Wavelets and Applications

Mattie Ji

Spring 2023

These are lecture notes from **APMA 1940Y: Wavelets and Applications** with Professor Mark Ainsworth at Brown University for the Spring 2023 semester. The most up-to-date version of the notes are maintained under my GitHub repository.

These notes are taken by Mattie Ji with gracious help and input from the instructor of this course. If you find any mistakes in these notes, please feel free to direct them via email to me or send a pull request on GitHub.

The notes are last updated March 20, 2023.

Table of Content

1   Lecture 1 - 1/25/2023

There's a limit of **18** people in this class, but there are only **14** seats...

1.1   Syllabus

- Instructor: Mark Ainsworth

- Best way to reach the instructor is by email

- Lectures are MWF from 10:00 am to 10:50 am

- Office hours are by appointment

The existential question every math major faces:

- Most of your math classes won't be useful in the real world - Professor Ainsworth used to study fluid mechanics in college, and it is VERY different from how they do it in industry... so why should you be studying mathematics?

- Well, mathematics studies things abstractly and allows us to solve problems faster. Even though engineers and physicists have a headstart, mathematicians can catch up really fast.

- But the interviewer won't stop you right there - they will ask you how you math degree helped you to accomplish this!

- This is the point of this class supposedly

This class is a WRIT class! So what is the WRIT component for - in the class, you will identify some applications of wavelets that interest you, and you will work a project on it and give a presentation on the selected topic in class. Professor Ainsworth will cover the theory part, and you will do the application part...

Certain applications of wavelets:

- The theory of Wavelets were all started in the 1990s.

- This is very important in the internet age! For instace, there's something called the **wavelet decomposition** of images that displays an image in finer and finer renditions. This is quite useful for data transmissions.

- You can use wavelets in seismic analysis

- A former student analyzed how to identify forgeries and watermarks in documents using wavelets

- Another student made a program that uses wavelets for music analysis

- Another student used wavelets for the stock market ... it didn't go too well

Wavelets are also really useful for theory!

- In Fourier Analysis (we will talk about this quite a bit), there's a certain defect with the subject we will exhibit that wavelets do not have.

Problem sets are due 7 days after the lecture it is assigned. You have to turn it before the START of the lecture, either in-person or electronically. You are supposed to work on your problem set yourself.

Lecture notes will not be handed out, unless you have a good reason to be absent. It's important to be able to take your own notes!

1.2   Time-Frequency Analysis

For the first 5 or 6 classes, we won't be knowing what a wavelet is! But we will try to build towards the subject. In order to do this, we will first show some other ways for signal analysis.

Definition 1.1. A signal is the set $\{f(t) : t \in \mathbb{R}\}$:

- Typically we will assume f is continuous, but they need not be.

- f could be either real-valued or complex-valued.

There are a few data associated to a signal:

- The **energy of signal** is

$$||f|| := (\int_{-\infty}^{\infty} |f(t)|^2 dt)^{1/2}$$

  This is pronounced "norm f". Note this varies linearly with $|f(t)|$

- The **center of signal** is

$$t^* := \frac{\int_{-\infty}^{\infty} t \cdot |f(t)|^2 dt}{\int_{\infty}^{\infty} |f(t)|^2 dt}$$

  This roughly measures where f is most of the time. The denominator makes sure that the center is invariant up to scaling. Note that a symmetric signal has the signal at the line of symmetry.

- The **radius of signal** is

$$\Delta^2 := \frac{\int_{-\infty}^{\infty} (t - t^*)^2 |f(t)|^2 dt}{\int_{\infty}^{\infty} |f(t)|^2 dt}$$

  This is roughly a measure of how spread out a signal is, so we want the spread to be translation invariant! The denominator is again for scale invariance. The actual radius is Δ, not Δ^2.

Remark 1.2. This has a lot of connections with probability and statistics! For example for any signal f, define

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

This is a probablity distribution as

$$\int_{-\infty}^{\infty} p(t) dt = 1$$

The mean is exactly the center

$$\mu = \int_{-\infty}^{\infty} tp(t) dt = t^*$$

The variance is exactly the radius

$$\sigma^2 = \int_{-\infty}^{\infty} (t - \mu)^2 p(t) dt = \Delta^2$$

2   Lecture 2 - 1/27/2023

2.1   Time Frequency Analysis - Continued

Let's test our measures on an example:

Example 2.1. Define the signal $f(t)$ where

$$f(t) = \begin{cases} \frac{1}{2\epsilon}, & -\epsilon \leq t \leq \epsilon \\ 0, & \text{otherwise} \end{cases}$$

where $\epsilon > 0$. We note that as $\epsilon \to 0$, f is more localized. Hence,

$$||f||^2 = \int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-\epsilon}^{\epsilon} (\frac{1}{2\epsilon})^2 dt = \frac{1}{2\epsilon} < \infty$$

$$t^* = \int_{-\epsilon}^{\epsilon} t(\frac{1}{2\epsilon})^2 dt = 0$$

$$\Delta^2 = \frac{1}{||f||^2} \int_{-\infty}^{\infty} t^2 |f(t)|^2 dt = \int_{-\epsilon}^{\epsilon} t^2 (\frac{1}{2\epsilon}) dt = \frac{\epsilon^2}{3}$$

$$\Delta = \frac{\epsilon}{\sqrt{3}}$$

We note that as $\epsilon \to 0$, $||f||^2 \to \infty$, $t^* = 0$, and $\Delta \to 0$ is more localized. As $\epsilon \to \infty$, $||f||^2 \to 0$, $t^* = 0$, and $\Delta \to \infty$ is less localized.

Remark 2.2. The energy, center, radius, are really just integrals of $|f(t)|^2$ multiplied with t^0, t^1, and t^2 respectively (roughly). This is called the 0-th, 1-st, and 2-nd moments of the function. Suppose we try to generalize this to higher moments for all $n \in \mathbb{N}$, can we reconstruct the signal out of the moments?

It turns out that the set of polynomial functions are dense in the continuous functions from $[a, b]$ to $\mathbb{R}$ (Stone-Weierstrass Theorem), but you can't uniformly approximate, for example, $e^x : \mathbb{R} \to \mathbb{R}$ using polynomials.

2.2   Fourier Transform

Given a signal f with finite energy, what can we do with f that tells us more information about it? Calculating all the moments is just unfeasible.

Let $\omega \in \mathbb{R}$, consider the term

$$e^{-i\omega t} f(t)$$

This term extracts the frequencies of $f(t)$, consider the function in terms of ω, $\hat{f}(\omega)$

$$\omega \mapsto \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt$$

Definition 2.3. The function $\omega \mapsto \hat{f}(\omega)$ is the **Fourier Transform** of f. This is a continuous version of Fourier Series we have dealt with before (plugging in integers to this function extracts our Fourier Coefficients).

Remark 2.4. There might be some issue with the integral not existing. Since this is an Applied Math class, all integrals exist...

Suppose we have what $\hat{f}(\omega)$ is for every ω, can we recover what f is?

Yes! This is the statement of the Fourier Inversion Theorem:

Theorem 2.5 (Fourier Inversion Theorem). Let $\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t)dt$, then

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{+i\omega t} \hat{f}(\omega)d\omega$$

The transformation back is called the Fourier Inversion formula.

