar. A stochastic process $X = (X(t))_{t\in\mathbb{R}}$ is self-similar of order H if

$$\forall\, a > 0\,, \quad (X(at))_{t\in\mathbb{R}} \overset{d}{=} a^H (X(t))_{t\in\mathbb{R}}$$

From the definition of fBm we see it is self-similar, as its mean function satisfies

$$\mathbb{E}[B_H(at)] = 0 = \mathbb{E}[a^H B_H(t)]$$

and its covariance function satisfies

$$\begin{aligned}\mathbb{E}[B_H(as)B_H(at)] &= \frac{1}{2}(|as|^{2H} + |at|^{2H} - |as - at|^{2H}) \\ &= \frac{a^{2H}}{2}(|s|^{2H} + |t|^{2H} - |t - s|^{2H}) \\ &= \mathbb{E}[a^H B_H(s) a^H B_H(t)]\end{aligned}$$

Furthermore, since it is a Gaussian process, it is completely determined by its mean function and covariance function.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 21 & 22: Time-Frequency Analysis of fBm

March 31, 2020 & April 2, 2020

Lecturer: Matthew Hirn

Fractional Brownian motions are interesting because they introduce several new behaviors relative to the Wiener process. For example, recall that the increments of regular Brownian motion, i.e. the Wiener process, are independent. This is not the case for fractional Brownian motion when $H \neq 1/2$. In fact the increments of fBm are negatively correlated for $H \in (0, 1/2)$ and positively correlated for $H \in (1/2, 1)$. To see this let $s_1 < t_1 < s_2 < t_2$ so that the intervals $[s_1, t_1]$ and $[s_2, t_2]$ are non-overlapping, and observe that

$$\begin{aligned}\mathrm{Cov}(B_H(t_1) - B_H(s_1), &B_H(t_2) - B_H(s_2)) \\ &= \mathbb{E}[(B_H(t_1) - B_H(s_1))(B_H(t_2) - B_H(s_2))] \\ &= \frac{1}{2}\left(|t_2 - s_1|^{2H} - |t_2 - t_1|^{2H} - (|s_2 - s_1|^{2H} - |s_2 - t_1|^{2H})\right)\end{aligned}$$

Now note that $t_2 - s_1 - (t_2 - t_1) = t_1 - s_1$ and $s_2 - s_1 - (s_2 - t_1) = t_1 - s_1$ and the function x^{2H} is concave when $H \in (0, 1/2)$ and convex when $H \in (1/2, 1)$. It follows that

$$\mathrm{Cov}(B_H(t_1) - B_H(s_1), B_H(t_2) - B_H(s_2)) \begin{cases} < 0 & H \in (0, 1/2) \\ > 0 & H \in (1/2, 1) \end{cases}$$

Therefore for $H \in (0, 1/2)$ the fBm is counter-persistent. That is, if it was increasing in the past, it is more likely to decrease in the future. On the other hand, for $H \in (1/2, 1)$, fBm is persistent. That is, the past trend is likely to continue in the future. We can this phenomenon in Figure 31, in which the the realization for $H = 0.15$ tends to go up down with a much higher frequency than than the realization for $H = 0.95$.

Visually, it would also appear the realizations become smoother with increasing Hurst parameter H. This is in fact the case, as the modulus of continuity of fractional Brownian motion is [8]:

$$\omega_{B_H}(\delta) = \delta^H |\log \delta|^{1/2}$$

Thus it is nearly H-Hölder, but not quite. Later on we will show the decay of the wavelet coefficients as the scale $s \to 0$ of fractional Brownian motion characterize the Hurst exponent H, and hence the regularity of B_H, even though realizations of B_H are nowhere differentiable.

Additionally, fBm have what is called long range dependence when $H \in (1/2, 1)$. Let us explain this in more detail, first by defining what long range dependence means and then by showing fBm possesses this property. We will say a stationary stochastic process has short range dependence if

$$\int_{\mathbb{R}} |R_X(\tau)| \, d\tau < +\infty$$

and a stationary stochastic process has long range dependence if

$$\int_{\mathbb{R}} |R_X(\tau)|\, d\tau = +\infty \tag{53}$$

Alternate definitions remove the absolute value, so that short range dependence means

$$\int_{\mathbb{R}} R_X(\tau)\, d\tau < +\infty$$

and long range dependence means

$$\int_{\mathbb{R}} R_X(\tau)\, d\tau = +\infty$$

Either way, recall if a stationary stochastic process is centered, i.e., $\mathbb{E}[X(t)] = 0$ for all $t \in \mathbb{R}$, then

$$\forall\, t \in \mathbb{R}\,, \quad R_X(\tau) = \mathbb{E}[X(t)X(t+\tau)]$$

Thus $R_X(\tau)$ measures the correlation between $X(t)$ and $X(t+\tau)$, which has time lag of τ. For stationary processes with short range dependence, this sum total (integration) of this correlation over all possible lags is finite, indicating these correlations must decay rapidly as the lag τ increases. On the other hand, stationary processes with long range dependence have correlations that persist even through large time lags, as indicated by (53). This behavior implies the process has "memory," which can be useful in many modeling situations.

Alternatively, one can say a stationary stochastic process has long range dependence if there exists a real number $\gamma \in (0,1)$ such that

$$\lim_{\tau \to +\infty} \tau^{\gamma} R_X(\tau) = c_X$$

for some constant $c_X > 0$. This is a characterization of long range dependence in the time domain, and it implies (53). We can also define long range dependence in the frequency domain. Indeed, from the frequency perspective, we say X has long range dependence if there exists a real number $\beta \in (0,1)$ and a constant $\tilde{c}_X > 0$ such that

$$\lim_{\omega \to 0} |\omega|^{\beta} \widehat{R}_X(\omega) = \tilde{c}_X \tag{54}$$

This frequency condition (54) also implies (53).

Notice the Ornstein-Uhlenbeck process has short range dependence since $R_X(\tau) = e^{-\theta|\tau|}$ and thus

$$\int_{\mathbb{R}} R_X(\tau)\, d\tau = \frac{2}{\theta}$$

or, from the time perspective,

$$\lim_{\tau \to +\infty} \tau^{\gamma} e^{-\theta\tau} = 0\,, \quad \forall\, \gamma \in (0,1)$$

Fractional Brownian motion for Hurst parameter $H \in (1/2, 1)$ is said to have long range dependence, but an fBm B_H is not stationary, so we need to make sense of this statement. One way to do so is to define a new random process based on the increments of B_H,

$$\widetilde{B}_H = (B_H(t+1) - B_H(t))_{t \in \mathbb{R}}$$

We will not take this path. Another path is to filter a stochastic process that has stationary increments with a wavelet transform. It turns out that the resulting process is stationary. As we mentioned earlier, since fBm is also self-similar, we will also be able to leverage the wavelet coefficients to characterize the self-similarity / regularity of B_H. First let us prove the following proposition.

Proposition 5.25. *Let X be a stochastic process with stationary increments and continuous sample paths. Let ψ be a continuous real valued wavelet with compact support. Then $X * \psi$ is a stationary process.*

Proof. Let $s, t \in \mathbb{R}$. Since ψ has zero average, we have:

$$\begin{aligned} X * \psi(t) &= \int_{\mathbb{R}} X(t-u)\psi(u)\,du \\ &= \int_{\mathbb{R}} X(t-u)\psi(u)\,du - X(t)\int_{\mathbb{R}} \psi(u)\,du \\ &= \int_{\mathbb{R}} [X(t-u) - X(t)]\psi(u)\,du \\ &\overset{d}{=} \int_{\mathbb{R}} [X(s-u) - X(s)]\psi(u)\,du \\ &= X * \psi(s) \end{aligned}$$

Letting $s = t + u$ we can apply the same argument to conclude:

$$(X * \psi(t+u))_{t \in \mathbb{R}} \overset{d}{=} (X * \psi(t))_{t \in \mathbb{R}}$$

□

Thus, in particular, $B_H * \psi_s$ is a stationary process for any scale parameter $s > 0$, since ψ_s is a wavelet. This means that its covariance function can be written as

$$\mathrm{Cov}_{B_H * \psi_s}(t, t+\tau) = R_{B_H * \psi_s}(\tau)$$

On the other hand, we cannot directly apply Theorem 5.21 because B_H is not stationary. Nevertheless, we have the following result:

Theorem 5.26. *Let B_H be a fractional Brownian motion with Hurst parameter H. Let ψ be continuous, real valued wavelet with compact support. Then $B_H * \psi$ is a stationary Gaussian process and*

$$\widehat{R}_{B_H * \psi}(\omega) = \frac{\lambda_H}{2|\omega|^{2H+1}} |\widehat{\psi}(\omega)|^2 \tag{55}$$

for some constant $\lambda_H > 0$.

Since $R_{B_H * \psi}$ is not integrable we must understand (55) in the sense of distributions. This means the proof must leverage the distributional definition of the Fourier transform. We give a brief overview now.

Recall the space of Schwartz class functions $\mathcal{S} = \mathcal{S}(\mathbb{R})$, which we originally defined in the proof of Theorem 2.18. The definition was:

$$\mathcal{S} = \left\{ \varphi \in \mathbf{C}^\infty(\mathbb{R}) : \forall\, m, n \in \mathbb{Z} \text{ with } m, n \geq 0\,,\, \sup_{t \in \mathbb{R}} |t|^m |\varphi^{(n)}(t)| < \infty \right\}$$

We note that if $\varphi \in \mathcal{S}$ then $\varphi^{(n)}$ has fast decay, that is

$$|\varphi^{(n)}(t)| \leq \frac{C_{m,n}}{1 + |t|^m}\,, \quad \forall\, m, n \geq 0$$

Now define the dual space of $\mathcal{S}$. It is denoted as $\mathcal{S}'$, and is referred to as the space of *tempered distributions*. It consists of all continuous linear functionals defined on $\mathcal{S}$:

$$\mathcal{S}' = \{T : \mathcal{S} \to \mathbb{C} : T \text{ is continuous and linear}\}$$

In order to understand what T "continuous" means, we need to place a metric on $\mathcal{S}$. To that end, define

$$\|\varphi\|_{m,n} = \sup_{t \in \mathbb{R}} |t|^m |\varphi^{(n)}(t)|$$

Each $\|\cdot\|_{m,n}$ defines a semi-norm on $\mathcal{S}$. We define the metric on $\mathcal{S}$ as

$$d(\varphi_1, \varphi_2) = \sum_{m,n \geq 0} \frac{1}{2^{m+n}} \cdot \frac{\|\varphi_1 - \varphi_2\|_{m,n}}{1 + \|\varphi_1 - \varphi_2\|_{m,n}}$$

Once can prove that $\mathcal{S}$ is complete with the metric $d(\varphi_1, \varphi_2)$. Furthermore, if $T \in \mathcal{S}'$ then there exists some $d \in \mathbb{Z}$ and constants $c_{m,n} \geq 0$ such that

$$|T(\varphi)| \leq \sum_{m=0}^{d} \sum_{n=0}^{d} c_{m,n} \|\varphi\|_{m,n} \tag{56}$$

Conversely, if T is a linear functional and (56) holds for some $d \in \mathbb{Z}$ and $c_{m,n} \geq 0$, then T is continuous and $T \in \mathcal{S}'$.

Now let us give some examples of tempered distributions $T \in \mathcal{S}'$.

Example 5.27. The Dirac distribution $\delta : \mathcal{S} \to \mathbb{C}$, defined as:

$$\delta(\varphi) = \varphi(0)$$

We can generalize it to $\delta_t : \mathcal{S} \to \mathbb{C}$,

$$\delta_t(\varphi) = \varphi(t)$$

Example 5.28. Let f be Lebesgue measurable and

$$|f(t)| \leq g(t)(1+|t|^m)$$

for some $m \geq 0$ with $g \in \mathbf{L}^1(\mathbb{R})$ and $g(t) \geq 0$. Then $T_f \in \mathcal{S}'$ where

$$T_f(\varphi) = \int_{\mathbb{R}} f(t)\varphi(t)\,dt$$

Indeed,

$$\begin{aligned}
|T_f(\varphi)| &\leq \int_{\mathbb{R}} g(t)(1+|t|^m)|\varphi(t)|\,dt \\
&\leq \sup_{u\in\mathbb{R}}(1+|u|^m)|\varphi(u)| \cdot \int_{\mathbb{R}} g(t)\,dt \\
&< \infty
\end{aligned}$$

Note that f does not have to be in $\mathbf{L}^1(\mathbb{R})$ or $\mathbf{L}^2(\mathbb{R})$.

Now we want to define the Fourier transform of a tempered distribution $T \in \mathcal{S}'$, which we will denote by $\widehat{T}$. We first note that for $\varphi \in \mathcal{S}$, we can use the $\mathbf{L}^1(\mathbb{R})$ definition of the Fourier transform to define $\widehat{\varphi}$:

$$\widehat{\varphi}(\omega) = \int_{\mathbb{R}} \varphi(t)e^{-i\omega t}\,dt$$

Since $\varphi \in \mathbf{C}^\infty(\mathbb{R})$ and $\varphi^{(n)}(t)$ has fast decay for each $n \geq 0$, $\widehat{\varphi} \in \mathbf{C}^\infty(\mathbb{R})$ and $\widehat{\varphi}^{(n)}(\omega)$ has fast decay for each $n \geq 0$ as well. Therefore $\widehat{\varphi} \in \mathcal{S}$. Furthermore, for $f, \varphi \in \mathcal{S}$, using Fubini's Theorem we have

$$\begin{aligned}
T_{\widehat{f}}(\varphi) = \int_{\mathbb{R}} \widehat{f}(t)\varphi(t)\,dt &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(\omega)e^{-i\omega t}\,d\omega\right]\varphi(t)\,dt \\
&= \int_{\mathbb{R}} f(\omega)\left[\int_{\mathbb{R}} \varphi(t)e^{-i\omega t}\,dt\right]\,d\omega \\
&= \int_{\mathbb{R}} f(\omega)\widehat{\varphi}(\omega)\,d\omega \\
&= T_f(\widehat{\varphi})
\end{aligned}$$

Inspired by this correspondence we make the following definition.

Definition 5.29. The Fourier transform of a tempered distribution $T \in \mathcal{S}'$ is the tempered distribution $\widehat{T} \in \mathcal{S}'$ defined as

$$\widehat{T}(\varphi) := T(\widehat{\varphi}), \quad \forall\, \varphi \in \mathcal{S}$$

Example 5.30. For the Dirac distribution

$$\widehat{\delta}(\varphi) = \delta(\widehat{\varphi}) = \widehat{\varphi}(0) = \int_{\mathbb{R}} \varphi(t)\, dt = T_{\chi_{\mathbb{R}}}(\varphi)$$

where $\chi_A(t) = 1$ if $t \in A \subseteq \mathbb{R}$ and $\chi_A(t) = 0$ if $t \notin A$. Hence the interpretation from earlier in the course that $\widehat{\delta}(\omega) = 1$ for all $\omega \in \mathbb{R}$.

Example 5.31. Let $f \in \mathbf{L}^1(\mathbb{R})$. Then using Fubini's Theorem:

$$\begin{aligned}
\widehat{T}_f(\varphi) = T_f(\widehat{\varphi}) &= \int_{\mathbb{R}} f(t)\widehat{\varphi}(t)\, dt \\
&= \int_{\mathbb{R}} f(t) \left[\int_{\mathbb{R}} \varphi(\omega) e^{-i\omega t}\, d\omega \right] dt \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(t) e^{-i\omega t}\, dt \right] \varphi(\omega)\, d\omega \\
&= \int_{\mathbb{R}} \widehat{f}(\omega)\varphi(\omega)\, d\omega = T_{\widehat{f}}(\varphi)
\end{aligned}$$

and thus the $\mathbf{L}^1(\mathbb{R})$ definition of the Fourier transform agrees with the distributional definition of the Fourier transform.

Our last example is more complicated, and needed for the proof of Theorem 5.26, so we collect it in the following lemma.

