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

Lecture 04: Uncertainty Principle and Introduction to DSP

January 21, 2019

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

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

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]. $\square$

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. $\square$

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:

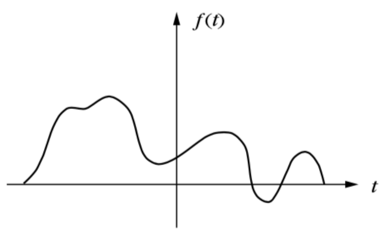
$$\text{Samples } = \{f(ns)\}_{n\in\mathbb{Z}}, \quad s^{-1} = \text{sampling rate} \tag{10}$$

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.

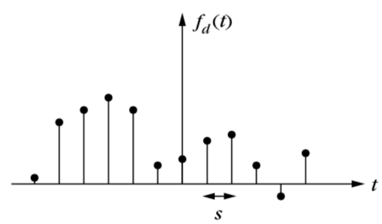
Assume that f is continuous, so that (10) is well defined. We want to know when we can recover $f(t)$ for all $t \in \mathbb{R}$ from the samples $\{f(ns)\}_{n\in\mathbb{Z}}$. We represent these discrete samples as a sum of weighted Diracs:

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

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:

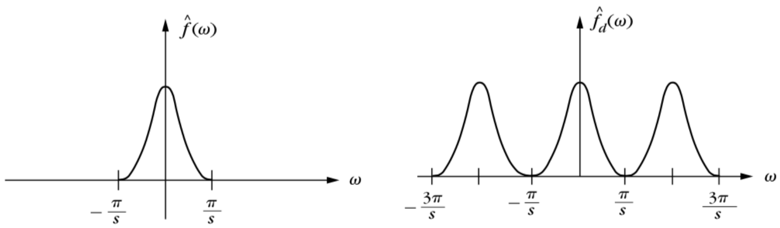
$$\widehat{f}_d(\omega) = \frac{1}{2\pi}\widehat{f} * \widehat{c}(\omega)$$

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.*

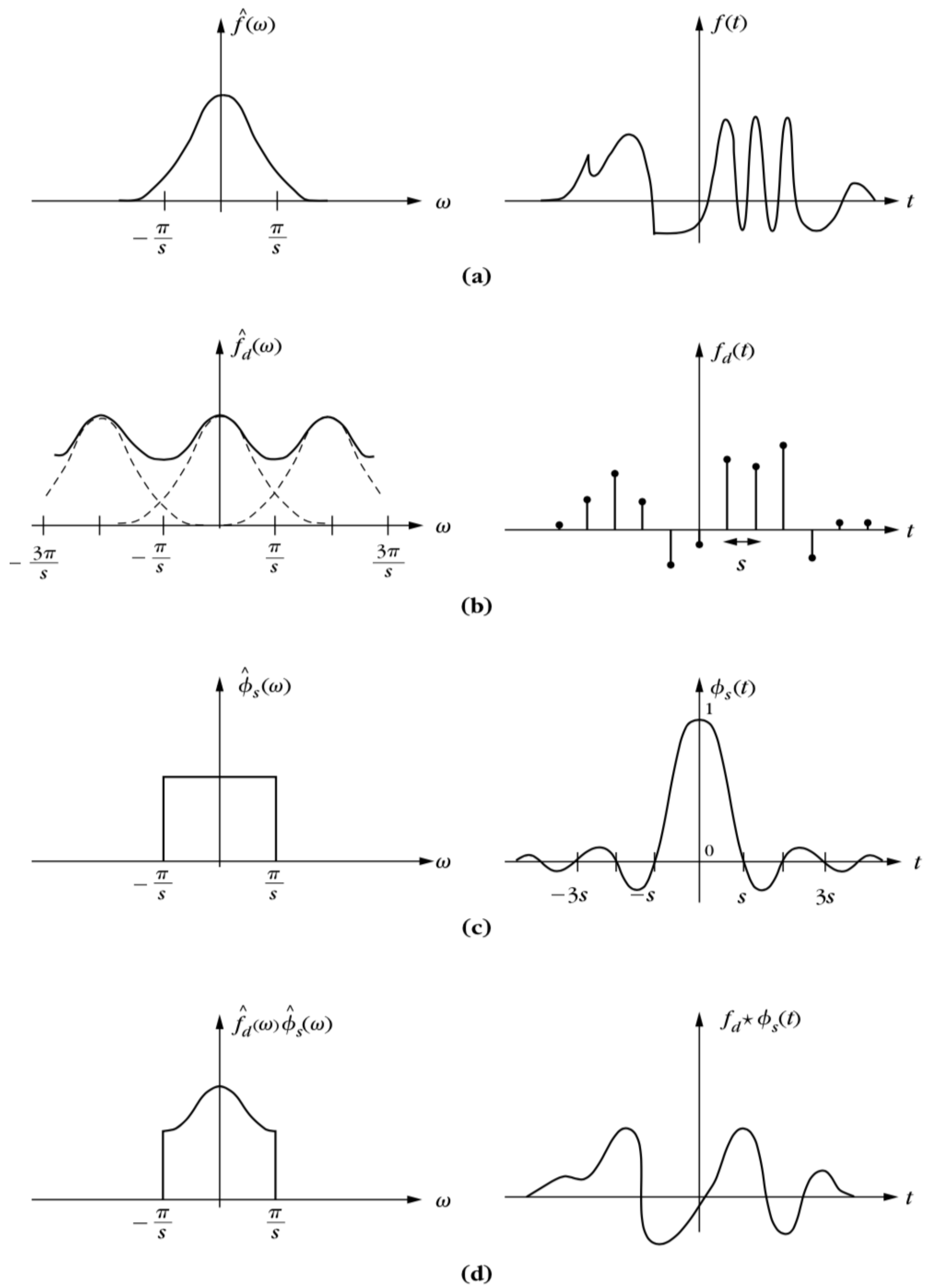


Figure 4: (a) Signal f and its Fourier transform $\widehat{f}$. (b) Aliasing produced by an overlapping of $\widehat{f}(\omega - 2k\pi/s)$ for different k, shown with dashed lines. (c) Low pass filter ϕ_s and its Fourier transform. (d) The filtering $f * \phi_s(t)$ which creates a low frequency signal 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*.

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.