Heuristically, $\hat{f}(\omega)$ represents the frequencies of the signals, and $f(t)$ represents the time series of the signal.

Proposition 2.6 (Parseval Identity). Let $f(t), g(t) : \mathbb{R} \to \mathbb{C}$, then

$$\int_{-\infty}^{\infty} f(t)\overline{g(t)} \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega)\overline{\hat{g}(\omega)} \, d\omega$$

Proposition 2.7. $\hat{f'}(\omega) = i\omega \hat{f}(\omega)$ (For less confusion, this is Fourier transform of f')

Proof. Let $f(t)$, $t \in \mathbb{R}$, consider $\hat{f'(t)}(\omega)$. Now recall that

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{+i\omega t} \hat{f}(\omega)d\omega$$

Taking the derivative on both sides, we have that

$$f'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d}{dt} e^{+i\omega t} \hat{f}(\omega)d\omega$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} i\omega e^{+i\omega t} \hat{f}(\omega)d\omega$$

Applying the Fourier Inversion Formula to $\hat{f'}(\omega)$, we can also write

$$f'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \hat{f'}(\omega)d\omega$$

But by the Fourier Inversion Formula, we have that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \underline{\hat{f'}(\omega)}d\omega = f'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{+i\omega t} \underline{i\omega \hat{f}(\omega)}d\omega$$

Hence we have that

$$\hat{f'}(\omega) = i\omega \hat{f}(\omega)$$

∎

Example 2.8. Suppose $f'(t) = g(t)$ where $t \in \mathbb{R}$ is in the time domain, hence we have that

$$\frac{df}{dt} = g(t)$$

What if we try to examine the frequency domain instead? Then consider taking the Fourier Transform on both sides:

$$\frac{\hat{df}}{dt}(\omega) = \hat{g}(\omega)$$

What is the Fourier Transform of a derivative? Well, we end up having that

$$i\omega \hat{f} = \hat{g}$$

Or equivalently we have

$$\hat{f}(\omega) = \frac{1}{i\omega}\hat{g}(\omega)$$

Computing the Fourier Inversion on both sides gives back our solution.

Remark 2.9. Lars Hörmander wrote a 5-volume book on PDES and dedicated Fourier Transformation to most of it.

3   Lecture 3 - 1/30/2023

Right now we have been going fairly slowly... Recall, so far, we have an underlying signal given either as

- (Time Domain Representation) - $\{f(t) \ : \ t \in \mathbb{R}\}$

- (Frequency Domain Representation) - $\{\hat{f}(\omega) \ : \ \omega \in \mathbb{R}\}$

These are essentially two sides of the same "coin" (the coin being the signal here).

In particular, we can convert time to frequency and backwards using the Fourier Transform:

$$\hat{f}(\omega) := \int_{-\infty}^{\infty} e^{-iwt} f(t) dt$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{+iwt} \hat{f}(\omega) d\omega$$

We also have this **Parseval Identity** that gives

$$\int_{-\infty}^{\infty} f(t)\overline{g(t)} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega)\overline{\hat{g}(\omega)} d\omega$$

Corollary 3.1. Choose $g = f$, then we have that

$$||f||^2 = \frac{1}{2\pi}||\hat{f}||^2$$

This gives a relationship between the energy of the signal and the L^2 norm of the frequency domain.

Example 3.2 (Frequency of the Step Function). Let $\epsilon > 0$ and define $f(t) = \begin{cases} \frac{1}{2\epsilon}, \ t \in [-\epsilon, \epsilon] \\ 0, \ \text{otherwise} \end{cases}$. What is $\hat{f}(\omega)$ in this case?

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

$$= \int_{-\epsilon}^{\epsilon} \frac{1}{2\epsilon} e^{-i\omega t} dt$$

$$= \frac{-1}{2i\omega\epsilon}[e^{i\omega\epsilon} - e^{-i\omega\epsilon}]$$

$$= \frac{1}{\omega\epsilon} \frac{e^{i\omega\epsilon} - e^{-i\omega\epsilon}}{2i}$$

$$= \frac{\sin(\omega\epsilon)}{\omega\epsilon}, \ \omega \in \mathbb{R} - \{0\}$$

Note when $\omega = 0$, this approaches 1. Hence we have

$$\hat{f}(\omega) = \begin{cases} \frac{\sin(\omega\epsilon)}{\omega\epsilon}, \ \omega \in \mathbb{R} - \{0\} \\ 1, \ \omega = 0 \end{cases}$$

We note that while $f(t)$ is compactly supported in the time domain, its frequency IS NOT compactly supported in the frequency domain but it does decary as $\|\omega\epsilon\| \to \infty$.

Having a frequency of 0 means the signal's not changing at that ω. In the language of Electrical Engineering, this means there's a DC component at these roots.

$\hat{f}(\omega)$ is what's called a **sink function**! Someone wrote an entire book about how to use sink functions to approximate solutions to PDEs. The time function can be thought of as the linearly combination of all the frequencies (super-position).

We also make the following notes:

1. As $\epsilon \to 0$, $f(t)$ becomes more localized. In other words, its spread Δ is getting smaller.

2. As $\epsilon \to 0$, $\hat{f}(\omega)$ becomes less localized, its correspondent spread $\hat{\Delta}$ is getting larger.

3. In other words, the more you know about the time localizations, the less you know about the frequency localization, and vice versa.

4. As $\epsilon \to \infty$, $f(t)$ is less localized.

5. As $\epsilon \to \infty$, $\hat{f}(\omega)$ decays faster and faster, and becomes more localizaed.

This observation is in fact true in general, which leads us to our next section:

3.1   The Uncertainty Principle

Let $\{f(t) : t \in \mathbb{R}\}$ be real-valued and assume $|f(t)|$ and $|\hat{f}(\omega)|$ decay sufficiently fast that:

- $||f|| < \infty$ (this implies $||\hat{f}||$ is finite)

- t^* (center of time domain) and ω^* (center of frequency domain) are well-defined.

- Δ and $\hat{\Delta}$ are well-defined.

You are not going to win a Field's Medal for proving these, but they do provide connections between the topics:

Proposition 3.3. If f is real-valued, then ω^* is ALWAYS 0.

Proof. Indeed we note that

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt$$

Taking the complex conjugate and since $f(t)$ is real valued

$$\overline{\hat{f}(\omega)} = \int_{-\infty}^{\infty} e^{+\omega t} f(t) dt = \hat{f}(-\omega)$$

Hence, we have that

$$|\hat{f}(\omega)|^2 = \hat{f}(\omega)\overline{\hat{f}(\omega)} = \hat{f}(\omega)\hat{f}(-\omega)$$

Hence the mapping $\omega \mapsto |\hat{f}(\omega)|^2$ is a even function. We know the average of an even function is just 0, hence

$$\omega^* \alpha \int_{-\infty}^{\infty} \omega |\hat{f}(\omega)|^2 d\omega = 0$$

∎

But this result tells us that knowing $\omega^* = 0$ gives us very little information about what $f(t)$ is.

Question 3.4. What about $\hat{\Delta}$? Does this tell us anything?