Lemma 5.32. *Let $f(t) = |t|^\alpha$ for $\alpha > 0$. Then, in the sense of distributions,*

$$\widehat{f}(\omega) = \lambda_\alpha |\omega|^{-(1+\alpha)}$$

That is

$$\widehat{T}_f(\varphi) = \lambda_\alpha \cdot \text{p.v.} \int_{\mathbb{R}} |\omega|^{-(1+\alpha)} \varphi(\omega)\, d\omega = \lim_{\epsilon \to 0^+} \int_{|\omega| > \epsilon} |\omega|^{-(1+\alpha)} \varphi(\omega)\, d\omega$$

Proof. For the purposes of this proof define $\varphi_s(t) = \varphi(st)$. A tempered distribution $T \in \mathcal{S}'$ is homogeneous of order α if

$$\forall\, s > 0\,,\ \varphi \in \mathcal{S}\,, \quad T(\varphi) = s^{1+\alpha} T(\varphi_s)$$

We first show that if T is homogeneous of order α, then $\widehat{T}$ is homogeneous of order $-(1+\alpha)$. Indeed we know:

$$\widehat{\varphi_s}(\omega) = s^{-1}\widehat{\varphi}(s^{-1}\omega) = s^{-1}\widehat{\varphi}_{s^{-1}}(\omega)$$

Therefore:

$$\begin{aligned}
\widehat{T}(\varphi_s) &= T(\widehat{\varphi_s}) \\
&= s^{-1} T(\widehat{\varphi}_{s^{-1}}) \\
&= s^{-1} s^{1+\alpha} T(\widehat{\varphi}) \\
&= s^\alpha \widehat{T}(\varphi)
\end{aligned}$$

Rearranging:

$$\widehat{T}(\varphi) = s^{-\alpha}\widehat{T}(\varphi_s) = s^{1-(1+\alpha)}\widehat{T}(\varphi)$$

Now observe that T_f with $f(t) = |t|^\alpha$ is homogeneous of order α since

$$T_f(\varphi_s) = \int_{\mathbb{R}} |t|^\alpha \varphi(st)\,dt = \int_{\mathbb{R}} |u/s|^\alpha\,\varphi(u)\frac{du}{s} = s^{-(1+\alpha)}\int_{\mathbb{R}} |u|^\alpha\varphi(u)\,du s^{-(1+\alpha)} T_f(\varphi)$$

Therefore $\widehat{T}_f$ must be homogeneous of order $-(1+\alpha)$. Additionally, $f(t) = |t|^\alpha$ is even, which means $T_f(\varphi)$ is "even," where the latter means

$$T_f(\varphi_{-1}) = T_f(\varphi)$$

Furthermore, since $f(t) = |t|^\alpha$ is real valued, $T_f(\varphi)$ is real valued for all real valued $\varphi \in \mathcal{S}$. It follows that $\widehat{T}(\varphi)$ must also be even and real valued if φ is real valued. But the only distributions which are homogeneous of order $-(1+\alpha)$, even, and real valued, are $c|\omega|^{-(1+\alpha)}$. □

Proof of Theorem 5.26. Set $f(t) = R_{B_H * \psi}(t)$. We compute:

$$\begin{aligned}
\widehat{T}_f &= T_f(\widehat{\varphi}) \\
&= \int_{\mathbb{R}} R_{B_H*\psi}(t)\widehat{\varphi}(t)\,dt \\
&= \int_{\mathbb{R}} \mathbb{E}[B_H * \psi(0) B_H * \psi(t)]\widehat{\varphi}(t)\,dt \\
&= \int_{\mathbb{R}} \mathbb{E}\left[\int_{\mathbb{R}} B_H(u)\psi(-u)\,du \cdot \int_{\mathbb{R}} B_H(v)\psi(t-v)\,dv\right]\widehat{\varphi}(t)\,dt \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{E}[B_H(u)B_H(v)]\psi(-u)\psi(t-v)\,du\,dv\right]\widehat{\varphi}(t)\,dt \\
&= \frac{1}{2}\int_{\mathbb{R}} \left[\int_{\mathbb{R}}\int_{\mathbb{R}} (|u|^{2H} + |v|^{2H} - |u-v|^{2H})\psi(-u)\psi(t-v)\,du\,dv\right]\widehat{\varphi}(t)\,dt \\
&= \frac{1}{2}\int_{\mathbb{R}} \Bigg[\int_{\mathbb{R}} |u|^{2H}\psi(-u)\underbrace{\int_{\mathbb{R}} \psi(t-v)\,dv}_{=0}\,du + \int_{\mathbb{R}} |v|^{2H}\psi(t-v)\underbrace{\int_{\mathbb{R}}\psi(-u)\,du}_{=0}\,dv - \cdots \\
&\qquad \cdots - \int_{\mathbb{R}}\int_{\mathbb{R}} |u-v|^{2H}\psi(-u)\psi(t-v)\,du\,dv\Bigg]\widehat{\varphi}(t)\,dt \\
&= -\frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} |u-v|^{2H}\psi(-u)\psi(t-v)\widehat{\varphi}(t)\,du\,dv\,dt \qquad (\text{CoV: } x = t-v) \\
&= -\frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} |t-(u+x)|^{2H}\psi(-u)\psi(x)\widehat{\varphi}(t)\,du\,dx\,dt \\
&= -\frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{R}} \psi(-u)\psi(x)\underbrace{\left[\int_{\mathbb{R}} |t-(u+x)|^{2H}\widehat{\varphi}(t)\,dt\right]}_{I}\,du\,dx
\end{aligned} \tag{57}$$

Now let us evaluate the integral I. First make a change of variables $y = t - (u + x)$. One obtains:

$$\begin{aligned} I &= \int_{\mathbb{R}} |y|^{2H} \widehat{\varphi}(y + u + x)\, dy \\ &= \int_{\mathbb{R}} |y|^{2H} \left[\int_{\mathbb{R}} \varphi(\omega) e^{-i\omega(y+u+x)}\, d\omega \right] dy \\ &= \int_{\mathbb{R}} |y|^{2H} \left[\int_{\mathbb{R}} e^{i\omega(u+x)} \varphi(\omega) e^{-i\omega y}\, d\omega \right] dy \\ &= \int_{\mathbb{R}} |y|^{2H} \mathcal{F}(M_{-(u+x)}\varphi)(y)\, dy \qquad \left[(M_v\varphi)(y) = e^{ivy}\varphi(y) \right] \\ &= -\int_{\mathbb{R}} \lambda_H |\omega|^{-(2H+1)} e^{-i\omega(u+x)} \varphi(\omega)\, d\omega \end{aligned} \tag{58}$$

where in the last line we used Lemma 5.32. Now plug (58) into (57) to obtain:

$$\begin{aligned} (57) &= \frac{\lambda_H}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} \psi(-u)\psi(x) \left[\int_{\mathbb{R}} |\omega|^{-(2H+1)} e^{-i\omega(u+x)} \varphi(\omega)\, d\omega \right] du\, dx \\ &= \frac{\lambda_H}{2} \int_{\mathbb{R}} |\omega|^{-(2H+1)} \varphi(\omega) \int_{\mathbb{R}} \psi(-u) e^{-i\omega u}\, du \int_{\mathbb{R}} \psi(x) e^{-i\omega x}\, dx\, d\omega \\ &= \int_{\mathbb{R}} \frac{\lambda_H}{2} |\omega|^{-(2H+1)} \widehat{\psi}^*(\omega) \widehat{\psi}(\omega) \varphi(\omega)\, d\omega \\ &= \int_{\mathbb{R}} \frac{\lambda_H}{2} |\omega|^{-(2H+1)} |\widehat{\psi}(\omega)|^2 \varphi(\omega)\, d\omega \end{aligned}$$

We conclude that, in the distributional sense,

$$\widehat{R}_{B_H * \psi}(\omega) = \frac{\lambda_H}{2|\omega|^{2H+1}} |\widehat{\psi}(\omega)|^2$$

□

It is tempting to think of the "power spectral density of B_H" as $(\lambda_H/2)|\omega|^{-(2H+1)}$ but this is not quite correct, and would lead, for example, to the wrong interpretation of its long range dependence property. Recall that since ψ is a wavelet,

$$\widehat{\psi}(\omega) = O(\omega) \text{ as } \omega \to 0$$

and thus $|\widehat{\psi}(\omega)|^2 = O(|\omega|^2)$. It follows that

$$\widehat{R}_{B_H * \psi}(\omega) = O(|\omega|^{1-2H})$$

Thus when $H \in (1/2, 1)$ we see that

$$\lim_{\omega \to 0} |\omega|^{2H-1} \widehat{R}_{B_H * \psi}(\omega) = c > 0$$

with $0 < 2H - 1 < 1$. Therefore we see that $B_H * \psi$ has long range dependence. Notice, however, for $H \in (0, 1/2)$ the same cannot be said since

$$\forall\, H \in (0, 1/2)\,, \quad \lim_{\omega \to 0} \widehat{R}_{B_H * \psi}(\omega) = c \cdot \lim_{\omega \to 0} |\omega|^{1-2H} = 0$$

Using the self-similarity of fBm, one can also show:

$$WB_H(u, s) \overset{d}{=} s^{H+1/2} WB_H\left(\frac{u}{s}, 1\right)$$

We leave the details as an exercise.

Exercise 54. Read Section 6.4 of *A Wavelet Tour of Signal Processing*.

Exercise 55. Let ψ be a continuous, compactly supported, real valued wavelet. Recall $Wf(u, s) = f * \overline{\psi}_s(u)$ with $\overline{\psi}(t) = \psi(-t)$. Prove:

$$\mathbb{E}[WB_H(u, s) WB_H(v, s)] = -\frac{s^{2H+1}}{2} \int_{\mathbb{R}} |t|^{2H} \psi * \overline{\psi}\left(\frac{u - v}{s} - t\right) dt$$

Observe that since B_H is a Gaussian process and

$$\mathbb{E}[WB_H(u, s)] = \mathbb{E}\left[\int_{\mathbb{R}} B_H(t) \psi_s(t - u)\, dt\right] = \int_{\mathbb{R}} \mathbb{E}[B_H(t)] \psi_s(t - u)\, dt = 0$$

you now have a direct proof that $(WB_H(u, s))_{u \in \mathbb{R}}$ is a stationary stochastic process.

Exercise 56. Let X be a second order stochastic process that is self-similar of order H with continuous sample paths, and let ψ be a continuous, compactly supported real valued wavelet.

(a) Prove:

$$X * \psi_s(u) \overset{d}{=} s^{H+1/2} X * \psi\left(\frac{u}{s}\right)$$

Conclude that if X also has stationary increments then:

$$\mathbb{E}[|X * \psi_s(u)|] = s^{H+1/2} \mathbb{E}[|X * \psi(0)|]$$

(b) Suppose X also has stationary increments. Prove:

$$\frac{\mathbb{E}[||X * \psi_{s_1}| * \psi_{s_2}(u)|]}{s_1^{1/2} \mathbb{E}[|X * \psi_{s_1}(u)|]} = \frac{\mathbb{E}[||X * \psi| * \psi_{s_2/s_1}(0)|]}{\mathbb{E}[|X * \psi(0)|]}$$

The numerator of the left hand side is called a wavelet scattering moment. Give an interpretation of this result.

Exercise 57. One can obtain realizations of fractional Brownian motion in MATLAB using the `wfbm` function (`https://www.mathworks.com/help/wavelet/ref/wfbm.html`) or in Python using the `fbm` package (available at: `https://pypi.org/project/fbm/`).

(a) Generate realizations of fractional Brownian motion for three different Hurst parameters H, one with $H < 1/2$, one with $H = 1/2$ (regular Brownian motion), and one with $H > 1/2$. Provide a plot of each realization. Using your code for the real valued wavelet transform from Exercise 44, compute the wavelet transform for each realization and plot the wavelet coefficients as in Figure 6.22(b) from the book.

Remark: You should generate long realizations of fBm with $N \geq 10000$.

(b) Now estimate the Hurst parameter H using the moments computed in Exercise 56(a). Do so by noting that Exercise 56(a) implies

$$F(\log_2 s) := \log_2 \mathbb{E}[|B_H * \psi_s(u)|] = (H + 1/2) \log_2 s + \log_2 \mathbb{E}[|B_H * \psi(0)|] \tag{59}$$

Since $|B_H * \psi_s|$ is stationary for each $s > 0$, the function $F(\log_2 s)$ on the left hand side of (59) does not depend on u and can be considered as a function of $\log_2 s$. The right hand side of (59) shows $F(\log_2 s)$ is linear with a slope of $H + 1/2$. Plot $F(\log_2 s) = \log_2 \mathbb{E}[|B_H * \psi_s(u)|]$ as a function of $\log_2 s$ for your three realizations from part (a). Estimate the slope numerically and compare it to the true value of H.

Remark: To estimate $\mathbb{E}[|B_H * \psi_s(u)|]$ note that $|B_H * \psi_s|$ is stationary. For a stationary process Y, one can estimate $\mathbb{E}[Y(t)] = \mathbb{E}[Y(0)]$ by computing:

$$\frac{1}{N} \sum_{i=1}^{N} Y(t_i) \approx \mathbb{E}[Y(0)]$$

for large N.

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 23 & 24: Introduction to Frames

April 7 & 9, 2020

Lecturer: Matthew Hirn

6 Frames

Chapter 5 of A Wavelet Tour of Signal Processing.

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 as 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 to wavelets possessing the framework to not only complete their story, but also the tools to chart a path forward into signal analysis via more general dictionaries.

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., $\mathbf{L}^2(\mathbb{R})$, $\boldsymbol{\ell}^2$, and $\mathbb{R}^N$ or $\mathbb{C}^N$. Consider a dictionary

$$\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma} \subset \mathcal{H}$$

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

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

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

$$A\|f\|^2 \leq \sum_{\gamma \in \Gamma} |\langle f, \phi_\gamma \rangle|^2 \leq B\|f\|^2 \tag{60}$$

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 image(Φ).

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 (60), we obtain:

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

Thus the only element of $\mathcal{H}$ orthogonal to every $\phi_\gamma \in \mathcal{D}$ is $f = 0$. 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 Φ analyzes a signal $f \in \mathcal{H}$ by testing it against the dictionary atoms ϕ_γ. The adjoint of Φ defines a synthesis operator, which we now explain. Consider the space of $\boldsymbol{\ell}^2$ sequences indexed by Γ:

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

Notice that the frame condition (60) guarantees that

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

Therefore Φ has an adjoint

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

which is defined through the following relation:

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

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

$$\begin{aligned}\langle a, \Phi f\rangle &= \sum_{\gamma\in\Gamma} a[\gamma]\langle f, \phi_\gamma\rangle^* \\ &= \sum_{\gamma\in\Gamma} a[\gamma]\langle \phi_\gamma, f\rangle \\ &= \sum_{\gamma\in\Gamma} \langle a[\gamma]\phi_\gamma, f\rangle \\ &= \left\langle \sum_{\gamma\in\Gamma} a[\gamma]\phi_\gamma, f \right\rangle\end{aligned}$$

from which it follows that

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

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

Notice that the frame condition (60) can be rewritten as:

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

where

$$\Phi^*\Phi f = \sum_{\gamma\in\Gamma} \langle f, \phi_\gamma\rangle \phi_\gamma \tag{61}$$

Notice that (61) looks exactly like the formula you get for expanding a vector f in an orthonormal basis. However, Φ here is a frame and so in general (61) will not return f but rather another element of $\mathcal{H}$. Back to the point at hand, it follows that we can take A and B as:

$$\begin{aligned}A &= \inf_{f\in\mathcal{H}} \frac{\langle \Phi^*\Phi f, f\rangle}{\|f\|^2} \\ B &= \sup_{f\in\mathcal{H}} \frac{\langle \Phi^*\Phi f, f\rangle}{\|f\|^2}\end{aligned}$$

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 Φ. The next theorem shows that if the frame analysis operator is stable (as defined by the frame condition (60)), then the frame synthesis operator obeys a similar stability condition.

Theorem 6.1. *A dictionary $\mathcal{D} = \{\phi_\gamma\}_{\gamma\in\Gamma}$ is a frame with bounds $0 < A \le B < \infty$ if and only if*

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

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

$$\inf_{f \in \mathcal{H}} \frac{\langle \Phi^* \Phi f, f \rangle}{\|f\|^2} = \inf_{a \in \text{image}(\Phi)} \frac{\langle \Phi \Phi^* a, a \rangle}{\|a\|^2} \tag{62}$$

and

$$\sup_{f \in \mathcal{H}} \frac{\langle \Phi^* \Phi f, f \rangle}{\|f\|^2} = \sup_{a \in \text{image}(\Phi)} \frac{\langle \Phi \Phi^* a, a \rangle}{\|a\|^2} \tag{63}$$

Let us first consider the case of a finite dimensional Hilbert space. In this case $\mathcal{H} \cong \mathbb{R}^N$ or $\mathcal{H} \cong \mathbb{C}^N$. Suppose that $\mathcal{D}$ is a frame and let λ be an eigenvalue of $\Phi^*\Phi$ with eigenvector $f_\lambda \neq 0$. Note the frame condition implies $\Phi^*\Phi$ is invertible and every eigenvalue satisfies $A \leq \lambda \leq B$. Furthermore $\Phi^*\Phi$ can be identified with an $N \times N$ matrix. We claim that $\Phi f_\lambda \in \text{image}(\Phi)$ is an eigenvector of $\Phi\Phi^*$ also with eigenvalue λ; 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 (60) imply that $\|\Phi f_\lambda\|^2 \geq A \|f_\lambda\|^2$. Since $\dim(\text{image}(\Phi)) = N$, we have shown the eigenvalues of $\Phi^*\Phi$ and $\Phi\Phi^*|_{\text{image}(\Phi)}$ are identical and we conclude that (62) and (63) hold.

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 Φ_N is the restriction of Φ 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:

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

The proof for the supremum is identical.

□

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 6.2. *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 58. Read Section 5.1.1 of *A Wavelet Tour of Signal Processing.*

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 Φ 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:

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

The next theorem computes the pseudo-inverse explicitly.

Theorem 6.3. *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 (60) as:

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

Thus

$$\Phi^*\Phi f = 0 \Longleftrightarrow f = 0$$

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 Φ. 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 $\text{null}(\Phi^*) \subseteq \text{image}(\Phi)^\perp$. Similarly, now let $a \in \text{image}(\Phi)^\perp$, so that:

$$\begin{aligned} a \in \text{image}(\Phi)^\perp &\Longrightarrow \langle a, \Phi f\rangle = 0\,, \quad \forall\, f \in \mathcal{H} \\ &\Longrightarrow \langle \Phi^* a, f\rangle = 0\,, \quad \forall\, f \in \mathcal{H} \\ &\Longrightarrow \Phi^* a = 0 \\ &\Longrightarrow a \in \text{null}(\Phi^*) \end{aligned}$$

Therefore $\text{image}(\Phi)^\perp \subseteq \text{null}(\Phi^*)$ and we conclude that $\text{image}(\Phi)^\perp = \text{null}(\Phi^*)$. But then

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

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

The pseudo-inverse implements a signal synthesis with the *(canonical) dual frame*, defined by:

$$\widetilde{\phi}_\gamma = (\Phi^*\Phi)^{-1}\phi_\gamma$$

which has associated frame analysis operator

$$\widetilde{\Phi} f(\gamma) = \langle f, \widetilde{\phi}_\gamma\rangle$$

The next theorem shows that the dual frame synthesis operator is indeed the pseudo-inverse of the original frame analysis operator, and that the dual frame is in fact a frame.

Theorem 6.4. *Let $\mathcal{D} = \{\phi_\gamma\}_{\gamma\in\Gamma}$ be a frame with frame bounds $0 < A \leq B < \infty$. Then the dual frame synthesis operator satisfies*

$$\widetilde{\Phi}^* = \Phi^\dagger \tag{64}$$

and thus

$$f = \sum_{\gamma\in\Gamma} \langle f, \phi_\gamma\rangle \widetilde{\phi}_\gamma = \sum_{\gamma\in\Gamma} \langle f, \widetilde{\phi}_\gamma\rangle \phi_\gamma \tag{65}$$

Furthermore, the dual dictionary

$$\widetilde{\mathcal{D}} = \{\widetilde{\phi}_\gamma\}_{\gamma\in\Gamma}$$

is a frame (hence the name dual frame) with frame bounds $0 < 1/B \leq 1/A < \infty$*, meaning that*

$$\frac{1}{B}\|f\|^2 \leq \sum_{\gamma\in\Gamma} |\langle f, \widetilde{\phi}_\gamma\rangle|^2 \leq \frac{1}{A}\|f\|^2, \quad \forall\, f \in \mathcal{H} \tag{66}$$

If the frame is tight (i.e., $A = B$*), then*

$$\widetilde{\phi}_\gamma = \frac{1}{A}\phi_\gamma$$

To prove this theorem, we will need the following lemma.

