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:

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

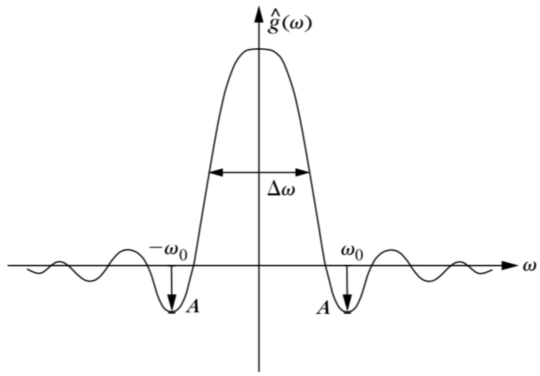

Figure 10: The energy spread of $\widehat{g}(\omega)$ is measured by its bandwidth and the maximum amplitude A of the first side lobes, located at $\pm\omega_0$.

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.

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.