Consider the following translation function of f

$$f_\alpha(t) := f(t - \alpha), t \in \mathbb{R}$$

Proposition 3.5. Let $\hat{f}_\alpha(\omega)$ denote the Fourier Transform of $f_\alpha(t)$, then

$$\hat{f}_\alpha(\omega) = e^{-i\omega\alpha}\hat{f}(\omega)$$

Proof. What is $\hat{f}_\alpha(\omega)$? Well

$$
\begin{aligned}
\hat{f}_\alpha(\omega) &= \int_{-\infty}^{\infty} e^{-iwt}f(t-\alpha)dt \\
&= \int_{-\infty}^{\infty} e^{-i\omega(s+\alpha)}f(s)ds && \text{Let } s = t - \alpha \\
&= e^{-i\omega\alpha}\int_{-\infty}^{\infty} e^{-i\omega s}f(s)ds \\
&= e^{-i\omega\alpha}\hat{f}(\omega)
\end{aligned}
$$

■

Thus, translation in the time domain corresponds to a phase shift (ie. some rotation) in the frequency domain!

Proposition 3.6. Let $\hat{\Delta}_\alpha^2$ be the spread of $\hat{f}_\alpha$, then

$$\hat{\Delta}_\alpha^2 = \hat{\Delta}^2$$

In other words, shfiting in the time domain has no effect on the spread of the frequency domain.

Proof. To compute $\hat{\Delta}_\alpha^2$, we have

$$
\begin{aligned}
\hat{\Delta}_\alpha^2 &= \int_{-\infty}^{\infty} (\omega - \omega^*)^2 |\hat{f}_\alpha(\omega)|^2 d\omega \\
&= \int_{-\infty}^{\infty} (\omega)^2 |\hat{f}_\alpha(\omega)|^2 d\omega \\
&= \int_{-\infty}^{\infty} (\omega)^2 |1||\hat{f}(\omega)| d\omega \\
&= \int_{-\infty}^{\infty} (\omega)^2 |\hat{f}(\omega)| d\omega && e^{iat} \text{ has complex norm of } 1 \\
&= \hat{\Delta}^2
\end{aligned}
$$

■

$\hat{\Delta}_\alpha = \hat{\Delta}$ is true for all α! So again, knowing $\hat{\Delta}$ tells us very little about the time domain. Knowing the mean and variance of the frequency domain tells us almost nothing about the time domain.

Remark 3.7. It turns out that there's a **fundamental limit** on what we can know about the time domain representation and the frequency domain representation simultaneously.

There is in fact a quantative statement about this, let $f(t)$ be a function, then

$$\Delta_f \hat{\Delta}_f \geq \frac{1}{2}$$

This is known as the **Time-Frequency Uncertainty Principle**.

4   Lecture 4 - 2/1/2023

4.1   Cauchy-Schwarz For Integrals

> **Theorem 4.1** (Cauchy-Schwarz for Integrals)**.** Let f, g be real-valued functions then
>
> $$(f, g) \leq ||f||||g||$$
>
> where $(f, g) := \int fg$.

Proof. Let f, g be real-valued and let $\lambda \in \mathbb{R}$, then

$$0 \leq \int (f - \lambda g)^2$$

Or equivalently we have that

$$0 \leq ||f||^2 - 2\lambda(f, g) + \lambda^2 ||g||^2$$

This is really a quadratic equation in variable λ. Now hoose λ to be the value that minimizes the expression $||f||^2 - 2\lambda(f, g) + \lambda^2 ||g||^2$. In deed, this is achieved when

$$\lambda = \frac{(f, g)}{||g||^2}$$

So we have that

$$0 \leq ||f||^2 - \frac{2(f, g)^2}{||g||^2} + \frac{(f, g)^2}{||g||^2}$$

Equivalently we have that

$$(f, g)^2 \leq ||f||^2 \cdot ||g||^2$$

∎

4.2   Deriving the Time Frequency Uncertainty Principle

> **Theorem 4.2.** Let f be real-valued signal (note the result still holds for complex valued signal), then
>
> $$\Delta_f \widehat{\Delta}_f \geq 1/2$$

Proof. If f doesn't decay fast enough, the inequality on the left side would be infinite. Otherwise, suppose f does decay fast enough, consider $||f||^2 = \int_{-\infty}^{\infty} |f(t)|^2 dt$. Let $u = |f(t)|^2$ and $dv = 1$, then $v = (t - t^*)$, and integration by parts gives

$$||f||^2 = \int_{-\infty}^{\infty} |f(t)|^2 dt$$

$$= [(t - t^*)|f(t)|^2]_{-\infty}^{\infty} - 2 \int_{-\infty}^{\infty} (t - t^*) f(t) f'(t) dt$$

Since t^* s finite, then $(t - t^*)|f(t)|^2$ is integrable, so we have that the first term $[(t - t^*)|f(t)|^2]_{-\infty}^{\infty} = 0$. So we have that

$$||f||^2 = |-2\int_{-\infty}^{\infty}(t - t^*)f(t)f'(t)dt|$$

$$\leq 2\sqrt{\int_{-\infty}^{\infty}(t - t^*)^2|f(t)|^2dt}\sqrt{\int_{-\infty}^{\infty}|f'(t)|^2dt} \qquad \text{Cauchy Schwartz Inequality}$$

$$= 2(\Delta_f||f||)\sqrt{\int_{-\infty}^{\infty}|f'(t)|^2dt}$$

What is $\int_{-\infty}^{\infty}|f'(t)|^2dt$? Well, recall we have previously proven that $\widehat{f'}(\omega) = i\omega\widehat{f}(\omega)$, then Parserval's Identity tells us that

$$\int_{-\infty}^{\infty}|f'(t)|^2dt = \frac{1}{2\pi}||i\omega\widehat{f}(\omega)||^2$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}(\omega)^2|\widehat{f}(\omega)|^2d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}(\omega - 0)^2|\widehat{f}(\omega)|^2d\omega$$

$$= \frac{1}{2\pi}\widehat{\Delta}_f^2||\widehat{f}||^2 \qquad \text{Since } \omega^* = 0 \text{ as } f \text{ is real valued}$$

$$= \widehat{\Delta}_f^2||f||^2$$

Thus, we have that

$$||f||^2 \leq 2\Delta_f||f||\widehat{\Delta}_f||f||$$

Hence we have that

$$\Delta_f\widehat{\Delta}_f \geq 1/2$$

∎

Proposition 4.3 (Principle of Maximum Laziness)**.** Mark Ainsworth's favorite principle - avoid all work at as much as possible.

Remark 4.4. Note that the inequality is a very tight inequality! Since the only part where it becomes an inequality is when we applied to Cauchy-Schwarz Inequality. Hence

$$1/2 = \Delta_f\widehat{\Delta}_f$$

if and only if the condition for Cauchy-Schwarz Inequality to become an equality is true. THis happens when $(t - t^*)f(t)$ and $f'(t)$ differs by a constant! This probably gives you an idea of what function you can use to make this an equality.

The mathematical proof of the Uncertainity Principle isn't too bad, what's tricky is how we would interpret this in the context of applications.

In the context of Heisenberg's Uncertainty Principle, we have that

$$\Delta_f \sim \text{position of particle}$$

$$\widehat{\Delta}_f \sim \text{momentum of particle}$$

Up to some appropriate scaling, we have that the constant that bounds below the product is the Planck's constant. This really has nothing to do with physics but rather a consequence of the Cauchy-Schwarz Inequality.