Lemma 6.5. *If* $L : \mathcal{H} \to \mathcal{H}$ *is a self-adjoint operator such that there exists* $0 < A \leq B < \infty$ *satisfying*

$$A\|f\|^2 \leq \langle Lf, f\rangle \leq B\|f\|^2, \quad \forall\, f \in \mathcal{H} \tag{67}$$

then L *is invertible and*

$$\frac{1}{B}\|f\|^2 \leq \langle L^{-1}f, f\rangle \leq \frac{1}{A}\|f\|^2, \quad \forall\, f \in \mathcal{H} \tag{68}$$

Proof. Suppose first that $\mathcal{H}$ is finite dimensional of dimension N. Since L is self-adjoint, it has an orthonormal set of eigenvectors $e_1, \ldots, e_N \in \mathcal{H}$ with eigenvalues $\lambda_1, \ldots, \lambda_N$ such that

$$Le_k = \lambda_k e_k, \quad \forall\, 1 \leq K \leq N$$

Equation (67) implies that $A \leq \lambda_k \leq B$ for each k. The operator L is therefore invertible, and it's eigenvalues are λ_k^{-1} with the same orthonormal eigenvectors e_k for $1 \leq k \leq N$. It follows that (68) must hold. The proof is extended to infinite dimensions using the same technique as in the proof of Theorem 6.1. □

Proof of Theorem 6.4. We first rewrite the dual analysis operator (noting that $\Phi^*\Phi$ is self-adjoint, and thus so is $(\Phi^*\Phi)^{-1}$):

$$\begin{aligned}\widetilde{\Phi}f(\gamma) = \langle f, \widetilde{\phi}_\gamma\rangle &= \langle f, (\Phi^*\Phi)^{-1}\phi_\gamma\rangle \\ &= \langle (\Phi^*\Phi)^{-1}f, \phi_\gamma\rangle \\ &= \Phi(\Phi^*\Phi)^{-1}f(\gamma)\end{aligned}$$

Thus

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

and we compute:

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

That proves (64).

Note that (65) can be written as:

$$I = \widetilde{\Phi}^*\Phi = \Phi^*\widetilde{\Phi}$$

where I is the identity operator. Since $\widetilde{\Phi}^* = \Phi^\dagger$, we have

$$\widetilde{\Phi}^*\Phi = \Phi^\dagger\Phi = I \tag{69}$$

Using the facts that $(\widetilde{\Phi}^*\Phi)^* = \Phi^*\widetilde{\Phi}$ and $I^* = I$, and taking the adjoint of both sides of (69), we obtain the second equality.

For the proof of (66), we use Lemma 6.5. Recall that the frame conditions can be rewritten as:

$$A\|f\|^2 \leq \langle \Phi^*\Phi f, f\rangle \leq B\|f\|^2, \quad \forall\, f \in \mathcal{H}$$

Applying Lemma 6.5 to $L = \Phi^*\Phi$ proves that

$$\frac{1}{B}\|f\|^2 \leq \langle (\Phi^*\Phi)^{-1} f, f\rangle \leq \frac{1}{A}\|f\|^2, \quad \forall\, f \in \mathcal{H}$$

Furthermore, using the first part of the proof we have:

$$\begin{aligned}\sum_{\gamma\in\Gamma} |\langle f, \widetilde{\phi}_\gamma\rangle|^2 &= \|\widetilde{\Phi} f\|^2 \\ &= \langle \Phi(\Phi^*\Phi)^{-1} f, \Phi(\Phi^*\Phi)^{-1} f\rangle \\ &= \langle \Phi^*\Phi(\Phi^*\Phi)^{-1} f, (\Phi^*\Phi)^{-1} f\rangle \\ &= \langle f, (\Phi^*\Phi)^{-1} f\rangle\end{aligned}$$

This proves (66).

If $A = B$, then

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

Thus the spectrum of $\Phi^*\Phi$ is only A, and we have $\Phi^*\Phi = AI$. It follows that $\widetilde{\phi}_\gamma = (\Phi^*\Phi)^{-1}\phi_\gamma = A^{-1}\phi_\gamma$. □

This theorem proves that one way to reconstruct a signal f from its frame coefficients $\Phi f(\gamma) = \langle f, \phi_\gamma\rangle$ is to use the dual frame $\widetilde{\phi}_\gamma$; equivalently, the synthesis coefficients of f in $\mathcal{D} = \{\phi_\gamma\}_{\gamma\in\Gamma}$ are the dual frame coefficients $\widetilde{\Phi} f(\gamma) = \langle f, \widetilde{\phi}_\gamma\rangle$. If the frame is tight, then we have the simple reconstruction formula:

$$f = \frac{1}{A}\sum_{\gamma\in\Gamma} \langle f, \phi_\gamma\rangle \phi_\gamma$$

which mirrors the reconstruction of a signal f in an orthonormal basis, except for the factor of A^{-1}.

If $\mathcal{D} = \{\phi_\gamma\}_{\gamma\in\Gamma}$ is a Riesz basis then the dictionary atoms are linearly independent, which implies that $\text{image}(\Phi) = \boldsymbol{\ell}^2(\Gamma)$; therefore the dual frame $\widetilde{\mathcal{D}} = \{\widetilde{\phi}_\gamma\}_{\gamma\in\Gamma}$ is also a Riesz basis. Inserting $f = \phi_n$ into (65) yields:

$$\phi_n = \sum_{\gamma\in\Gamma} \langle \phi_n, \widetilde{\phi}_\gamma \rangle \phi_\gamma$$

The linear independence of $\mathcal{D}$ implies that the only expansion of ϕ_n in $\mathcal{D}$ is the trivial expansion $\phi_n = \phi_n$, which implies that

$$\langle \phi_n, \widetilde{\phi}_\gamma \rangle = \begin{cases} 1 & n = \gamma \\ 0 & n \neq \gamma \end{cases}$$

Thus the frame and dual frame are *biorthogonal bases* for $\mathcal{H}$. Furthermore, if the Riesz basis is normalized so that $\|\phi_\gamma\| = 1$ for all $\gamma \in \Gamma$, then using the dual frame bounds (66) and the biorthogonality we have:

$$\frac{1}{B} = \frac{1}{B}\|\phi_n\|^2 \leq \sum_{\gamma\in\Gamma} |\langle \phi_n, \widetilde{\phi}_\gamma \rangle|^2 = 1 \leq \frac{1}{A}\|\phi_n\|^2 = \frac{1}{A}$$

This shows that

$$A \leq 1 \leq B$$

for a Riesz basis with normalized atoms.

Exercise 59. Read Section 5.1.2 of *A Wavelet Tour of Signal Processing.*

Exercise 60. Prove that if $K \in \mathbb{R}$ with $|K| \geq 1$, then

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

is a tight frame for $\mathbf{L}^2[0,1]$. Compute the frame bound. Prove the result cannot hold in general for $|K| < 1$ by finding a specific such K and proving $\mathcal{D}$ is not a frame for that K.

Exercise 61. Prove that a finite set of N vectors $\{\phi_n\}_{1\leq n\leq N}$ is always a frame for the space $\mathbf{V}$ defined by:

$$\mathbf{V} = \text{span}\{\phi_n\}_{1\leq n\leq N}$$

Exercise 62. Let $\phi_p \in \mathbb{R}^N$ be defined as:

$$\phi_p[n] = \delta[(n-p) \bmod N] - \delta[(n-p-1) \bmod N], \quad 0 \leq p < N$$

and define $\mathbf{V}$ as:

$$\mathbf{V} = \left\{ f \in \mathbb{R}^N : \sum_{n=0}^{N-1} f[n] = 0 \right\}$$

Prove that the dictionary $\mathcal{D} = \{\phi_p\}_{0\leq p<N}$ is a translation invariant frame for $\mathbf{V}$; that is, prove there exists some filter $h \in \mathbb{R}^N$ such that $(\Phi f)(p) = (f \circledast h)(p)$ for all $f \in \mathbf{V}$, where $(\Phi f)(p) = \langle f, \phi_p \rangle$ is the frame analysis operator. Compute the frame bounds. Is it a numerically stable frame when N is large?

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 25 & 26: Frame Projections and Semi-Discrete Frames

April 14 & 16, 2020

Lecturer: Matthew Hirn

6.1.3 Dual Frame Analysis and Synthesis Computations

Section 5.1.3 of A Wavelet Tour of Signal Processing.

To compress and denoise a signal f we will project the signal onto a closed subspace $\mathbf{V} \subset \mathcal{H}$ that is generated from the span of a subset dictionary atoms from a larger dictionary. We thus need to study projections onto $\mathbf{V}$. As is well known from linear algebra, the best linear approximation of $f \in \mathcal{H}$ in $\mathbf{V}$ is the orthogonal projection of f onto $\mathbf{V}$. To make clear the setup, we let $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma} \subset \mathcal{H}$ be a dictionary in $\mathcal{H}$, but which is a frame *only on* $\mathbf{V}$, i.e.,

$$A\|g\|^2 \leq \sum_{\gamma \in \Gamma} |\langle g, \phi_\gamma \rangle|^2 \leq B\|g\|^2, \quad \forall\, g \in \mathbf{V}$$

The analysis operator Φ is still defined on all of $\mathcal{H}$, but it may not behave "nicely" off of $\mathbf{V}$. The next theorem shows how to compute the orthogonal projection of $f \in \mathcal{H}$ onto $\mathbf{V}$ with the dual frame.

Theorem 6.6. *Let $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$ be a frame for $\mathbf{V} \subset \mathcal{H}$, and $\widetilde{\mathcal{D}} = \{\widetilde{\phi}_\gamma\}_{\gamma \in \Gamma}$ its dual frame on $\mathbf{V}$. The orthogonal projection of $f \in \mathcal{H}$ onto $\mathbf{V}$ is*

$$P_{\mathbf{V}} f = \sum_{\gamma \in \Gamma} \langle f, \phi_\gamma \rangle \widetilde{\phi}_\gamma = \sum_{\gamma \in \Gamma} \langle f, \widetilde{\phi}_\gamma \rangle \phi_\gamma \tag{70}$$

Proof. To show that $P_{\mathbf{V}}$ is a projection, we must show that $P_{\mathbf{V}} g = g$ for all $g \in \mathbf{V}$. But since $\mathcal{D}$ is a frame for $\mathbf{V}$, we have the synthesis formula given by (65) which proves that $P_{\mathbf{V}} g = g$ for all $g \in \mathbf{V}$.

To show that $P_{\mathbf{V}}$ is an orthogonal projection, we must verify that

$$\langle f - P_{\mathbf{V}} f, \phi_n \rangle = 0, \quad \forall\, n \in \Gamma$$

Note that (65) implies that

$$\phi_n = \sum_{\gamma \in \Gamma} \langle \phi_n, \widetilde{\phi}_\gamma \rangle \phi_\gamma$$

Therefore we compute:

$$\begin{aligned}
\langle f - P_{\mathbf{V}} f, \phi_n \rangle &= \langle f - \sum_{\gamma \in \Gamma} \langle f, \phi_\gamma \rangle \widetilde{\phi}_\gamma, \phi_n \rangle \\
&= \langle f, \phi_n \rangle - \sum_{\gamma \in \Gamma} \langle f, \phi_\gamma \rangle \langle \widetilde{\phi}_\gamma, \phi_n \rangle \\
&= \langle f, \phi_n \rangle - \sum_{\gamma \in \Gamma} \langle f, \langle \widetilde{\phi}_\gamma, \phi_n \rangle^* \phi_\gamma \rangle \\
&= \langle f, \phi_n - \sum_{\gamma \in \Gamma} \langle \widetilde{\phi}_\gamma, \phi_n \rangle^* \phi_\gamma \rangle \\
&= \langle f, \phi_n - \sum_{\gamma \in \Gamma} \langle \phi_n, \widetilde{\phi}_\gamma \rangle \phi_\gamma \rangle \\
&= \langle f, \phi_n - \phi_n \rangle = 0
\end{aligned}$$

□

Since $\mathcal{D}$ is a frame for a subspace $\mathbf{V} \subset \mathcal{H}$, Φ is only invertible on this subspace and the definition of the pseudo-inverse is now:

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

Let $\Phi_{\mathbf{V}}$ be the restriction of the frame analysis operator to $\mathbf{V}$. The operator $\Phi^* \Phi_{\mathbf{V}}$ is invertible on $\mathbf{V}$ and we write $(\Phi^* \Phi_{\mathbf{V}})^{-1}$ as its inverse on $\mathbf{V}$. One can verify that

$$\Phi^\dagger = (\Phi^* \Phi_{\mathbf{V}})^{-1} \Phi^* = \widetilde{\Phi}^*$$

Let $f \in \mathcal{H}$. Theorem 6.6 and (70) give two ways in which to compute orthogonal projections onto $\mathbf{V}$. In a dual synthesis scenario, the orthogonal projection $P_{\mathbf{V}} f$ is computed from the frame analysis coefficients with the dual frame synthesis operator:

$$P_{\mathbf{V}} f = \widetilde{\Phi}^* \Phi f = \sum_{\gamma \in \Gamma} \langle f, \phi_\gamma \rangle \widetilde{\phi}_\gamma \tag{71}$$

If the frame $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$ does not depend on the signal f, then the dual frame vectors are precomputed:

$$\widetilde{\phi}_\gamma = (\Phi^* \Phi_{\mathbf{V}})^{-1} \phi_\gamma$$

and the signal $P_{\mathbf{V}} f$ is synthesized with (71).

However, in many applications the frame vectors depend on the signal f. In this case the dual frame vectors $\widetilde{\phi}_\gamma$ cannot be computed in advance, and it is highly inefficient to compute them directly for each new signal f. In this case, we have already computed Φf and we want to compute $P_{\mathbf{V}} f$. We compute first:

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

Let L be the linear operator defined as

$$Lh = \Phi^* \Phi_{\mathbf{V}} h, \quad \forall\, h \in \mathbf{V}$$

We then compute $P_{\mathbf{V}} f$ via:

$$L^{-1} y = (\Phi^* \Phi_{\mathbf{V}})^{-1} \Phi^* \Phi f = \widetilde{\Phi}^* \Phi f = P_{\mathbf{V}} f$$

We have already encountered several situations which would lead to something similar to the above scenario. For example, when we studied instantaneous frequencies we focused on the ridge points of either the windowed Fourier transform $Sf(u, \xi)$ or the wavelet transform $Wf(u, s)$. While these are not frames according to our current definition (since the index set (u, ξ) or (u, s) is uncountable), this is something we will remedy shortly. The subspace $\mathbf{V}$ then depends on the signal f since it is the subspace of $\mathcal{H}$ generated by the span of the $g_{u,\xi}$ or the $\psi_{u,s}$ that correspond to the ridge points of f in either the windowed Fourier or wavelet representation. Computing $P_{\mathbf{V}} f$ then synthesizes a signal $\widetilde{f}$ from only the ridge information of f. One can do something similar (and we will in a bit) when analyzing signals with isolated singularities and generating $\mathbf{V}$ as the span of the $\psi_{u,s}$ that correspond to the wavelet modulus maxima. As we shall see the synthesized signal $\widetilde{f} = P_{\mathbf{V}} f \approx f$, thus indicating that these local maxima points carry the majority of information in such signals.

The alternate scenario is a dual analysis, in which $P_{\mathbf{V}} f$ is computed as

$$P_{\mathbf{V}} f = \Phi^* \widetilde{\Phi} f = \sum_{\gamma \in \Gamma} \langle f, \widetilde{\phi}_\gamma \rangle \phi_\gamma$$

Similarly to before, if Φ does not depend upon f, then the dual frame vectors $\widetilde{\phi}_\gamma$ can be precomputed.

It is also possible in this case to view $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$ as a subset of a larger frame, which has been obtained by solving for a sparse approximation of f in the larger frame.

When $\mathcal{D}$ depends on f, we again circumvent computing the dual frame directly. Let

$$a[\gamma] = \widetilde{\Phi} f(\gamma) = \langle f, \widetilde{\phi}_\gamma \rangle$$

and note that

$$P_{\mathbf{V}} f = \Phi^* a = \sum_{\gamma \in \Gamma} a[\gamma] \phi_\gamma$$

Since $\Phi P_{\mathbf{V}} f = \Phi f$, we have that

$$\Phi \Phi^* a = \Phi f$$

Let $\Phi^*_{\mathrm{Im}(\Phi)}$ be the restriction of Φ^* to image(Φ). Since $\Phi \Phi^*_{\mathrm{Im}(\Phi)}$ is invertible on image(Φ), we have

$$a = (\Phi \Phi^*_{\mathrm{Im}(\Phi)})^{-1} \Phi f$$

Notice that a is obtained by computing $a = L^{-1} y$, where in this case $y = \Phi f$ and $L = \Phi \Phi^*_{\mathrm{Im}(\Phi)}$.

Exercise 63. Read Section 5.1.3 of *A Wavelet Tour of Signal Processing.*

Exercise 64. Read Section 5.1.4 of *A Wavelet Tour of Signal Processing.*

6.1.4 Translation Invariant Frames

Section 5.1.5 of A Wavelet Tour of Signal Processing.