In the context of time-frequency analysis, this means that the more narrow you examine in the time domain, the less you know about how the frequency at the place works, and vice versa. This is a HUGE problem in signal processing. You know everything with Fourier Analysis if you have what $f(t)$ or $\widehat{f}(\omega)$ are, but in the real world, you don't know what $f(t)$ or $\widehat{f}(\omega)$ are, you only have localized data about the waves.

4.3   Bandwidths and Filtering

Suppose we have a signal with time domain representation $\{f(t) : t \in \mathbb{R}\}$ with f being real-valued.

Therefore, the frequency representation

$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt$$

Ideally a frequency representation would capture all the frequencies. Unfortunately, due to limitations in the real world, we can't do that, our machines just can't measure low and high frequencies up to some number, so we are dealing with the following type of signals:

Definition 4.5. A signal $f(t)$ is **band limited** if

$$\widehat{f}(\omega) = 0, \ \forall \omega \text{ such that } |\omega| \geq \alpha > 0$$

The smallest of such α is the **bandwidth** of the signal.

Example 4.6. Given $\alpha > 0$, we set

$$\widehat{f_\alpha}(\omega) := \begin{cases} \widehat{f}(\omega), & |\omega| \leq \alpha \\ 0, & |\omega| > \alpha \end{cases}$$

, then $\widehat{f_\alpha}(\omega)$ is band limited with bandwidth α.

Question 4.7. What can we infer about $f(t)$ given $\widehat{f}(\omega)$?

5   Lecture 5 - 2/3/2023

5.1   Bandwidths and Filtering

Given a signal $\{f(t) \; : \; t \in \mathbb{R}\}$ or $\widehat{f}(\omega) \; : \; \omega \in \mathbb{R}$, we recall that we said the signal is **band-limited** with **bandwidth** α if

$$\widehat{f}(\omega) = 0, \; \forall |\omega| > \alpha$$

where α is the smallest such value.

Given a frequency $\widehat{f}(\omega)$ and $\alpha > 0$, we can construct the **cut-off frequencies** at α as:

$$\widehat{f_\alpha}(\omega) = \begin{cases} \widehat{f}(\omega), \; |\omega| < \alpha \\ 0, \; |\omega| > \alpha \end{cases}$$

Question 5.1. What happens in time domain? What can we infer about $f(t)$ based on $\widehat{f_\alpha}(\omega)$.

We observe that $\widehat{f_\alpha}(\omega)$ decreases the radius of the frequency space and localizes this, by the Uncertainty Principle, this should intuitive make the time domain less localized.

Remark 5.2. A negative frequency really just represents a wave in another direction. In the interpretation in its Fourier transform, a negative and positive frequency are respectively e^{-it} and e^{it}, their sum is exactly $\cos(t)$, which is what we interpret to be as the standing wave.

Assuming all signals decay fast enough, we observe that

$$
\begin{aligned}
f_a(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f_a}(\omega) e^{i\omega t} d\omega \\
&= \frac{1}{2\pi} \int_{-a}^{a} \widehat{f}(\omega) e^{i\omega t} d\omega \\
&= \frac{1}{2\pi} \int_{-a}^{a} e^{i\omega t} d\omega \int_{-\infty}^{\infty} e^{-i\omega s} f(s) ds \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(s) ds \int_{-a}^{a} e^{i\omega(t-s)} dt \qquad \text{Fubini's Theorem}
\end{aligned}
$$

We observes that

$$\frac{1}{2\pi} \int_{-a}^{a} e^{i\omega(t-s)} dt = \frac{1}{2\pi} [\frac{e^{i\omega(t-s)}}{i(t-s)}]_{-a}^{a} = \frac{1}{\pi(t-s)} \frac{e^{ia(t-s)} - e^{-a(t-s)}}{2i} = \ell_a(t-s)$$

We define

$$\ell_a(t) = \frac{\sin at}{\pi t}$$

as the **Shannon Sampling Function**.

Thus, we have that

$$f_a(t) = \int_{-\infty}^{\infty} f(s) \ell_a(t-s) ds = (f * \ell_a)(t)$$

This is indeed the convolution of $f(t)$ and $\ell_a(t)$.

- Every value of $f(s)$ affects $f_a(t)$, apart from when $\sin \alpha t = 0$ (which is a discrete set)

- $f_a(t)$ is most affected by $f(t)$

> **Question 5.3.** What happens if we vary the value of α?

For α getting larger, we should expect that $\widehat{f_a}(\omega)$ gets closer and closer to $\widehat{f}(\omega)$, so $f_a(t)$ should be closer to $f(t)$. In fact, $\ell_a(t)$ becomes more and more like the Dirac Delta function as $a \to \infty$.

For α getting smaller, we should expect $f_a(t)$ getting less and less localized.

We also have a corresponding formulation of filtering by considering the time signal $f(t)$ and

> **Definition 5.4.** We have a corresponding formulation of filtering - $\widehat{f} \to \widehat{f_\alpha}(\omega)$, where we can define
>
> $$\widehat{f_\alpha}(\omega) := \widehat{\ }(\omega)H(\omega)$$
>
> where $H(\omega)$ is the indicator function of $[-\alpha, \alpha]$. H is called the **frequency representation of the filter**.

In the Frequency Domain, we have that
$$\widehat{f} \mapsto \widehat{f_a} = \widehat{f}H$$

In the Time Domain, we have that
$$f \mapsto f_a = f * h$$

where we note that $H = \widehat{h}$!

> **Remark 5.5.** So who is Claude Shannon? His contribution to signal processing was figuring out - given a bandlimited signal, how many points do you need to sample, so that others can reconstruct the signal based on the points. This is sometimes called the Shannon Sampling Theorem, and we will discuss more about this later.

5.2   Filter Design

Recall we have $\widehat{f} \to \widehat{f_a}(\omega) = H(\omega)\widehat{f}(\omega) \sim \widehat{f}(\omega)$.

How should we choose an appropriate $H(\omega)$? Well we ideally want

- We know that $\widehat{f}(0) = \int_{-\infty}^{\in} f(t)dt$ (average value of f), hence we want f_a and f to have the same average, hence
$$\widehat{f_a}(0) = H(0)\widehat{f}(0)$$
  This is true if and only if $H(0) = 1$.

- We want to perserve moments as before, so we want to perserve higher order moments, for example $\int_{-\infty}^{\infty} tf(t)dt$. We observe that
$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} f(t)dt$$
  Thus we have that
$$\frac{d\widehat{f}}{d\omega}(\omega) = \int_{\infty}^{\infty} -ite^{i\omega t}tf(t)dt$$
  Hence we have that
$$\widehat{f}'(0) = \int_{-\infty}^{\infty} tf(t)dt$$
  So we really want to perserve $\widehat{f_a'}(0) = \widehat{f}'(0)$, which we have
$$\widehat{f_a'}(0) = H(0)\widehat{f}'(0) + H'(0)\widehat{f}(0)$$
  The equality we want is perserved if and only if $H(0) = 1$ and $H'(0) = 0$.

- By Engineering's Induction, (you can also generalize this rigorously), we have that H perserves all momemnts up to order m if and only if

$$H(0) = 1, \ H'(0) = H''(0) = ... = H^{(m)}(0) = 0$$

This type of filter is called a **BUTTERWORTH FILTER** of order m.