Let $\{\phi_\gamma\}_{\gamma\in\Gamma} \subset \mathbf{L}^2(\mathbb{R}^d)$ be a countable family of time frequency atoms. Recall that a translation invariant dictionary is a dictionary $\mathcal{D}$ of the form

$$\mathcal{D} = \{\phi_{u,\gamma}\}_{u\in\mathbb{R},\,\gamma\in\Gamma}$$

where

$$\phi_{u,\gamma}(x) = \phi_\gamma(x-u)$$

The analysis operator associated to $\mathcal{D}$ acts upon $f \in \mathbf{L}^2(\mathbb{R}^d)$ and is defined as

$$\Phi f(u,\gamma) = \langle f, \phi_{u,\gamma}\rangle = f * \bar{\phi}_\gamma(u), \quad \bar{\phi}_\gamma(x) = \phi^*_\gamma(-x)$$

Since the index set of $\mathcal{D}$ is $\mathbb{R}^d \times \Gamma$ is not countable, it is thus not strictly speaking a frame by the definition we have utilized up to this point. However, we can consider the energy of the transform $\Phi f(u,\gamma)$, which is defined as

$$\|\Phi f\|^2 = \sum_{\gamma\in\Gamma} \|\Phi f(\cdot,\gamma)\|_2^2 = \sum_{\gamma\in\Gamma} \int |\Phi f(u,\gamma)|^2\, du$$

If there exist $0 < A \le B < \infty$ such that

$$A\|f\|_2^2 \le \sum_{\gamma\in\Gamma} \|\Phi f(\cdot,\gamma)\|_2^2 = \sum_{\gamma\in\Gamma} \|f * \bar{\phi}_\gamma\|_2^2 \le B\|f\|_2^2 \tag{72}$$

then all of the frame theory results we have studied thus far still apply. We will refer to such dictionaries as *semi-discrete frames*, since their index set is the cross product of $\mathbb{R}^d$ and Γ, where Γ is discrete but of course $\mathbb{R}^d$ is not. The next theorem shows that the semi-discrete frame condition (72) is equivalent to a condition on the Fourier transforms of the generators ϕ_γ.

Theorem 6.7. *Let $\{\phi_\gamma\}_{\gamma\in\Gamma} \subset \mathbf{L}^2(\mathbb{R}^d)$ be a family of generator functions. Then there exist $0 < A \le B < \infty$ such that*

$$A \le \sum_{\gamma\in\Gamma} |\widehat{\phi}_\gamma(\omega)|^2 \le B, \quad \textit{for almost every } \omega \in \mathbb{R}^d, \tag{73}$$

if and only if $\mathcal{D} = \{\phi_{u,\gamma}\}_{u\in\mathbb{R}^d,\,\gamma\in\Gamma}$ is a semi-discrete frame with frame bounds A and B. Any family $\{\widetilde{\phi}_\gamma\}_{\gamma\in\Gamma}$ that satisfies

$$\sum_{\gamma\in\Gamma} \widehat{\phi}^*_\gamma(\omega)\widehat{\widetilde{\phi}}_\gamma(\omega) = 1$$

defines a left inverse

$$f = \sum_{\gamma\in\Gamma} \Phi f(\cdot,\gamma) * \widetilde{\phi}_\gamma = \sum_{\gamma\in\Gamma} f * \bar{\phi}_\gamma * \widetilde{\phi}_\gamma$$

and are thus the generators of the dual frame. They are defined in frequency as

$$\widehat{\widetilde{\phi}}_\gamma(\omega) = \frac{\widehat{\phi}_\gamma(\omega)}{\sum_{n\in\Gamma} |\widehat{\phi}_n(\omega)|^2}$$

Proof. Let $H : \mathbf{L}^2(\mathbb{R}^d) \to \mathbf{L}^2(\mathbb{R}^d)$ be defined as $Hf = f * \bar{h}$ for some filter h, where $\bar{h}(x) = h^*(-x)$. We first prove that $H^*g = g * h$. Indeed, using the Parseval formula (Theorem 2.12) and the convolution formula we have:

$$\begin{aligned}
\langle g, Hf\rangle &= \int_{\mathbb{R}^d} g(x)(f * \bar{h})^*(x)\, dx \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{g}(\omega)\widehat{f^*}(\omega)\widehat{\bar{h}}^*(\omega)\, d\omega \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{g}(\omega)\widehat{h}(\omega)\widehat{f^*}(\omega)\, d\omega \\
&= \int_{\mathbb{R}^d} g * h(x) f^*(x)\, dx \\
&= \langle g * h, f\rangle
\end{aligned}$$

Now assume that $\mathcal{D}$ is a semi-discrete frame with frame bounds A and B, and let $\Phi_\gamma f = \Phi f(\cdot, \gamma) = f * \bar{\phi}_\gamma$. Since $\mathcal{D}$ is a semi-discrete frame, each $\Phi_\gamma : \mathbf{L}^2(\mathbb{R}^d) \to \mathbf{L}^2(\mathbb{R}^d)$ and by the above computation $\Phi_\gamma^* g = g * \phi_\gamma$. The analysis operator is $\Phi : \mathbf{L}^2(\mathbb{R}^d) \to \boldsymbol{\ell}^2(\Gamma, \mathbf{L}^2(\mathbb{R}^d))$ which can be written as $\Phi f = (\Phi_\gamma f)_{\gamma\in\Gamma}$. Let $G = (g_\gamma)_{\gamma\in\Gamma} \in \boldsymbol{\ell}^2(\Gamma, \mathbf{L}^2(\mathbb{R}^d))$ and now compute the adjoint of Φ:

$$\begin{aligned}
\langle G, \Phi f\rangle &= \sum_{\gamma\in\Gamma} \langle g_\gamma, \Phi_\gamma f\rangle \\
&= \sum_{\gamma\in\Gamma} \langle \Phi_\gamma^* g_\gamma, f\rangle \\
&= \left\langle \sum_{\gamma\in\Gamma} \Phi_\gamma^* g_\gamma, f \right\rangle \\
&= \left\langle \sum_{\gamma\in\Gamma} g_\gamma * \phi_\gamma, f \right\rangle
\end{aligned}$$

It follows that

$$\Phi^* G = \sum_{\gamma\in\Gamma} g_\gamma * \phi_\gamma$$

and furthermore

$$\Phi^*\Phi f = \sum_{\gamma\in\Gamma} f * \bar{\phi}_\gamma * \phi_\gamma$$

The semi-discrete frame condition (72) is equivalent to

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

We can rewrite $\langle \Phi^*\Phi f, f\rangle$:

$$\begin{aligned}
\langle \Phi^*\Phi f, f\rangle &= \int_{\mathbb{R}^d} \sum_{\gamma\in\Gamma} f * \bar{\phi}_\gamma * \phi_\gamma(x) f^*(x)\, dx \\
&= \sum_{\gamma\in\Gamma} \int_{\mathbb{R}^d} f * \bar{\phi}_\gamma * \phi_\gamma(x) f^*(x)\, dx \\
&= \sum_{\gamma} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \widehat{f}(\omega)\widehat{\phi}^*_\gamma(\omega)\widehat{\phi}_\gamma(\omega)\widehat{f}^*(\omega)\, d\omega \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \left(\sum_{\gamma\in\Gamma} |\widehat{\phi}_\gamma(\omega)|^2\right) d\omega
\end{aligned}$$

Suppose by contradiction there exists $E \subset \mathbb{R}^d$ with finite but nonzero Lebesgue measure, i.e., $0 < |E| < \infty$, and for which

$$\sum_{\gamma} |\widehat{\phi}_\gamma(\omega)|^2 > B, \quad \forall\, \omega \in E$$

Let $\widehat{f}(\omega) = (2\pi)^{d/2}\mathbf{1}_E(\omega)$. We have that $\|\widehat{f}\|^2 = (2\pi)^d|E|$ and thus $f \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|f\|^2 = |E|$. But then

$$\begin{aligned}
\langle \Phi^*\Phi f, f\rangle &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} (2\pi)^d \mathbf{1}_E(\omega) \left(\sum_{\gamma\in\Gamma} |\widehat{\phi}_\gamma(\omega)|^2\right) d\omega \\
&> B \int_E d\omega = B|E| = B\|f\|^2
\end{aligned}$$

which contradicts $\langle \Phi^*\Phi f, f\rangle \leq B\|f\|^2$. A similar argument proves the lower bound, and thus we have shown that

$$\text{for a.e. } \omega \in \mathbb{R}^d, \quad A \leq \sum_{\gamma\in\Gamma} |\widehat{\phi}_\gamma(\omega)|^2 \leq B$$

Now assume that (73) holds. Let $f \in \mathbf{L}^2(\mathbb{R}^d)$ and multiply through by $(2\pi)^{-d}|\widehat{f}(\omega)|^2$ and integrate over $\mathbb{R}^d$ to obtain:

$$\frac{A}{(2\pi)^d} \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2\, d\omega \leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \sum_{\gamma\in\Gamma} |\widehat{\phi}_\gamma(\omega)|^2\, d\omega \leq \frac{B}{(2\pi)^d} \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2\, d\omega$$

which is equivalent to

$$A\|f\|^2 \leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\widehat{f}(\omega)|^2 \sum_{\gamma\in\Gamma} |\widehat{\phi}_\gamma(\omega)|^2\, d\omega \leq B\|f\|^2 \tag{74}$$

We rewrite the inner part:

$$\begin{aligned}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}|\widehat{f}(\omega)|^2\sum_{\gamma\in\Gamma}|\widehat{\phi}_\gamma(\omega)|^2\,d\omega &= \sum_{\gamma\in\Gamma}\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\widehat{f}(\omega)\widehat{\phi}_\gamma^*(\omega)\widehat{f}^*(\omega)\widehat{\phi}_\gamma(\omega)\,d\omega\\ &= \sum_{\gamma\in\Gamma}\int_{\mathbb{R}^d} f*\bar{\phi}_\gamma(x)(f*\bar{\phi}_\gamma)^*(x)\,dx\\ &= \sum_{\gamma\in\Gamma}\|f*\bar{\phi}_\gamma\|^2\\ &= \sum_{\gamma\in\Gamma}\|\Phi f(\cdot,\gamma)\|^2\end{aligned}$$

Plugging this into (74) proves that $\mathcal{D}$ is a semi-discrete frame.

Now let $\{\widetilde{\phi}_\gamma\}_{\gamma\in\Gamma}$ be a family of functions that satisfies

$$\sum_{\gamma\in\Gamma}\widehat{\phi}_\gamma^*(\omega)\widehat{\widetilde{\phi}}_\gamma(\omega)=1 \tag{75}$$

First, it is clear that such functions are defined in frequency as:

$$\widehat{\widetilde{\phi}}(\omega)=\frac{\widehat{\phi}_\gamma(\omega)}{\sum_{n\in\Gamma}|\widehat{\phi}_n(\omega)|^2} \tag{76}$$

by simply plugging (76) into the left hand side of (75) and verifying that the sum is equal to one. Now define

$$g(x)=\sum_{\gamma\in\Gamma}\Phi(\cdot,\gamma)*\widetilde{\phi}_\gamma(x)=\sum_{\gamma\in\Gamma}f*\bar{\phi}_\gamma*\widetilde{\phi}_\gamma(x)$$

The Fourier transform of g is:

$$\widehat{g}(\omega)=\sum_{\gamma\in\Gamma}\widehat{f}(\omega)\widehat{\phi}_\gamma^*(\omega)\widehat{\widetilde{\phi}}_\gamma(\omega)=\widehat{f}(\omega)\sum_{\gamma\in\Gamma}\widehat{\phi}_\gamma^*(\omega)\widehat{\widetilde{\phi}}_\gamma(\omega)=\widehat{f}(\omega)$$

It follows that $g=f$, which completes the proof. □

Exercise 65. Read Section 5.1.5 of *A Wavelet Tour of Signal Processing.*

6.2 Translation Invariant Dyadic Wavelet Transform

Section 5.2 of A Wavelet Tour of Signal Processing.

Recall that a continuous wavelet transform computes

$$Wf(u,s)=\langle f,\psi_{u,s}\rangle=f*\bar{\psi}_s(u),\quad\forall(u,s)\in\mathbb{R}\times(0,\infty) \tag{77}$$

where

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \quad \text{and} \quad \bar{\psi}_s(t) = \frac{1}{\sqrt{s}}\psi^*\left(-\frac{t}{s}\right)$$

The operator W, as defined in (77), does not define an analysis operator of a semi-discrete frame because the scale parameter s takes values over the entire interval $(0, \infty)$, which is not discrete.

A semi-discrete wavelet frame is generated by sampling the scale parameter s along an exponential sequence $\{a^j\}_{j\in\mathbb{Z}}$ for some $a > 1$. In many applications (but not all!), we take $a = 2$. In this case the generating family is $\{\psi_j\}_{j\in\mathbb{Z}}$ with

$$\psi_j(t) = 2^{-j}\psi(2^{-j}t)$$

and the translation invariant dictionary is given by:

$$\mathcal{D} = \{\psi_{u,j}\}_{u\in\mathbb{R},\, j\in\mathbb{Z}}, \quad \psi_{u,j}(t) = \psi_j(t-u) = 2^{-j}\psi(2^{-j}(t-u))$$

The resulting analysis operator defines the dyadic wavelet transform:

$$Wf(u,j) = \langle f, \psi_{u,j}\rangle = f * \bar{\psi}_j(u), \quad \bar{\psi}_j(t) = 2^{-j}\psi^*(-2^{-j}t)$$

Notice that rather than normalizing the dilated wavelets by $2^{-j/2}$, which would be analogous to the normalization $s^{-1/2}$ in the continuous wavelet transform, we normalize by 2^{-j}. This is to simplify the following presentation. It simply means that the normalization preserves the $\mathbf{L}^1$ norm of ψ as opposed to the $\mathbf{L}^2$ norm, that is, $\|\psi_j\|_1 = \|\psi\|_1$. Notice as well that $\widehat{\psi}_j(\omega) = \widehat{\psi}(2^j\omega)$ with this normalization.

Applying Theorem 6.7 shows that $\mathcal{D}$ is a semi-discrete frame if and only if there exists $0 < A \leq B < 0$ such that

$$A \leq \sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2 \leq B, \quad \forall \omega \in \mathbb{R} \setminus \{0\} \tag{78}$$

In this case $W : \mathbf{L}^2(\mathbb{R}) \to \boldsymbol{\ell}^2(\mathbf{L}^2(\mathbb{R}))$ when the scales are restricted to $s = 2^j$. Notice that if ψ is a complex analytic wavelet (meaning that $\widehat{\psi}(\omega) = 0$ for all $\omega \leq 0$), then it is impossible for (78) to hold. We will come back to this in a bit. For now assume that ψ is a real valued wavelet. The standard semi-discrete frame condition, which is equivalent to (78), is written as:

$$A\|f\|_2^2 \leq \sum_{j\in\mathbb{Z}} \|f * \bar{\psi}_j\|_2^2 \leq B\|f\|_2^2$$

Equation (78) shows that if the frequency axis is completely covered by dilated dyadic wavelets, then a dyadic wavelet transform defines a complete and stable representation of $f \in \mathbf{L}^2(\mathbb{R})$; see Figure 32.

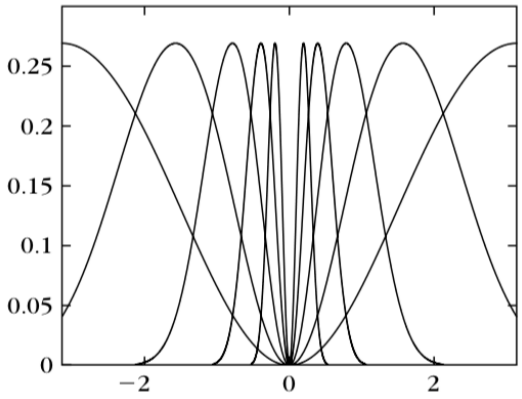


Figure 32: The squared Fourier transform modulus $|\widehat{\psi}(2^j\omega)|^2$ of a real valued spline wavelet, for $1 \leq j \leq 5$ and $\omega \in [-\pi, \pi]$.

Remark 6.8. Recall for the continuous wavelet transform, we had the following admissibility condition for a real valued wavelet:

$$C_\psi = \int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|^2}{\omega} < \infty$$

In fact (78) is closely related to the admissibility condition, as the following calculation shows (let $\omega_0 > 0$):

$$\begin{aligned} C_\psi &= \int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|^2}{\omega}\, d\omega\,, \quad (\text{CoV: } \omega = 2^\lambda \omega_0 \Rightarrow d\omega = (\log 2) 2^\lambda \omega_0 d\lambda) \\ &= \int_{\mathbb{R}} \frac{|\widehat{\psi}(2^\lambda \omega_0)|^2}{2^\lambda \omega_0} (\log 2) 2^\lambda \omega_0\, d\lambda \\ &= (\log 2) \int_{\mathbb{R}} |\widehat{\psi}(2^\lambda \omega_0)|^2\, d\lambda \end{aligned}$$

Note that $\omega_0 > 0$ was arbitrary and since $|\widehat{\psi}(\omega)| = |\widehat{\psi}(-\omega)|$ for any ω, in fact it holds for any $\omega_0 \neq 0$. Thus we see (78) is a discrete version of the wavelet admissibility condition. This calculation also explains why switching to a an $\mathbf{L}^1(\mathbb{R})$ normalization for the wavelet is a good idea.

In the case of complex analytic wavelets, one option is to use a larger set of generating wavelets given by:

$$\{\psi_{j,\epsilon}\}_{j\in\mathbb{Z},\, \epsilon\in\{1,-1\}}, \quad \psi_{j,\epsilon}(t) = 2^{-j}\psi(\epsilon 2^{-j} t)$$

In this case for suitably chosen wavelets it is possible for (78) to hold. However, it is unnecessary to double the number of generating wavelets as in the above. Indeed, we can instead replace (78) with

$$2A \leq \sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2 + \sum_{j\in\mathbb{Z}} |\widehat{\psi}(-2^j\omega)|^2 \leq 2B, \quad \forall\, \omega \in \mathbb{R} \setminus \{0\} \tag{79}$$

which, due to the wavelet ψ being complex analytic, is equivalent to

$$2A \leq \sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2 \leq 2B, \quad \forall\, \omega \in (0,\infty)$$

Let $f \in \mathbf{L}^2(\mathbb{R})$ be real valued and let f_a be the analytic part of f. Recall that $\widehat{f_a}(\omega) = 2\widehat{f}(\omega)$ for $\omega > 0$ and $2\|f\|_2^2 = \|f_a\|_2^2$. Then:

$$\begin{aligned}
\sum_{j\in\mathbb{Z}} \|f * \bar{\psi}_j\|_2^2 &= \sum_{j\in\mathbb{Z}} \int_{\mathbb{R}} |f * \bar{\psi}_j(t)|^2\, dt \\
&= \frac{1}{2\pi} \sum_{j\in\mathbb{Z}} \int_{\mathbb{R}} |\widehat{f}(\omega)|^2 |\widehat{\psi}(2^j\omega)|^2\, d\omega \\
&= \frac{1}{2\pi} \int_0^{+\infty} |\widehat{f}(\omega)|^2 \sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2\, d\omega \\
&= \frac{1}{4}\frac{1}{2\pi} \int_0^{+\infty} |\widehat{f_a}(\omega)|^2 \sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2\, d\omega \\
&\geq \frac{A}{2}\frac{1}{2\pi} \int_0^{+\infty} |\widehat{f_a}(\omega)|^2\, d\omega \\
&= \frac{A}{2}\|f_a\|_2^2 \\
&= A\|f\|_2^2
\end{aligned}$$

A similar argument shows that $\sum_j \|f * \bar{\psi}_j\|_2^2 \leq B\|f\|_2^2$. Therefore the dyadic wavelet transform with a complex analytic wavelet defines a semi-discrete frame with frame bounds A and B if (79) holds.

Now suppose we only want to compute the dyadic wavelet transform up to a maximum scale 2^j for $j < J$. The lost low frequency information is captured by a single scaling function (or low pass filter) whose Fourier transform is concentrated around the origin. Let $\phi \in \mathbf{L}^2(\mathbb{R})$ be a low pass filter and let $\phi_J(t) = 2^{-J}\phi(2^{-J}t)$ and let ψ be a real valued wavelet. The dyadic wavelet transform in this case is defined as:

$$W_J f = \{f * \bar{\phi}_J(u),\ f * \bar{\psi}_j(u)\}_{u\in\mathbb{R},\, j<J}$$

The operator W_J is the analysis operator of a semi-discrete frame if

$$A \leq |\widehat{\phi}(2^J\omega)|^2 + \sum_{j<J} |\widehat{\psi}(2^j\omega)|^2 \leq B$$

If the family $\{\psi_j\}_{j\in\mathbb{Z}}$ are the generators of a semi-discrete frame, meaning that (78) holds, then one can define ϕ in frequency as:

$$|\widehat{\phi}(\omega)|^2 = \begin{cases} (A+B)/2, & \omega = 0 \\ \sum_{j\geq 0} |\widehat{\psi}(2^j\omega)|^2, & \omega \neq 0 \end{cases}$$

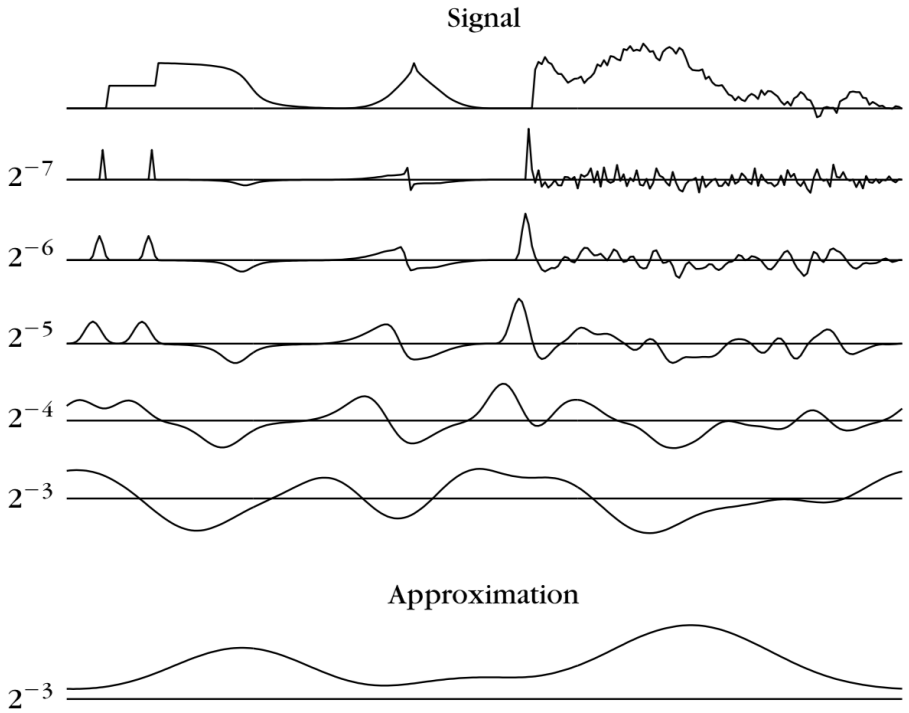


Figure 33: The dyadic wavelet transform $W_J f$ computed with $J = -2$ and $-7 \leq j \leq -3$. The top curve is $f(t)$, the next five curves are $f * \bar{\psi}_j(u)$, and the bottom curve is $f * \bar{\phi}_J$.

Figure 33 plots the dyadic wavelet transform $W_J f$ for the signal f from Figure 13.

A dual wavelet for a semi-discrete dyadic wavelet frame (without scaling function) is computed in frequency as:

$$\widehat{\widetilde{\psi}}(\omega) = \frac{\widehat{\psi}(\omega)}{\sum_{k\in\mathbb{Z}} |\widehat{\psi}(2^k\omega)|^2}$$

and the generators of the dual semi-discrete dictionary are given by the dilations of $\widetilde{\psi}$, namely $\{\widetilde{\psi}_j\}_{j\in\mathbb{Z}}$. From this definition it follows that the Fourier transform of $\widetilde{\psi}_j$ satisfies:

$$\widehat{\widetilde{\psi}}_j(\omega) = \widehat{\widetilde{\psi}}(2^j\omega) = \frac{\widehat{\psi}(2^j\omega)}{\sum_{k\in\mathbb{Z}} |\widehat{\psi}(2^{j+k}\omega)|^2} = \frac{\widehat{\psi}(2^j\omega)}{\sum_{k\in\mathbb{Z}} |\widehat{\psi}(2^k\omega)|^2}$$

We thus have

$$\sum_{j\in\mathbb{Z}} \widehat{\psi}_j^*(\omega)\widehat{\widetilde{\psi}}_j(\omega) = \sum_{j\in\mathbb{Z}} \widehat{\psi}^*(2^j\omega)\widehat{\widetilde{\psi}}(2^j\omega) = 1, \quad \forall\, \omega \in \mathbb{R}\setminus\{0\}$$

and so by Theorem 6.7 the following reconstruction formula holds:

$$f(t) = \sum_{j\in\mathbb{Z}} f * \bar{\psi}_j * \widetilde{\psi}_j(t)$$

Things are simplified when the semi-discrete dyadic wavelet frame is tight. In this case

$$\widetilde{\psi}_{u,j}(t) = \frac{1}{A}\psi_{u,j}(t) = \frac{1}{A}2^{-j}\psi(2^{-j}(t-u))$$

and signal synthesis is computed as:

$$f(t) = \frac{1}{A}\sum_{j\in\mathbb{Z}} f * \bar{\psi}_j * \psi_j(t)$$

Notice that we must then have

$$\sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2 = A\,, \quad \forall\, \omega \in \mathbb{R}\setminus\{0\}$$

if the wavelet ψ is real valued (with a similar condition for complex analytic wavelets). This is a Littlewood-Paley type condition, and implies the Fourier transforms of the dilations of the wavelet ψ evenly cover the frequency axis.

Exercise 66. Read Section 5.2 of *A Wavelet Tour of Signal Processing.*

Exercise 67. Read Section 5.3 of *A Wavelet Tour of Signal Processing.*

Exercise 68. Read Section 5.4 of *A Wavelet Tour of Signal Processing.*

Exercise 69. Let h be a filter with $\widehat{h}(0) = \sqrt{2}$ and let $\phi \in \mathbf{L}^2(\mathbb{R})$ be a low pass filter with the following Fourier transform:

$$\widehat{\phi}(\omega) = \frac{1}{\sqrt{2}}\widehat{h}(\omega/2)\widehat{\phi}(\omega/2)$$

Let g be a filter with $\widehat{g}(0) = 0$ and let ψ be a wavelet with Fourier transform:

$$\widehat{\psi}(\omega) = \frac{1}{\sqrt{2}}\widehat{g}(\omega/2)\widehat{\phi}(\omega/2)$$

Prove that if there exist $0 < A \leq B < \infty$ such that

$$A(2 - |\widehat{h}(\omega)|^2) \leq |\widehat{g}(\omega)|^2 \leq B(2 - |\widehat{h}(\omega)|^2)$$

then the family $\{\psi_j\}_{j\in\mathbb{Z}}$ are the generators of a semi-discrete frame.

Exercise 70. Let $X = (X(t))_{t\in\mathbb{R}}$ be a second order stationary stochastic process with continuous sample paths. Let ψ be a real valued, continuous, compactly supported wavelet for which

$$\sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2 = 1\,, \quad \forall\, \omega \neq 0$$

Prove:

$$\sum_{j\in\mathbb{Z}} \mathbb{E}\left[|X * \overline{\psi}_j(t)|^2\right] = \mathrm{Var}_X(0) = \mathbb{E}[(X(0) - m_X)^2]\,, \quad \forall\, t \in \mathbb{R}$$

Exercise 71. This exercise is about representations of signals f that are invariant to translation of f.

(a) Let $f \in \mathbf{L}^2(\mathbb{R})$ and let $\phi \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{C}^1(\mathbb{R})$ be a low pass filter with $\phi' \in \mathbf{L}^1(\mathbb{R})$. Let $f_u(t) = f(t-u)$ be the translation of f by u. Prove there exists a universal constant $C > 0$ such that:

$$\|f * \phi_J - f_u * \phi_J\|_2 \leq C2^{-J}|u|\|\phi'\|_1\|f\|_2$$

(b) Let ϕ be as in part (a) and suppose $\psi \in \mathbf{L}^1(\mathbb{R})$ is a wavelet for which

$$|\widehat{\phi}(2^J\omega)|^2 + \sum_{j<J} |\widehat{\psi}(2^j\omega)|^2 = 1\,, \quad \forall\, \omega \in \mathbb{R}$$

Part (a) shows that $f * \phi_J$ is a representation of f that is invariant to translations of f so long as $|u| \ll 2^J$. However, $f * \phi_J$ only keeps the low frequencies of f. A representation that keeps more information from f is:

$$S_J f = \{f * \phi_J\,, |f * \psi_j| * \phi_J : j < J\} \in \boldsymbol{\ell}^2(\mathbf{L}^2(\mathbb{R}))$$

Prove this representation is also translation invariant in the same sense, meaning there exists a constant $C > 0$ such that:

$$\|S_J f - S_J f_u\|_{\boldsymbol{\ell}^2(\mathbf{L}^2(\mathbb{R}))} \leq C2^{-J}|u|\|\phi'\|_1\|f\|_2$$

Exercise 72. THIS EXERCISE IS OPTIONAL! JUST IF YOU WANT AN ADDITIONAL CHALLENGE. The *Zak transform* maps any $f \in \mathbf{L}^2(\mathbb{R})$ to:

$$Zf(u,\xi) = \sum_{l\in\mathbb{Z}} e^{2\pi i l\xi} f(u-l)$$

(a) Prove that $Z : \mathbf{L}^2(\mathbb{R}) \to \mathbf{L}^2[0,1]^2$ is a unitary operator, i.e. show that

$$\int_{\mathbb{R}} f(t)g^*(t)\,dt = \int_0^1 \int_0^1 Zf(u,\xi)Zg^*(u,\xi)\,du\,d\xi$$

One approach is the following: Let $g(t) = \mathbf{1}_{[0,1]}(t)$ and consider

$$\mathcal{B} = \{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}, \quad g_{n,k}(t) = g(t-n)e^{2\pi ikt}$$

Verify that $\mathcal{B}$ is an orthonormal basis for $\mathbf{L}^2(\mathbb{R})$, and then show that $\{Zg_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is an orthonormal basis for $\mathbf{L}^2[0,1]^2$.

(b) Prove that the inverse Zak transform is defined by:

$$Z^{-1}h(u) = \int_0^1 h(u,\xi)\,d\xi, \quad \forall\, h \in \mathbf{L}^2[0,1]^2$$

(c) Now let $g \in \mathbf{L}^2(\mathbb{R})$ be arbitrary and consider

$$\mathcal{D} = \{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}, \quad g_{n,k}(t) = g(t-n)e^{2\pi ikt}$$

Prove that $\mathcal{D}$ is a frame for $\mathbf{L}^2(\mathbb{R})$ with frame bounds $0 < A \le B < \infty$ if and only if

$$A \le |Zg(u,\xi)|^2 \le B, \quad \forall\,(u,\xi) \in [0,1]^2 \tag{80}$$

(d) Prove that if (80) holds, then the dual window $\widetilde{g}$ of the dual frame $\widetilde{D}$ is defined by

$$Z\widetilde{g}(u,\xi) = \frac{1}{Zg^*(u,\xi)}$$

Math 994-002: Applied and Computational Harmonic Analysis, MSU, Spring 2020

Lecture 27 & 28: 2D Directional Wavelet Frames

April 21 & 23, 2020

Lecturer: Matthew Hirn

6.2.1 Dyadic Maxima Representation

Section 6.2.2 of A Wavelet Tour of Signal Processing

Let us now return to the analysis of pointwise singularities of signals f via the decay of $Wf(u,s)$ as $s \to 0$. Let ψ be a real valued wavelet, and recall that a wavelet modulus maxima is defined as a point (u_0, s_0) such that $|Wf(u,s_0)|$ is locally maximum at $u = u_0$.

All of the results regarding wavelet coefficient decay and the pointwise regularity of $f(t)$ (including, in particular, Theorems 5.5, 5.7, and 5.8) hold for dyadic wavelet semi-discrete frames by restricting $s = 2^j$ for $j \in \mathbb{Z}$. Let (u_0, j) be a modulus maxima point of $Wf(u,j)$, meaning that

$$\frac{\partial Wf}{\partial u}(u_0, j) = 0 \tag{81}$$

Since $Wf(u,j) = f * \bar{\psi}_j(u)$, $\bar{\psi}_j(t) = 2^{-j}\psi(-2^{-j}t)$ and

$$\frac{d}{dt}\bar{\psi}_j(t) = -2^{-j}2^{-j}\psi'(-2^{-j}t) = -2^{-j}\overline{\psi'}_j(t)$$

equation (81) is equivalent to

$$f * \overline{\psi'}_j(u_0) = 0$$

Figure 34 shows the dyadic wavelet transform of a signal and the corresponding wavelet modulus maxima.

Let Λ denote the wavelet modulus maxima of f:

$$\Lambda = \{(u,j) \in \mathbb{R} \times \mathbb{Z} : f * \overline{\psi'}_j(u) = 0\}$$

Recall that the dictionary $\mathcal{D}$ of a dyadic wavelet transform is:

$$\mathcal{D} = \{\psi_{u,j}\}_{(u,j)\in\mathbb{R}\times\mathbb{Z}}$$

The set Λ defines a sub-dictionary of $\mathcal{D}$:

$$\mathcal{D}_\Lambda = \{\psi_{u,j}\}_{(u,j)\in\Lambda}$$

Furthermore, the completion of the span of $\mathcal{D}_\Lambda$ defines a closed subspace $\mathbf{V}_\Lambda$ of $\mathbf{L}^2(\mathbb{R})$:

$$\mathbf{V}_\Lambda = \overline{\operatorname{span}\mathcal{D}_\Lambda}$$

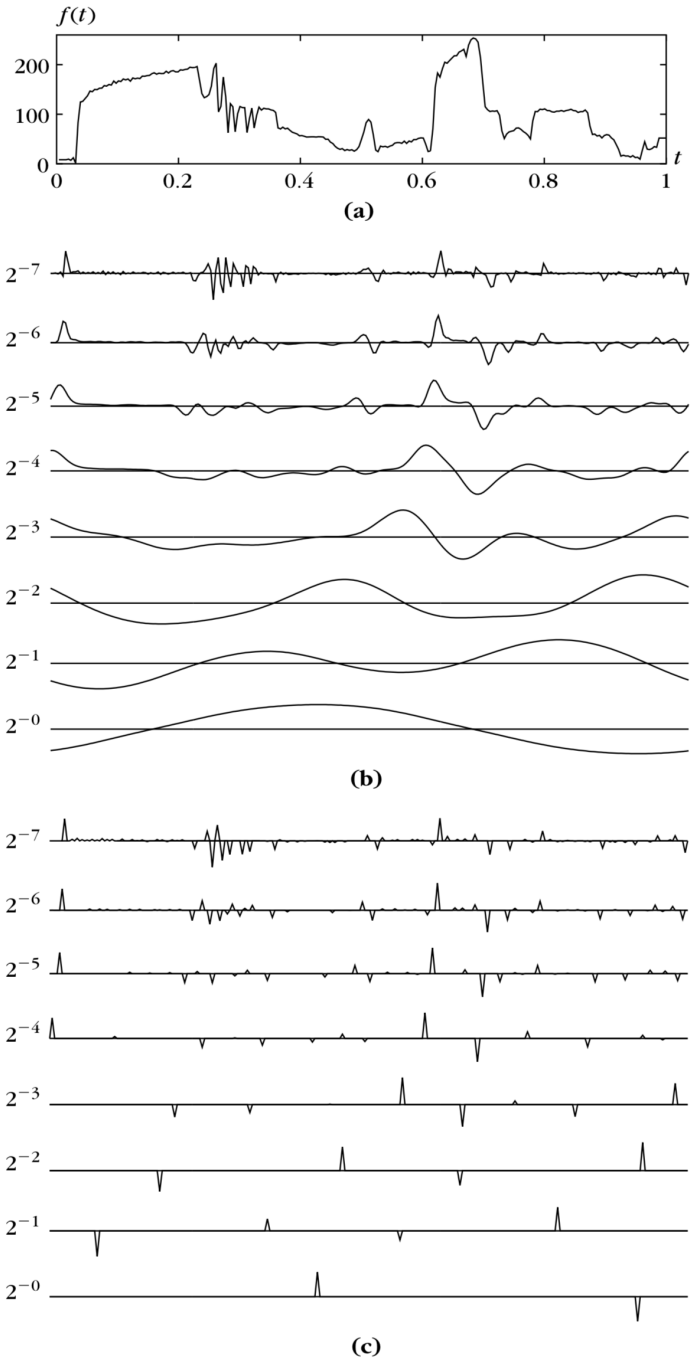


Figure 34: (a) The signal $f(t)$. (b) Dyadic wavelet transform computed with a wavelet $\psi = -\theta'$. (c) Modulus maxima of the dyadic wavelet transform.