6   Lecture 6 - 2/6/2023

6.1   Gábor Transform

Recall in last lecture, we have previously demonstrated the following equivalence:

Given a function f in the time domain, where h is a real-valued filter with the assumption:

- h is real-valued

- $t^* = 0$

- $\Delta_h < \infty$

Then what get's filtered in the time domain is
$$g = h * f$$
In the frequency world, we achieve the same with $\widehat{g} = H\widehat{f}$, where $H = \widehat{h}$ (Note our convolution becomes multiplication).

We also have the uncertainty principle, which tells us that there's a limit on the localization we can get simultaneously in time and frequency. What does this mean for filtering?

We have
$$g(\alpha) = \int_{-\infty}^{\infty} f(t)h(\alpha - t)dt, \ \alpha \in \mathbb{R}$$

where $h(\alpha - t)$ representations the localization around α.

We know the Fourier Transform of $h(\alpha - t)$ is just

$$\int_{-\infty}^{\infty} e^{-i\omega t}h(\alpha - t)dt = \int_{-\infty}^{\infty} e^{-\omega(w-s)}h(s)ds \qquad \text{Let } s = \alpha - t$$
$$= e^{-i\omega\alpha}\overline{\int_{-\infty}^{\infty} e^{-i\omega s}h(s)ds}$$
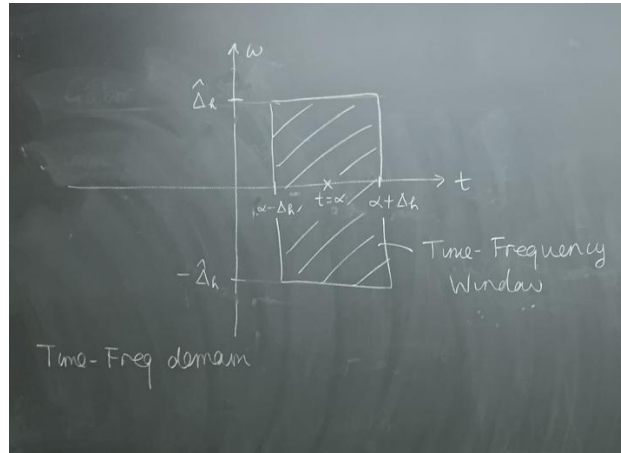$$= e^{-i\omega\alpha}\overline{H(\omega)}$$

Parserval's Identity tells us that

$$\int_{-\infty}^{\infty} f(t)h(\alpha - t)dt = g(\alpha) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(\omega)e^{i\omega\alpha}H(\omega)d\omega$$

In the time domain, we have the localization of $f(t)$ around $t = \alpha$ over a radius Δ_h (the role of $h(\alpha-t)$ is the sampler).

In the frequency domain, this is the sampling of $\widehat{f}$ over a radius $\widehat{\Delta}_h$ centered around $\omega = 0$.

What happens in the time-frequency domain:

This is called the Time-Frequency Window on $(\alpha - \Delta_h, \alpha + \Delta_h) \times (-\widehat{\Delta}_h, \widehat{\Delta}_h)$. In other words, filtering gives information about $f(t)$ and $\widehat{f}(\omega)$ over a time-frequency window.

We note that the area of this window is bounded by

$$(2\Delta_h)(2\widehat{\Delta}_h) \geq 2$$

We have two significant issues:

- The area of the time-frequency window is limited

- We have flexibility to shift the time-window by varying α, but there's no flexibility to vary the frequency as it is always centered on $\omega = 0$.

Question 6.1. Can we try to limit the area of the time-frequency window?

Example 6.2. Here are some possible attempts:

1. Try $h(t) = \begin{cases} \frac{1}{2\epsilon}, & |t| \leq \epsilon \\ 0, & \text{otherwise} \end{cases}$, then $\widehat{h}(\omega) \sim \frac{\sin(\omega t)}{\omega t}$. Then we have that

$$\begin{aligned} \widehat{\Delta}_h &= \lim_{R \to \infty} \int_{-\infty}^{\infty} |\omega||\widehat{f}(\omega)|^2 d\omega \\ &= \infty \end{aligned}$$

So the radius is not good.

2. Let's try the Gaussian filter $h(t) = e^{-t^2/2}$, then $\Delta_h \cdot \widehat{\Delta}_h = 1/2$. This suggests that a Gaussian filter should be optimal, but this function is globally supported. We do have rapid decay, and

$$h(\alpha - t) = e^{-(t-\alpha)^2/2} \approx 0 \text{ when } |t - \alpha| \text{ is "large"}$$

In practice, we would just approximate using a Gaussian filter.

Question 6.3. Can we try to move the time-frequency window vertically so that it can be centered on a given frequency ξ?

Answer. Yes! This is solved by what's known as the Gabor transform. Gabor actually won a Nobel Prize for discovering holograms. He was also a member of Fellows of the Royal Society. ∎

Now, let's consider

$$g(t) = (h * f)(t) \text{ or } \widehat{g}(\omega) = H(\omega)\widehat{f}(\omega)$$

given information around $\omega = 0$.

Idea 6.4. Let's try to shift the spectrum of f before filtering!

Recall we have that

$$f(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega s} \widehat{f}(\omega) d\omega$$

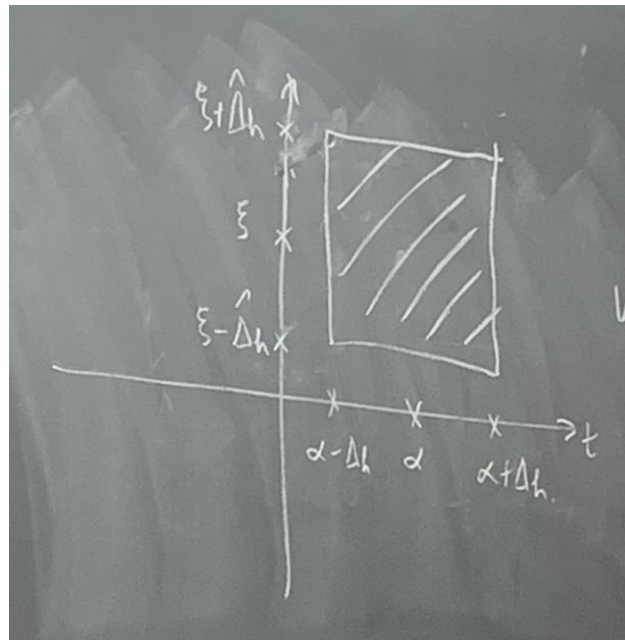This is a super-position of frequencies. Hence

$$e^{-i\xi s} f(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\omega - \xi)s} \widehat{f}(\omega) d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega s} \widehat{f}(\omega + \xi) d\omega$$

This shows that the spectrum of f is shifted by an amount ξ (sine $\widehat{f}(\omega) \to \widehat{f}(\omega + \xi)$).

Definition 6.5. Applying the filter to the resulting function gives the **Gabor transform**:

$$G_h(\alpha, \xi) := (h * e^{-i\xi s} f(s))(\alpha) = \int_{-\infty}^{\infty} e^{-i\xi s} f(s) h(\alpha - s) ds$$

In the time-frequency window, we now have

This is also a special case of what's known as a "short time Fourier transform".