We can therefore project f onto $\mathbf{V}_\Lambda$. Doing so amounts to computing an approximation f_Λ of f which is the signal synthesized from only the wavelet modulus maxima of f. It is computed with a dual synthesis as:

$$f_\Lambda = P_{\mathbf{V}_\Lambda} f = \sum_{(u,j)\in\Lambda} \langle f, \psi_{u,j}\rangle \widetilde{\psi}_{u,j}$$

For general dyadic wavelets, $f_\Lambda \neq f$. However, signals with the same modulus maxima differ from each other by small amplitude errors introducing no oscillations, so in numerical experiments $f_\Lambda \approx f$. If f is band-limited (meaning it has a compactly supported Fourier transform) and ψ is as well, then the wavelet modulus maxima define a complete representation of f and in this case $f_\Lambda = f$.

Figure 35 computes the projection f_Λ for the signal first introduced in Figure 34. The signal is not bandlimited, so the reconstruction is not perfect. However, Figure 35(b) shows that the approximation is quite good, and the relative error is approximately 2.5%. Figure 35 reconstructs the signal using only the top 50% of the wavelet modulus maxima coefficients. The sharpest signal transitions have been preserved, since they have the largest amplitude responses, however small texture variations are removed since the wavelet modulus maxima there have relatively small amplitudes. The resulting signal appears to be piecewise regular. In either case, we have achieve a lossy compression of f in which there is a large compression and the loss is not too large.

Exercise 73. Read Section 6.2.2 of *A Wavelet Tour of Signal Processing.*

6.3 Multiscale Directional Frames for Images

Section 5.5 of A Wavelet Tour of Signal Processing.

6.3.1 Directional Wavelet Frames

Section 5.5.1 of A Wavelet Tour of Signal Processing.

We now consider two dimensional wavelet semi-discrete frames for image analysis. Such semi-discrete frames are constructed with wavelets that have directional sensitivity, providing information on the direction of sharp transitions such as edges and textures.

Let $x = (x_1, x_2) \in \mathbb{R}^2$. A directional wavelet $\psi_\alpha(x)$ of angle $\alpha \in [0, 2\pi)$ is a wavelet having p directional vanishing moments along any one dimensional line of direction $\alpha + \pi/2$ in the plane but does not have directional vanishing moments along the direction α. The former condition means that:

$$\int_{\mathbb{R}} \psi_\alpha(\rho\cos\alpha - u\sin\alpha, \rho\sin\alpha + u\cos\alpha) u^k \, du = 0, \quad \forall\, \rho \in \mathbb{R},\ 0 \leq k < p$$

Such a wavelet oscillates in the direction $\alpha + \pi/2$ but not in the direction α.

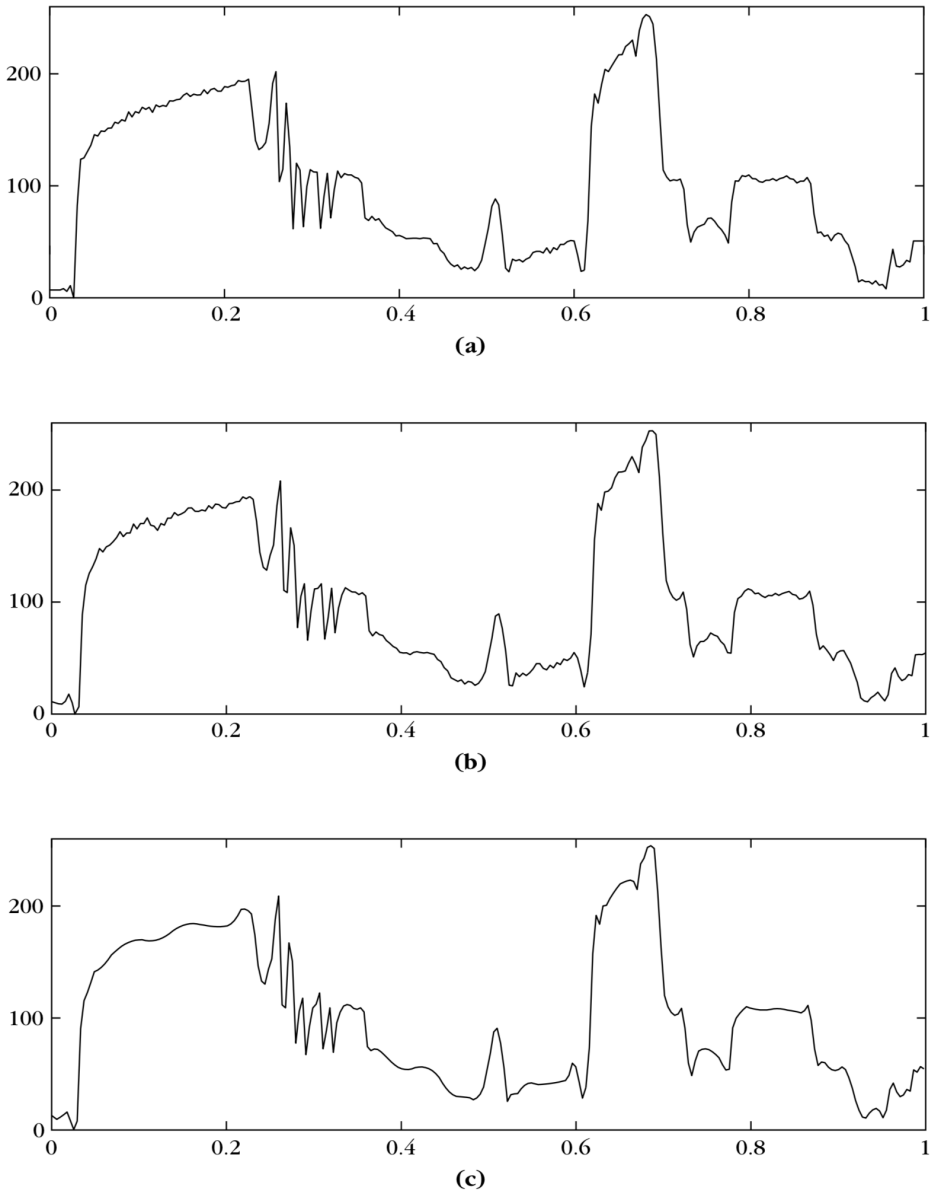


Figure 35: (a) The signal $f(t)$. (b) Signal approximation $f_\Lambda(t)$ using the dyadic wavelet modulus maxima shown in Figure 34. (c) Approximation recovered using only the largest 50% of the wavelet modulus maxima.

Let $\Theta \subset [0, \pi)$ denote the set of angles α. Typically Θ is a uniform sampling:

$$\Theta = \{\alpha = 2\pi k/K : 0 \leq k < K\}$$

The generators of a translation invariant dictionary are the dyadic dilations of each directional wavelet:

$$\{\psi_{j,\alpha}\}_{j\in\mathbb{Z},\,\alpha\in\Theta}, \quad \psi_{j,\alpha}(x) = 2^{-2j}\psi_\alpha(2^{-j}x)$$

Often the directional wavelets ψ_α are obtained by rotating a single mother wavelet ψ; we will come back to this shortly when we define two dimensional Gabor and Morlet wavelets. For real valued directional wavelets, Theorem 6.7 proves that the generating wavelets generate a semi-discrete frame if and only if there exists $0 < A \leq B < \infty$ such that

$$A \leq \sum_{j\in\mathbb{Z}} \sum_{\alpha\in\Theta} |\widehat{\psi}_\alpha(2^j\omega)|^2 \leq B, \quad \forall\, \omega \in \mathbb{R}^2 \setminus \{(0,0)\}$$

If the generating wavelets ψ_α are complex valued analytic wavelets, then they generate a semi-discrete frame if and only if

$$2A \leq \sum_{j\in\mathbb{Z}} \sum_{\alpha\in\Theta} |\widehat{\psi}_\alpha(2^j\omega)|^2 + \sum_{j\in\mathbb{Z}} \sum_{\alpha\in\Theta} |\widehat{\psi}_\alpha(-2^j\omega)|^2 \leq 2B, \quad \forall\, \omega \in \mathbb{R}^2 \setminus \{(0,0)\} \tag{82}$$

When the above semi-discrete frame conditions holds, the dyadic directional wavelet transform is a map $W : \mathbf{L}^2(\mathbb{R}^2) \to \boldsymbol{\ell}^2(\mathbf{L}^2(\mathbb{R}^2))$ defined as:

$$Wf = \{f * \bar{\psi}_{j,\alpha}(u) : j \in \mathbb{Z},\, \alpha \in \Theta,\, u \in \mathbb{R}^2\}, \quad \bar{\psi}_{j,\alpha}(x) = \psi^*_{j,\alpha}(-x)$$

A wavelet $\psi_{u,j,\alpha}(x) = \psi_{j,\alpha}(x-u)$ has support dilated by 2^j, located in a neighborhood of u and oscillates in the direction $\alpha + \pi/2$. If $f(x)$ is constant over the support of $\psi_{j,\alpha}(x-u)$ along lines of direction $\alpha + \pi/2$, then $f * \bar{\psi}_{j,\alpha}(u) = 0$ because of its directional vanishing moments. In particular, the wavelet coefficient vanishes in the neighborhood of an edge having a tangent in the direction of $\alpha + \pi/2$. If the edge angle deviates from $\alpha + \pi/2$, then it produces large amplitude coefficients, with a maximum typically when the edge has direction α. Figure 36 illustrates the idea.

Two dimensional Gabor wavelets are directional wavelets generated from a single complex valued mother wavelet. The mother wavelet is defined as:

$$\psi(x) = g_\sigma(x) e^{i\xi\cdot x}$$

where $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$ is the central frequency of ψ and $g(x)$ is a Gaussian, which we take as

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

The mother Gabor wavelet oscillates along the angle $\arccos(\xi_1/|\xi|)$, and thus in the more general language of directional wavelets it has angle $\alpha = \arccos(\xi_1/|\xi|) - \pi/2$. The Fourier transform of ψ is

$$\widehat{\psi}(\omega) = e^{-\sigma^2|\omega-\xi|^2/2}$$

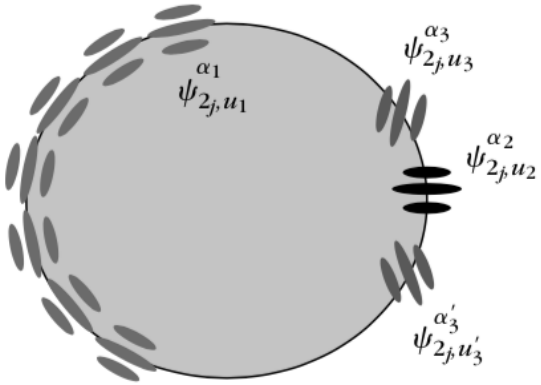


Figure 36: A cartoon image of a disk, with a regular edge. When the wavelet direction α is orthogonal to the tangent of the edge, the coefficients vanish as indicated by the black wavelet response (α_2). When the wavelet direction α aligns with the tangent of the curve (as on the left with α_1), the wavelet coefficients have large amplitude. When the tangent of the curve is not aligned with the wavelet, but is not orthogonal either (as in α_3 and α'_3), wavelet coefficients may have non-negligible amplitude but generally not as large as the α_1 coefficients.

For appropriate choices of $\sigma \in \mathbb{R}$ and $\xi \in \mathbb{R}^2$ the wavelet has nearly zero average and is almost analytic. The generators of a Gabor wavelet semi-discrete frame are obtained from dilations and rotations of the mother wavelet:

$$\psi_{j,\theta}(x) = 2^{-2j}\psi(2^{-j}R_\theta^{-1}x), \quad j \in \mathbb{Z},\, \theta \in \Theta$$

where R_θ is the two dimensional rotation matrix by the angle θ,

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

A simple computation shows that

$$\psi_{j,\theta}(x) = g_{2^j\sigma}(x)e^{i2^{-j}R_\theta\xi\cdot x}$$

Thus $\psi_{j,\theta}$ changes the essential support of ψ from a ball of radius σ to a ball of radius $2^j\sigma$, the direction of oscillation is rotated by θ radians, and the magnitude of the frequency of this oscillation is now $2^{-j}|\xi|$. In frequency we have:

$$\widehat{\psi}_{j,\theta}(\omega) = \widehat{\psi}(2^jR_\theta^{-1}\omega) = e^{-(2^j\sigma)^2|\omega-2^{-j}R_\theta\xi|^2/2}$$

Thus the essential support of $\widehat{\psi}_{j,\theta}(\omega)$ is a ball of radius $(2^j\sigma)^{-1}$ centered at $2^{-j}R_\theta\xi$. These frequency supports will cover the upper half plane for appropriate choices of K (the number

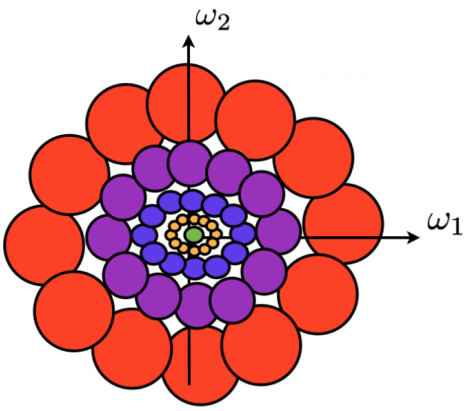


Figure 37: Frequency supports of the two dimensional oriented Gabor wavelets. Scales 2^j are indicated with different colors. The central green ball corresponds to the frequency support of a two dimensional scaling function.

of angles θ) and ξ; reflections will then cover the lower half plane, and so (82) will be satisfied. Figure 37 illustrates this frequency covering.

As in one dimension, Gabor wavelets can be amended to have precisely zero average:

$$\psi(x) = g_\sigma(x)(e^{i\xi\cdot x} - C), \quad C \text{ chosen so that } \int_{\mathbb{R}^2} \psi(x)\,dx = 0$$

which gives a directional Morlet wavelet. Figure 38 plots the real part of a Morlet wavelet at different scales and orientations; Figure 39 plots the imaginary part; and Figure 40 plots their Fourier transforms.

A dyadic Gabor/Morlet wavelet transform computes:

$$Wf = \{f * \psi_{j,\theta}(u) : j \in \mathbb{Z},\, \theta \in \Theta,\, u \in \mathbb{R}^2\}$$

Figure 41 shows the result of computing the dyadic Gabor wavelet transform of an image consisting of one texture embedded in another texture. The middle texture is relatively smooth along vertical lines, but has significant variations in the horizontal direction. It results that a Gabor wavelet transform with two directional angles $\theta = 0$ and $\theta = \pi/2$ will have large magnitude responses for $\theta = 0$, and negligible response for $\theta = \pi/2$. The outside texture, on the other hand, has the most variation along the angle $\alpha = \pi/4$. The Gabor wavelet coefficients along the directions $\theta = 0, \pi/2$ should be negligible for this outside texture, and indeed they are.

Figure 42 computes the Morlet wavelet transform of a black and white image of a butterfly, where the directional edge detection properties of the transform are exhibited, particularly in the wing of the butterfly.

As a point of comparison, we can define a non-directional wavelet $\psi(x)$ as the Laplacian of a Gaussian:

$$\psi(x) = -(\Delta g)(x)\,, \quad g(x) = \frac{1}{2\pi\sigma^2} e^{-|x|^2/2\sigma^2}$$

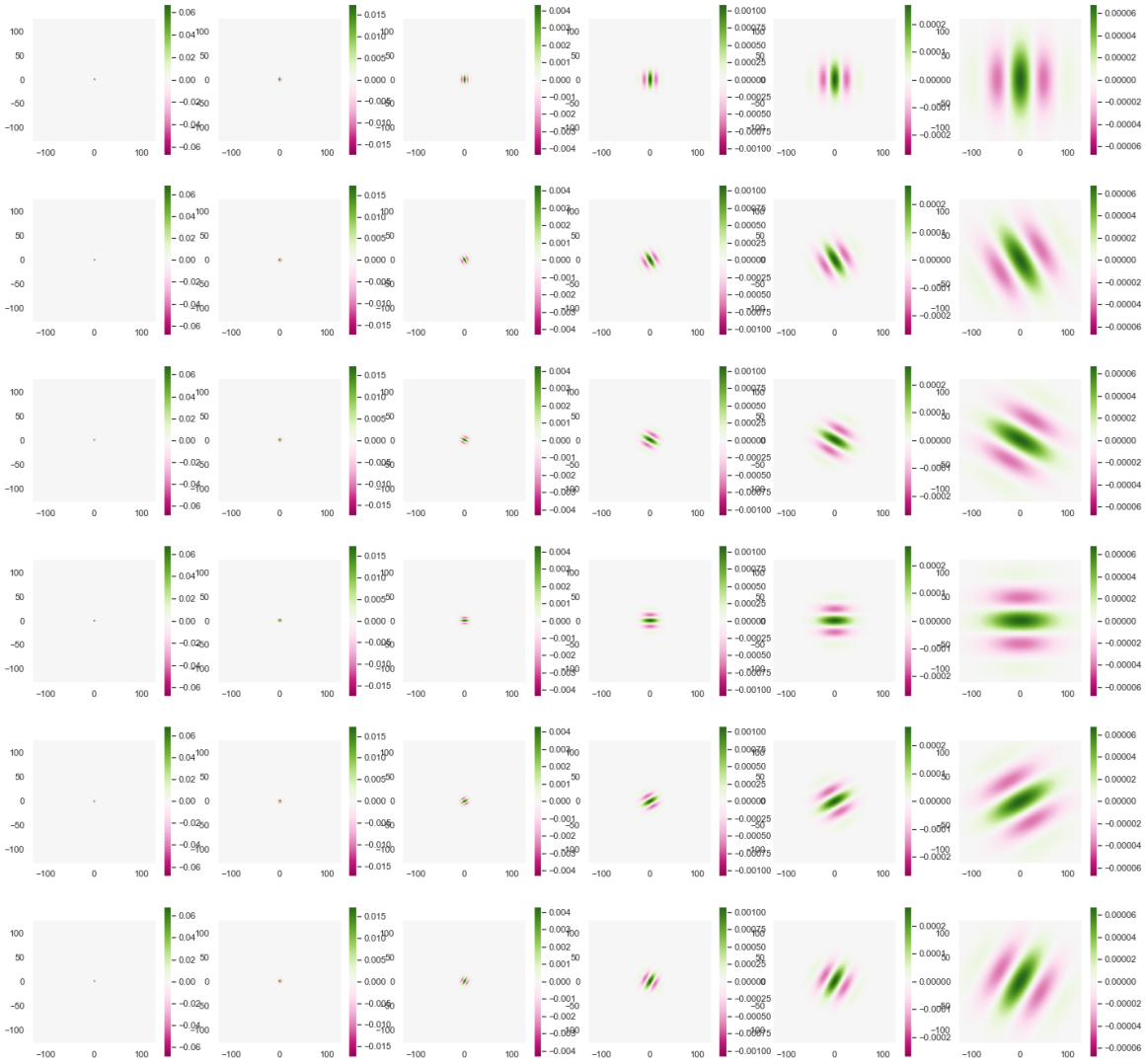


Figure 38: Real part of Morlet wavelets. Increasing scale left to right, and increasing angle in $[0, \pi)$ from top to bottom. Green is positive, pink is negative, and white is zero.

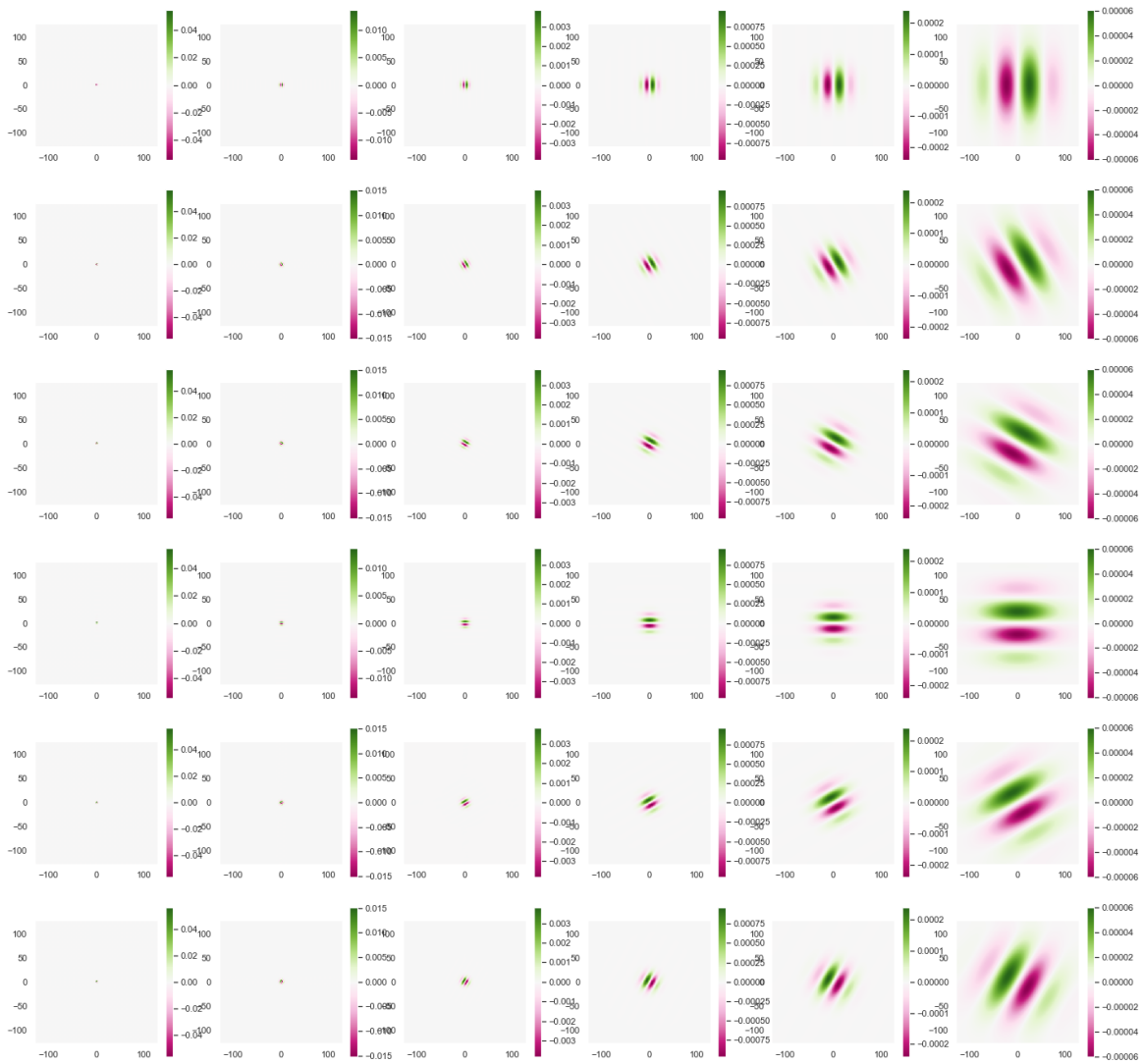


Figure 39: Imaginary part of Morlet wavelets. Increasing scale left to right, and increasing angle in $[0, \pi)$ from top to bottom. Green is positive, pink is negative, and white is zero.

Figure 40: Fourier transforms of Morlet wavelets. Increasing scale left to right, and increasing angle in $[0, \pi)$ from top to bottom. Green is positive, pink is negative, and white is zero.

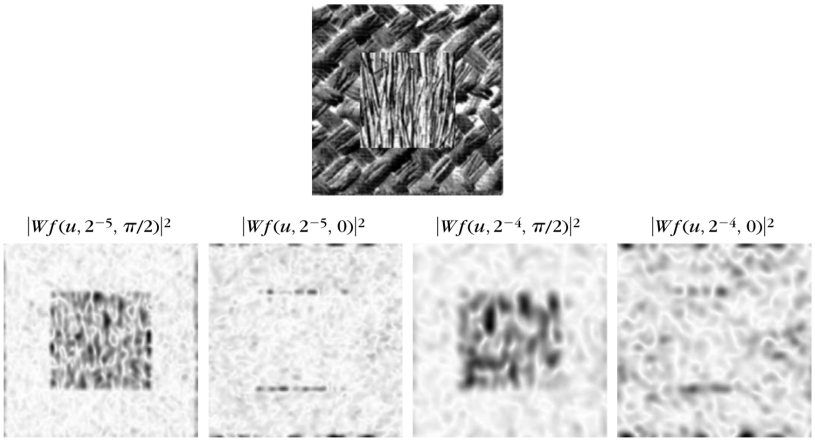


Figure 41: Top: Image of one texture embedded in another texture. Bottom: The magnitude of the Gabor wavelet transform $|Wf(u, j, \theta)| = |f * \psi_{j,\theta}(u)|$ for $j = -4, -5$ and $\theta = 0, \pi/2$. Images taken from Figure 5.9 of [1].

Note we have

$$\widehat{\psi}(\omega) = |\omega|^2 \widehat{g}(\omega) = |\omega|^2 e^{-\sigma^2 |\omega|^2/2}$$

and so ψ is indeed a wavelet, and is radially symmetric, meaning it has no directionality. The wavelet transform for this wavelet computes:

$$Wf = \{f * \psi_j(u) : j \in \mathbb{Z}\,,\, u \in \mathbb{R}^2\}$$

Even though this wavelet has no directionality, it can still pick up edges at small scales and meso- and macroscopic patterns at larger scales, agnostic of the direction. See Figure 43 for pictures and more details.

Exercise 74. Read Section 5.5.1 of *A Wavelet Tour of Signal Processing.*

Exercise 75. Read Section 5.5.2 of *A Wavelet Tour of Signal Processing.*

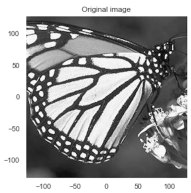
6.4 Multiscale Edge Detection

Section 6.3 of A Wavelet Tour of Signal Processing.

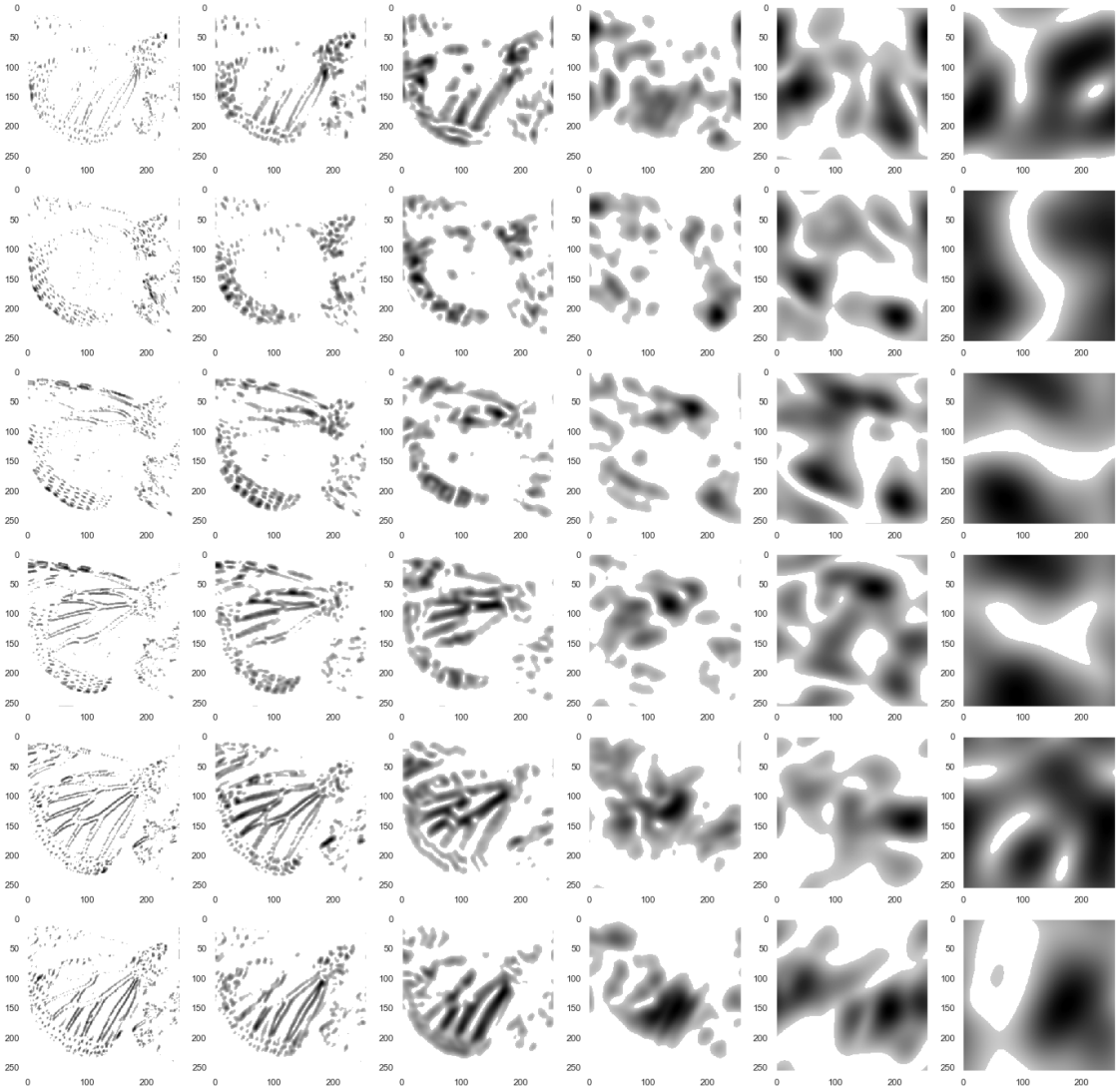
6.4.1 Wavelet Maxima for Images

Section 6.3.1 of A Wavelet Tour of Signal Processing.

Taken directly from [1]: *Image edges are often important for pattern recognition. This is clearly illustrated by our visual ability to recognize an object from a drawing that gives a*



(a) Butterfly



(b) Modulus of Morlet wavelet coefficients, $|f * \psi_{j,\theta}(u)|$.

Figure 42: (a) Black and white image of a butterfly. (b) Modulus of the Morlet wavelet coefficients, with increasing scale from left to right and increasing angle from top to bottom. Small wavelet coefficients have been set to zero to better illustrate the large amplitude coefficients.

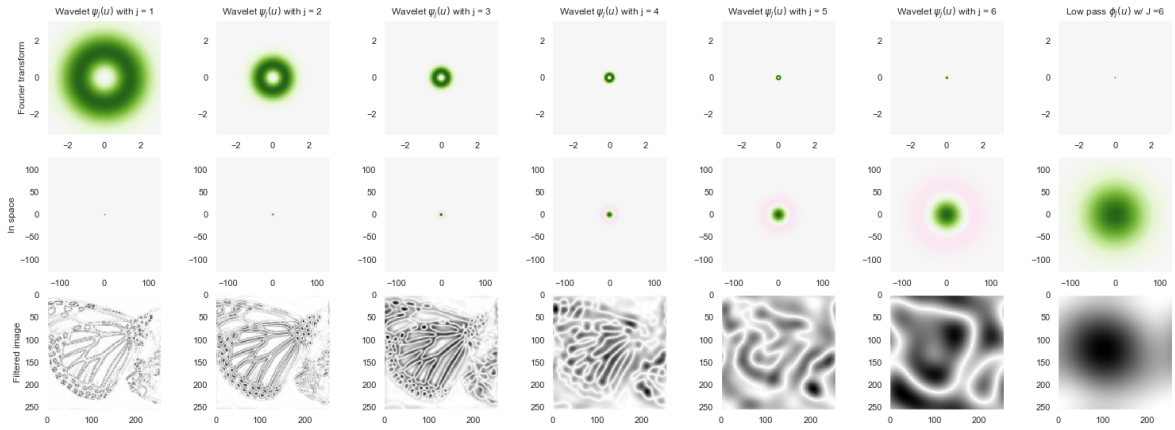


Figure 43: Non-directional wavelet transform based upon $\psi(x) = -(\Delta g)(x)$, for $g(x)$ a Gaussian. Scale increases from left to right; green is positive, pink is negative, white is zero in the first two rows. In the last row, grey/black is positive, white is zero. Top row: The Fourier transforms $\widehat{\psi}_j(\omega)$, which are like donuts. Middle row: The wavelets $\psi_j(x)$, which oscillate radially out from the origin. Bottom row: The absolute value of the wavelet coefficients, $|f * \psi_j(u)|$, for the butterfly image from Figure 42(a). Notice at small scales the wavelet still detects edges, but agnostic to the direction.

rough outline of contours. But, what is an edge? It could be defined as points where the image intensity has sharp transitions. A closer look shows that this definition is often not satisfactory. Image textures do have sharp intensity variations that are often not considered as edges. When looking at a brick wall, we may decide that the edges are the contours of the wall whereas the bricks define a texture. Alternatively, we may include the contours of each brick in the set of edges and consider the irregular surface of each brick as a texture. The discrimination of edges versus textures depends on the scale of analysis.

Figure 44 displays an image of bricks form the Brodatz image database.

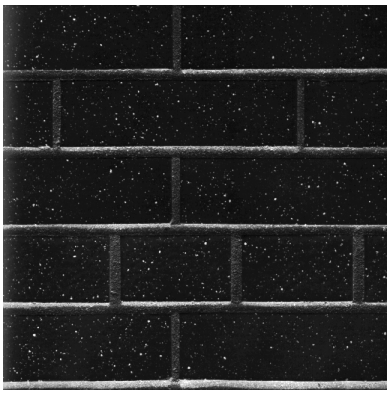


Figure 44: Bricks

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a two dimensional signal, such as an image. A Canny edge detection algorithm detects points of sharp variation in f by calculating the norm of the gradient vector of f. Recall the gradient of f is:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x) \right)$$

Let $\vec{n} = (\cos\beta, \sin\beta)$ be a unit vector in $\mathbb{R}^2$. The directional derivative of f in the direction $\vec{n}$ is defined as:

$$\frac{\partial f}{\partial \vec{n}}(x) = \nabla f(x) \cdot \vec{n} = \frac{\partial f}{\partial x_1}(x) \cos\beta + \frac{\partial f}{\partial x_2}(x) \sin\beta$$

The absolute value of $\partial f / \partial \vec{n}$ is maximum if $\vec{n}$ is colinear to ∇f. Since $\partial f / \partial \vec{n}(x)$ will be maximum in the direction of most change around x, this shows that the gradient $\nabla f(x)$ points in the direction of the most rapid increase around the point $x \in \mathbb{R}^2$; see Figure 45 for an illustration. The magnitude of the variation at x (or edge strength) is given by the norm of the gradient:

$$|\nabla f(x)| = \sqrt{\left(\frac{\partial f}{\partial x_1}(x) \right)^2 + \left(\frac{\partial f}{\partial x_2}(x) \right)^2}$$

Additionally the direction of the gradient is given by:

$$Af(x) = \begin{cases} \tan^{-1}\left(\frac{\partial f/\partial x_2(x)}{\partial f/\partial x_1(x)} \right) & \frac{\partial f}{\partial x_1}(x) \geq 0 \\ \pi + \tan^{-1}\left(\frac{\partial f/\partial x_2(x)}{\partial f/\partial x_1(x)} \right) & \frac{\partial f}{\partial x_1}(x) < 0 \end{cases} \tag{83}$$

A point $y \in \mathbb{R}^2$ is defined as an edge if $|\nabla f(x)|$ is locally maximum at $x = y$. Figure 46 shows the performance of the Canny edge detector (a refined version of it) on a color image.