7   Lecture 7 - 2/8/2023

Recall last lecture - given a special filter $\{h(t)\ :\ t \in \mathbb{R}\}$, then we define the Gabor transform as

$$(G_h f)(\alpha, \xi) = \int_{-\infty}^{\infty} ds\, f(s) e^{-i\xi s} h(\alpha - s)$$

The idea is that - by varying α and ξ, we can shift the window horizontally and vertically across the time-frequency domain, around spatial location α and frequency ξ.

> **Proposition 7.1.** The Gabor transform gives a time-frequency window
>
> $$(\alpha - \Delta_h, \alpha + \Delta_h) \times (\xi - \widehat{\Delta}_h, \xi + \widehat{\Delta}_h)$$
>
> ie. window has area
>
> $$2\Delta_h 2\widehat{\Delta}_h \geq \frac{2 \cdot 2}{2} = 2$$

Proof. We note that $h(\alpha - s)$ is centered on α with spread being Δ_h and $e^{-i\xi s}$ has no contribution to the time domain, thus clearly the temporal window is given by $(\alpha - \Delta_h, \alpha + \Delta_h)$.

By Parserval's Identity, we note that

$$(G_h f)(\alpha, \xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \widehat{f}(\omega) \overline{?}$$

where the term under the complex conjugation is exactly the Fourier transform of $e^{+i\omega s} h(\alpha - s)$, which is

$$
\begin{aligned}
\int_{-\infty}^{\infty} e^{-i\omega s} e^{i\xi s} h(\alpha - s) ds &= \int_{-\infty}^{\infty} e^{-i(-\xi)s} h(\alpha - s) ds \\
&= \int_{-\infty}^{\infty} e^{-i(\omega - \xi)\alpha} e^{i(\omega - \xi)t} h(t) dt && \text{Let } t = \alpha - s \\
&= e^{-i(\omega - \xi)\alpha} int_{-\infty}^{\infty} e^{-i(\xi - \omega)t} h(t) dt \\
&= e^{-i(\omega - \xi)\alpha} \widehat{h}(\xi - \omega)
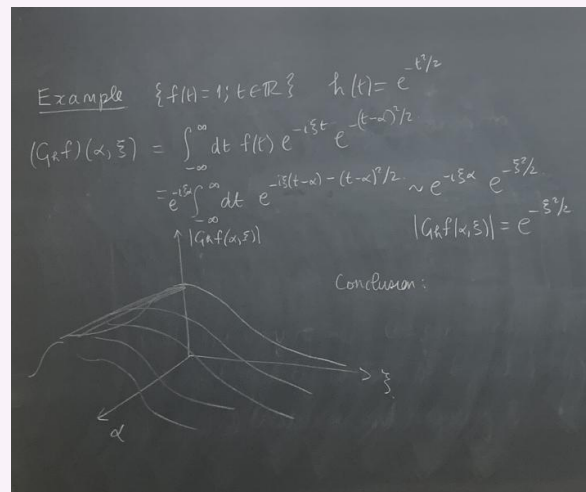\end{aligned}
$$

Thus, we have that

$$(G_h f)(\alpha, \xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \widehat{f}(\omega) e^{+i(\omega - \xi)\alpha} \overline{\widehat{h}(\xi - \omega)}$$

Hence, we have centering on $\omega = \xi$ on the frequency domain and width of window $\widehat{\Delta}_h$. ∎

> **Example 7.2.** Consider a signal $\{f(t) = 1\ :\ t \in \mathbb{R}\}$ and $h(t) = e^{-t^2/2}$, then
>
> $$
> \begin{aligned}
> (G_h f)(\alpha, \xi) &= \int_{-\infty}^{\infty} dt\, f(t) e^{-i\xi t} e^{-(t-\alpha)^2/2} \\
> &= e^{-i\xi\alpha} \int_{-\infty}^{\infty} e^{-(t-\alpha)^2/2 - i\xi t} \\
> &\sim e^{-i\xi\alpha} e^{-\xi^2/2}
> \end{aligned}
> $$

In particular, we note that $|G_h f| = e^{-\xi^2/2}$ with the graph:

Just analyzing the graph, we see that $|G_h f|$ is not varying in intensity in time, so there's no spatial localization at all. $|G_h f|$ is largest when $\xi = 0$, so the largest frequency content of the signal is around where $\xi = 0$, hence the dominant frequency is $\xi = 0$.

Question 7.3. How to compute/approximate $G_h f$ when f is not so simple?

We always make $h(t) = e^{-t^2/2}$ as our filter, to compute a numerical approximation, here are a few issues:

1. $f(t)$ may be given only partially, e.g. only at certain points or on a limited range. Suppose we have $\{f(t) : -w < t < w\}$ for some w. So the first approximation we need is

$$(G_h f)(\alpha, \xi) \approx \int_{-w}^{w} dt f(t) e^{-i\omega t} e^{-(t-\alpha)^2/2}$$

   This is a good approximation if $f(t)$ decays rapidly outside. We also expect this to be a good approximation even if f does not decay, provided that α is located "well-inside" $(-w, w)$.

2. We still can't compute analytically. Let's try to use the Trapezoidal Rule with nodes indexed by $-N, -N + 1, ..., 0, ..., N$ (choose N large enough), then

$$\approx \frac{W}{N} \sum_{m=-N}^{N} {}'' e^{-i\xi t_m} e^{-(t_m-\alpha)^2/2} f(t_m)$$

   where $t_m = mW/N$ for $m = -N, ..., N$ (the $\sum''$ means the first and the last term are halved).

3. This could be very expensive if $N >> 1$ and/or $W >> 1$, we note that

$$e^{-t^2/2} < 10^{-6} \text{ when } |t| > 5.6$$

   Therefore we are only going to sum over N such that $|t_m - \alpha| < 6$, ie. only when

$$-6N/W < m - \frac{N\alpha}{W} < 6N/W$$

   Hence combining our approximations gives

$$(G_h f)(\alpha, \xi) \approx \frac{W}{N} \sum_{m=Lo}^{m=Hi} e^{-i\xi t_m} e^{-(t_m-\alpha)^2/2} f(t_m)$$

where

$$Lo := \max(-N, \lfloor fracNW(\alpha - 6) \rfloor)$$
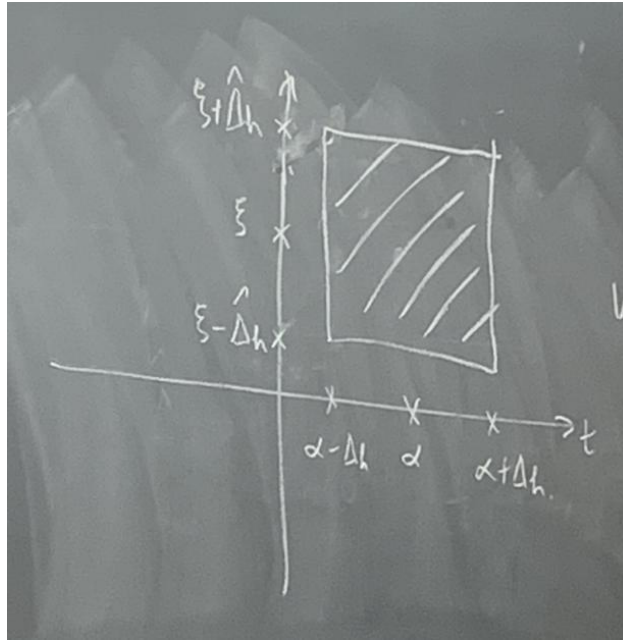$$Hi = \min(N, \lceil N/W(\alpha + 6) \rceil)$$
$$t_m = mW/N$$