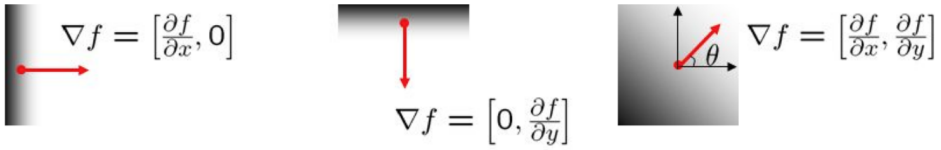


Figure 45: Illustration of the gradient vector for simple images.

A multiscale version of the Canny edge detector is implemented by smoothing the image with a convolution kernel $\theta(x)$ that is dilated at different scales. This is computed with two wavelets that are partial derivatives of θ:

$$\psi_1 = -\frac{\partial \theta}{\partial x_1}, \qquad \psi_2 = -\frac{\partial \theta}{\partial x_2}$$

These wavelets are dilated at the scales 2^j for $j \in \mathbb{Z}$:

$$\psi_{j,k}(x) = 2^{-2j}\psi_k(2^{-j}x)$$

and the dyadic wavelet transform computes:

$$Wf = \{f * \bar{\psi}_{j,k}(u) : j \in \mathbb{Z},\ k = 1, 2,\ u \in \mathbb{R}^2\}, \quad \bar{\psi}_{j,k}(x) = \psi_{j,k}(-x)$$

We write $Wf(u, j)$ as

$$Wf(u, j) = \begin{pmatrix} f * \bar{\psi}_{j,1}(u) \\ f * \bar{\psi}_{j,2}(u) \end{pmatrix}$$

and we think of it as two-dimensional vector valued function, $Wf(u, j) \in \mathbb{R}^2$. The wavelets $\psi_{j,1}$ measure variations in the horizontal direction at the scale 2^j, while the wavelets $\psi_{j,2}$ measure variations in the vertical direction at the scale 2^j.

Denote

$$\theta_j(x) = 2^{-2j}\theta(2^{-j}x), \qquad \bar{\theta}_j(x) = \theta_j(-x)$$

Figure 46: The result of the Canny edge detector applied to the image on the left. Taken from https://en.wikipedia.org/wiki/Canny_edge_detector

The dyadic wavelets can be written as:

$$\bar{\psi}_{j,k} = 2^j \frac{\partial \bar{\theta}_j}{\partial x_k}$$

It follows that

$$f * \bar{\psi}_{j,k}(u) = 2^j \frac{\partial}{\partial u_k}(f * \bar{\theta}_j)(u)$$

and so

$$Wf(u,j) = \begin{pmatrix} f * \bar{\psi}_{j,1}(u) \\ f * \bar{\psi}_{j,2}(u) \end{pmatrix} = 2^j \nabla (f * \bar{\theta}_j)(u)$$

Therefore $Wf(u,j)$ is proportional (up to a factor 2^j) to the gradient of the smooth version of f at the scale 2^j. The norm of $Wf(u,j)$, defined as:

$$|Wf(u,j)| = \sqrt{|f * \bar{\psi}_{j,1}(u)|^2 + |f * \bar{\psi}_{j,2}(u)|^2}$$

is thus proportional to $|\nabla(f * \bar{\theta}_j)(u)|$ as well. We can also compute the angle $AWf(u,j)$ of $Wf(u,j)$ using the definition (83).

The unit vector

$$\vec{n}_j(u) = (\cos AWf(u,j), \sin AWf(u,j))$$

is colinear to $\nabla(f * \bar{\theta}_j)(u)$. An edge point at the scale 2^j is a point $v \in \mathbb{R}^2$ such that $|Wf(u,j)|$ is locally maximum at at $u = v$. These points are two dimensional wavelet modulus maxima. Individual wavelet modulus maxima are chained together to form a maxima curve that follows an edge.

Figure 47 computes the Canny dyadic wavelet transform of image of a disc. The wavelet modulus maxima curves are along the boundary of the disc.

As in 1D the decay of the 2D Canny dyadic wavelet coefficients depends upon the local regularity of f. Let $0 \leq \alpha \leq 1$ denote the Lipschitz regularity. A function $f : \mathbb{R}^2 \to \mathbb{R}$ is Lipschitz α at $v \in \mathbb{R}^2$ if there exists $K > 0$ such that for all $x \in \mathbb{R}^2$,

$$|f(x) - f(v)| \leq K|x - v|^\alpha$$

where $|x|$ is the norm of $x \in \mathbb{R}^2$. As in one dimension, the local Lipschitz regularity of f is related to the asymptotic decay of $Wf(u,j)$. Indeed, this regularity is controlled by $|Wf(u,j)|$. Let $\Omega \subset \mathbb{R}^2$ be a bounded domain. One can prove that f is uniformly Lipschitz α inside Ω if and only if there exists $A > 0$ such that

$$|Wf(u,j)| \leq A 2^{j\alpha}, \quad \forall\, u \in \Omega$$

Also analogously to the 1D setting, we can synthesize high fidelity approximations to images using only their wavelet modulus maxima. Similarly to before, let Λ denote the set

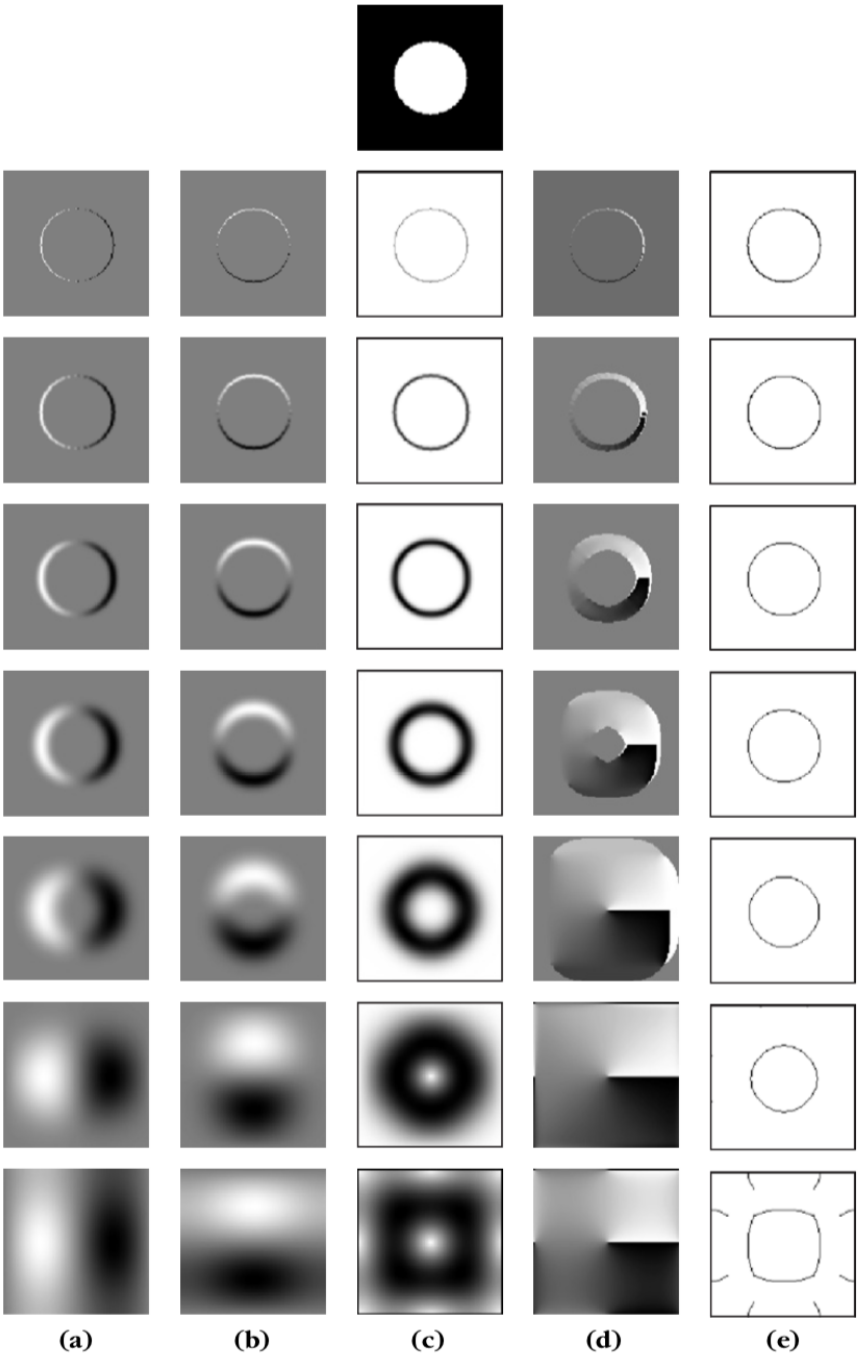


Figure 47: Top image: The disc. (a) Horizontal wavelet transform: $f * \bar{\psi}_{j,1}(u)$ for $-6 \leq j \leq 0$. (b) Vertical wavelet transform: $f * \bar{\psi}_{j,2}(u)$ for $-6 \leq j \leq 0$. (c) Norm of wavelet coefficients $|Wf(u,j)|$. (d) Angles $AWf(u,j)$ where $|Wf(u,j)| \neq 0$. (e) The wavelet modulus maxima curves.

of wavelet modulus maxima of f using the Canny dyadic wavelet transform. The synthesized image is computed using the dual synthesis:

$$f_\Lambda = \sum_{(u,j)\in\Lambda} \sum_{k=1,2} \langle f, \psi_{u,j,k}\rangle \widetilde{\psi}_{u,j,k}$$

One can also de-noise images by thresholding wavelet modulus maxima. Suppose that we have only a noisy version of f, modeled as

$$\widetilde{f}(x) = f(x) + \varepsilon(x), \quad \varepsilon(x) \sim \mathcal{N}(0, \sigma^2)$$

where $\varepsilon(x)$ is sampled independently and identically from the normal distribution; ε is referred to as white noise. The wavelet coefficients of $\widetilde{f}$ are:

$$W\widetilde{f} = Wf + W\varepsilon$$

At the large scales the averaging by $\bar{\theta}_j$ kills much of $W\varepsilon$ since ε has zero average. However, at small scales, the wavelets respond against ε and $W\varepsilon$ masks Wf. The image can be de-noised by thresholding the wavelet coefficients $W\widetilde{f}$, which results in a "cartoon" version of the original image if the variance σ^2 of the noise is not too large; Figure 48 illustrates the idea.

Exercise 76. Read Section 6.3 of *A Wavelet Tour of Signal Processing.*

Exercise 77. OPTIONAL Using your code from previous exercises compute the dyadic wavelet transform of the signal in Figure 13. Compute the wavelet modulus maxima as well. Implement a dual synthesis projection (however you like) and compute f_Λ, i.e., the signal synthesized from only the wavelet modulus maxima coefficients. Threshold the wavelet modulus maxima coefficients and synthesize a signal only from the largest ones. Turn in plots of the wavelet coefficients, the wavelet modulus maxima, and the synthesized signals. Explain your results.

Exercise 78. OPTIONAL Let $K \in \mathbb{N}$ with $K \geq 2$ and define $\vec{n}_k = (\cos(2\pi k/K), \sin(2\pi k/K)) \in \mathbb{R}^2$.

(a) Prove that $\mathcal{D} = \{\vec{n}_k : 0 \leq k < K\}$ is a tight frame of K vectors in $\mathbb{R}^2$ and that for any $\omega \in \mathbb{R}^2$, it satisfies

$$\sum_{k=0}^{K-1} |\omega \cdot \vec{n}_k|^2 = \frac{K|\omega|^2}{2}$$

(b) Let

$$\psi_k = \frac{\partial\theta}{\partial\vec{n}_k}$$

be the directional derivative of $\theta(x)$ in the direction $\vec{n}_k$. Define dilations of ψ_k as:

$$\psi_{j,k}(x) = 2^{-2j}\psi_k(2^{-j}x), \quad j \in \mathbb{Z}$$

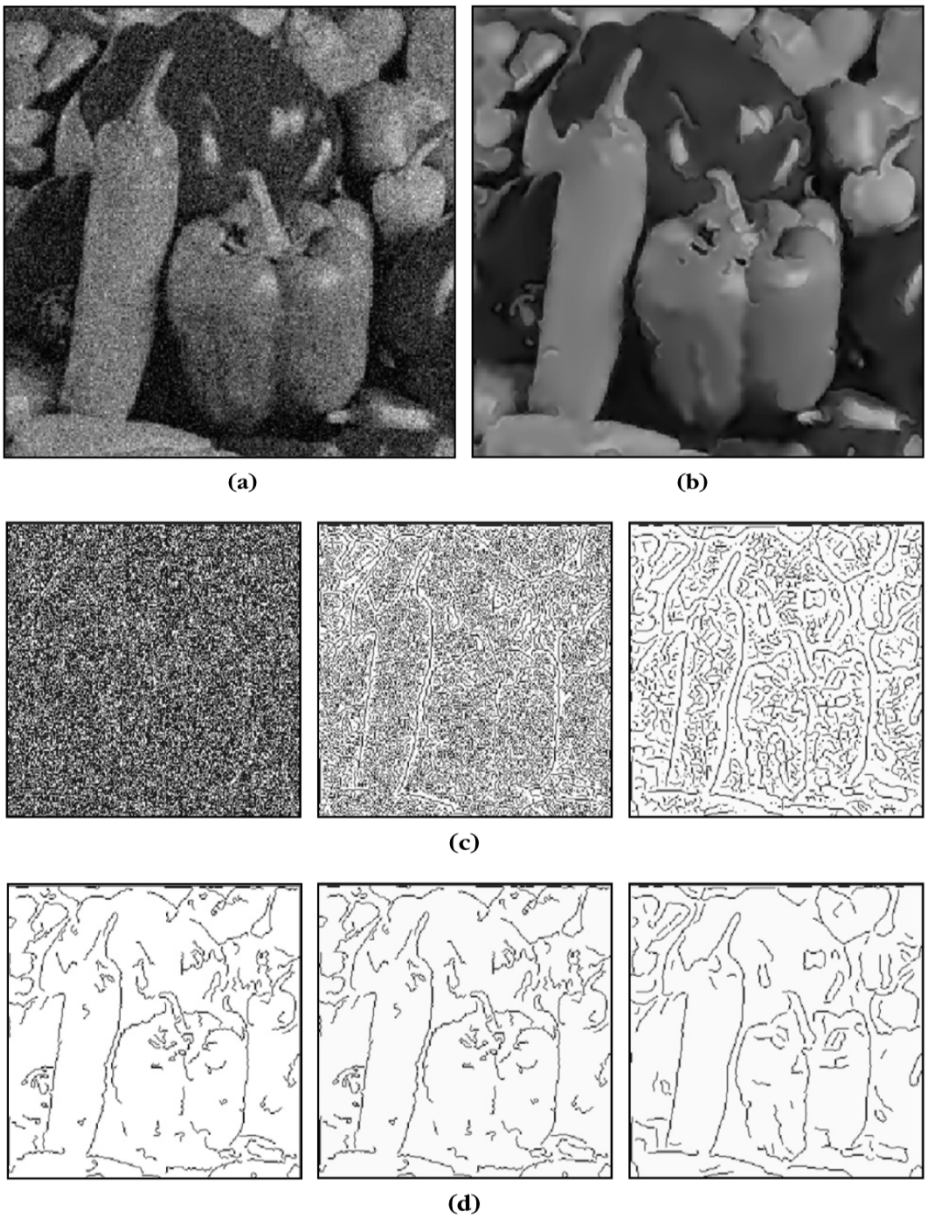


Figure 48: (a) Noisy peppers image. (b) Restored peppers image from the thresholding the maxima curves shown in (d). (c) The wavelet modulus maxima points of the noisy image for scales $-7 \leq j \leq -5$. (d) The thresholded wavelet modulus maxima.

If $\theta(x)$ is rotationally invariant (i.e., $\theta(x)$ depends only on $|x|$), then prove that $\mathcal{D} = \{\psi_{j,k} : j \in \mathbb{Z}, 0 \leq k < K\}$ are the generators of translation invariant semi-discrete frame if and only if

$$\frac{2A}{K} \leq \sum_{j \in \mathbb{Z}} 2^{2j} |\omega|^2 |\widehat{\theta}(2^j \omega)|^2 \leq \frac{2B}{K}, \quad \text{a.e. } \omega \in \mathbb{R}^2$$

Exercise 79. **OPTIONAL** In this problem you will compute the wavelet transform of two dimensional textures.

(a) Implement a dyadic 2D Morlet wavelet transform. Visualize your wavelets and their Fourier transforms, as in the plots posted in the #in-class channel of the course Slack, and turn these visualizations in.

(b) Take an image of your choice, and compute the dyadic Morlet wavelet transform of the image. Turn in plots of the real and imaginary parts of the wavelet coefficients for each (j, θ), and the modulus of the wavelet coefficients for each (j, θ). Do you see the directional responses at different scales?

References

[1] Stéphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way.* Academic Press, 3rd edition, 2008.

[2] Elias M. Stein and Rami Shakarchi. *Fourier Analysis: An Introduction.* Princeton Lectures in Analysis. Princeton University Press, 2003.

[3] John J. Benedetto and Matthew Dellatorre. Uncertainty principles and weighted norm inequalities. *Contemporary Mathematics*, 693:55–78, 2017.

[4] Yves Meyer. *Wavelets and Operators*, volume 1. Cambridge University Press, 1993.

[5] Karlheinz Gröchenig. *Foundations of Time Frequency Analysis.* Springer Birkhäuser, 2001.

[6] Steven Shreve. *Stochastic Calculus for Finance II.* Springer-Verlag New York, 2004.

[7] Hermine Biermé. Introduction to random fields and scale invariance. hal-01493834, 2018.

[8] Georgiy Shevchenko. Fractional Brownian motion in a nutshell. arXiv:1406.1956, 2014.