8   Lecture 8 - 2/10/2023

8.1   Gabor Transform (Continued)

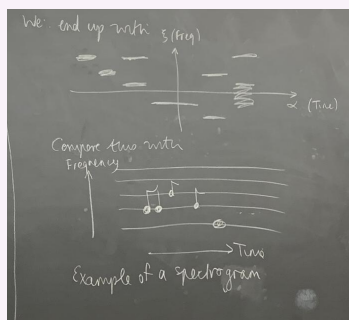We recall that the Gabor transform was defined

$$(G_h f)(\alpha, \xi) = \int_{-\infty}^{\infty} f(t) e^{-i\xi t} e^{-(t-\alpha)^2/2} dt$$

The Gabor Transform enable us to do time-frequency analysis. In particular we can plot the following **spectrogram** and center on some window (α, ξ) with width $2\Delta_h$ and height $2\widehat{\Delta}_h$:

But since we have this fundamental restriction on the area of the window, $(G_h f)(\alpha, \xi) \sim$ average of f and $\hat{f}$ over the window.

> **Example 8.1.** We can't concentrate this around points because of the fundamental limit on the area of the window, hence the spectrogram has some imperfections going on:
>
>
>
> We can compare this with a music sheet. In particular, a music sheet is really an example of a spectrogram and is in fact a perfect spectrogram, as you know exactly what frequencies should occur at each time.

Unfortunately, Gabor transform is limited by the time-frequency uncertainty principle, but we can change the aspect ratio of the time-frequency window to focus on localization in either time or frequency.

To decrease Δ_h, we want to decay our $e^{-t^2/2}$ faster, to do this we can consider $e^{-\epsilon t^2/2}$. If we choose $\epsilon > 1$, we tighten the time window. If $\epsilon < 1$, we tighten the frequency window.

A discontinuity at time t corresponds to the existence of all frequencies around time t.

Claim 8.2. Varying ϵ is more useful than varying ξ and/or α.

> **Question 8.3.** Why don't we try to sacrifice a parameter?

8.2   Varying $t - \omega$ Aspect Ratio Transform

Let $\{h(t) : t \in \mathbb{R}\}$ be a given filter and define a new transform

$$W_h(\alpha, a) = \int_{-\infty}^{\infty} f(t) h\left(\frac{\alpha - t}{a}\right) dt$$

By varging 'a', we can change the aspect ratio of the window (we are not normalizing yet)

$$\Delta_h^2 \sim \int_{-\infty}^{\infty} (t - t^*) |h(t)|^2 dt \ \ (a = 1)$$

$$\Delta_{h,a}^2 \sim \int_{-\infty}^{\infty} (t - t^*)^2 |h(t/a)|^2 dt$$

Without loss of generality, let's say $t^* = 0$, then

$$\Delta_{h,a}^2 \sim \int_{-\infty}^{\infty} t^2 |h(t/a)|^2 dt$$

$$= a^3 \int_{-\infty}^{\infty} (t/a)^2 |h(t/a)|^2 d(t/a)$$

$$= a^3 \int_{-\infty}^{\infty} s^2 |h(s)|^2 ds$$

Recall we also have to normalize this, and

$$\|h(t/a)\|^2 = a \int_{-\infty}^{\infty} |h(t/a)|^2 \frac{dt}{a} = a\|h\|^2$$

Hence, we have that

$$\Delta_{h,a}^2 \sim a^2 \Delta_h^2$$

Let's try to rescale h so that $h\left(\frac{\bullet}{a}\right)$ is independent of $|a|$, ie. let's define

$$h_a(t) = \frac{1}{\sqrt{a}} h(t/a)$$

Hence we note that

$$\|h_a(t)\|^2 = \frac{1}{a} \int_{-\infty}^{\infty} |h(t/a)|^2 dt = \int_{-\infty}^{\infty} |h(s)|^2 ds$$

This value is now independent of a!
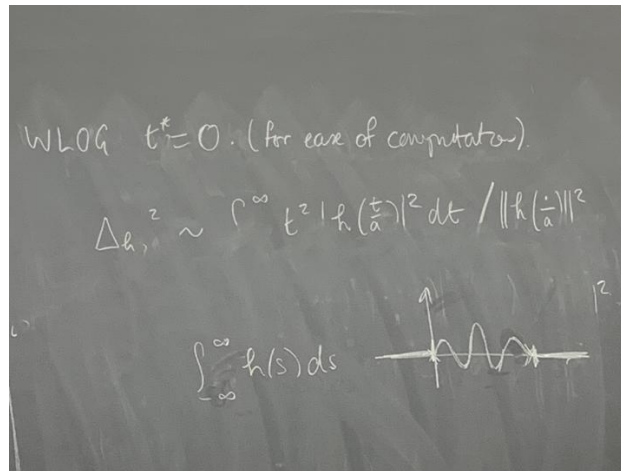
We now arrive at a new type of transform

$$(W_h f)(\alpha, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) h(\frac{\alpha - t}{a}) dt$$

This is also called the **continuous Wavelet transform**. In particular, h is the **analyzing wavelet**! We want to choose h so that it doesn't pick up any DC components, namely that

$$\int_{-\infty}^{\infty} h(s) ds = 0$$

For covenience, we also want $h(s)$ to be compactly supported. So we want something like:

Which looks like the shape of a wavelet!

9   Lecture 9 - 2/13/2023

9.1   Continuous Wavelet Transform

At the end of the last class, we claim up with the following

Definition 9.1. Let ψ be a real-valued function with $\psi(t)$, $|t|^{1/2}\psi(t)$, and $t\psi(t)$ all integrable with finite energy (ie. square integrable), and that $\widehat{\psi}(0) = 0$ (note this is true if and only if $\int_{-\infty}^{\infty} \psi(t)dt = 0$). Then the **continuous wavelet transform (CWT)** of a signal $f(t)$ is

$$(W_\psi f)(\alpha, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi(\frac{\alpha - t}{a})dt, \ \alpha \in \mathbb{R}, a > 0$$

The landpass filter ψ is called the "analyzing wavelet".

Question 9.2. What are the Time-Frequency properties of this?

In the time domain, we get a window around $t = \alpha$ and radius is given by

$$\Delta_\psi^2 = \int_{-\infty}^{\infty} t^2\psi(\frac{\alpha - t}{a})^2 dt / \int_{-\infty}^{\infty} \psi(\frac{\alpha - t}{a})^2 dt$$

$$= a^2 \int_{-\infty}^{\infty} s^2\psi(s)^2 ds / \int_{-\infty}^{\infty} \psi(s)^2 ds \qquad\qquad \text{Let } s = (\alpha - t)/a$$

$$= (a\Delta_\psi)^2$$

So the time window is $(\alpha - a\Delta_\psi, \alpha + a\Delta_\psi)$.

In the frequency domain, we use Parserval's Identity to write $W_\psi f$ in terms of $\widehat{f}(\omega)$. To do this, we need the Fourier Transform of $t \mapsto \psi(\frac{\alpha - t}{a})$. Using techniques in Homework 1, we get

$$\int_{-\infty}^{\infty} e^{-i\omega t}\psi(\frac{\alpha - t}{a})dt = ae^{-i\omega\alpha}\overline{\widehat{\psi}(a\omega)}$$

Hence, by Parserval's Identity, we have that

$$(W_\psi f)(\alpha, a) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(\omega)e^{i\omega\alpha}\widehat{\psi}(a\omega)d\omega$$

This shows that the frequency window is centered on $\omega = 0$ and has radius, with a very similar calculation as before

$$\widehat{\Delta}_{\psi,a}^2 = \frac{1}{a^2}\widehat{\Delta}_\psi^2$$

Hence we have the frequency window on $(-\frac{1}{a}\widehat{\Delta}_\psi^2, +\frac{1}{a}\widehat{\Delta}_\psi^2)$.

Our time frequency window looks like:
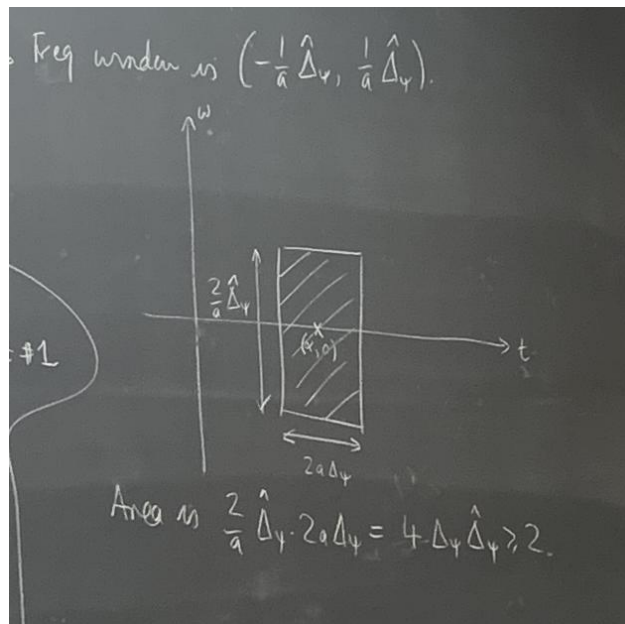
with area being $\frac{2}{a}\Delta_\psi(2a\widehat{\Delta}_\psi) = 4\Delta_\psi\widehat{\Delta}_\psi \geq 2$. We didn't change the area of the window, but we can change the aspect ratio by varying the value of a.

The flexibility to vary a allows us to change the scale at which resolve time and/or frequency. In a sense, a is similar to a "zoom factor" on a microscope that allows us to focus in a local feature.

Definition 9.3. A band-pass filter $\psi(t)$ is a filter with dc component being 0, ie $\int_{-\infty}^{\infty}\psi(t)dt = 0$

Example 9.4. The Haar Wavelet is
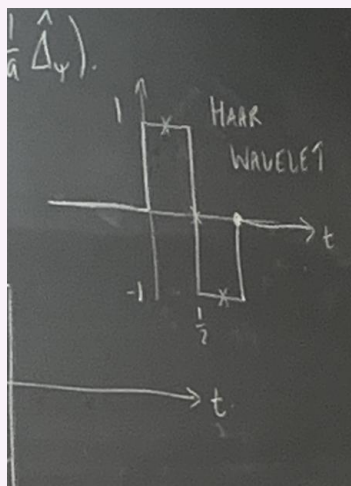
This is in some sense the simplest wavelet. It is locally supported, band-pass, etc.

9.2   Smoothness of Functions

The actual value of the filter function doesn't matter, what we care about more is how the function varies over time! This concerns with the "smoothness" of a given function f, ie. we want to know where it is not smooth.

Question 9.5. How do we measure smoothness?

Definition 9.6. We say f is Lipschitz-continuous at s with exponent μ if there exists $K < \infty$ and $\mu \in (0, 1)$

$$|f(s) - f(t)| \leq K|s - t|^{\mu}$$

, for t close to s. The idea here is that the larger μ is, the more "differentiable" the function is.

Question 9.7. Can we figure out s and μ by looking at $\widehat{f}(\omega)$ is?

Theorem 9.8. If $\widehat{f}$ satisfies $\int_{-\infty}^{\infty} |\widehat{f}(\omega)|(1 + |\omega|^{\mu})d\omega < \infty$, then f is Lipschitz continuous every where with exponent μ.

Proof. Let's consider $\frac{|f(s) - f(t)|}{|s - t|^{\mu}}$ and figure out how large μ can be, so

$$f(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f(\omega)} e^{+i\omega s} d\omega$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f(\omega)} e^{+i\omega t} d\omega$$

$$f(s) - f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(\omega)(e^{i\omega s} - e^{i\omega t}) d\omega$$

We observe that

$$|e^{i\omega s} - e^{i\omega t}| \leq 2 \qquad\qquad\qquad \text{By Triangle's Inequality, this is best bound when } \omega >> 0$$

On the other hand, we have that

$$|e^{i\omega s} - e^{i\omega t}| = |[e^{i\omega\theta}]_t^s|$$
$$= |\int_t^s i\omega e^{i\omega\theta} d\theta|$$
$$\leq |\int_t^s d\theta|\omega||$$
$$\leq |\omega| \cdot |s - t| \qquad\qquad\qquad \text{Good when } s \text{ is close to } t \text{ and } \omega \text{ is small}$$

Now, we have that

$$\frac{|e^{i\omega s} - e^{i\omega t}|}{|s - t|^{\mu}} \leq \begin{cases} 2|s - t|^{-\mu}, \\ |\omega||s - t|^{1-\mu} \end{cases}$$

If $|s - t| > |\omega|^{-1}$, then $|s - t|^{-1} < |\omega|$, hence $|s - t|^{-\mu} < |\omega|^{\mu}$.

If $|s - t| < |\omega|^{-1}$, then $|s - t|^{1-\mu} < |\omega|^{\mu-1}$, hence $|\omega||s - t|^{1-\mu} < |\omega|^{\mu}$, thus we always have that

$$\frac{|e^{i\omega s} - e^{i\omega t}|}{|s - t|^{\mu}} \leq C|\omega|^{\mu}$$

for some constant C, hence we have that

$$\begin{aligned}
\frac{|f(s) - f(t)|}{|s - t|^{\mu}} &\le \frac{1}{2\pi} \int_{-\infty}^{\infty} |\widehat{f}(\omega)| \cdot \frac{|e^{i\omega s} - e^{i\omega t}|}{|s - t|^{\mu}} d\omega \\
&\le \frac{C}{2\pi} \int_{-\infty}^{\infty} |\widehat{f}(\omega)|(1 + |\omega|^{\mu}) d\omega \\
&< \infty
\end{aligned}$$

∎

Good news - looking at the decay of $|\widehat{f}(\omega)|$ as $|\omega| \to \infty$ tells us about the smoothness of f.

Bad news - this tells us nothing about the local smoothness, ie. how doe sthe smooth vary with respect to position! (μ is the lower bound on the smoothness of all the points).

Recall that for $t \mapsto f(t - \alpha)$, the Fourier transform is $\omega \mapsto e^{i\alpha\omega}\widehat{f}(\omega)$, but the fourier transform DOES NOT translate anything, so its modulus is the same as $|\widehat{f}(\omega)|$. Hence no matter how we try to tweak around the theorem, we won't have the local information.

10   Lecture 10 - 2/15/2023

Recall that we have established, last time, that:

- Locally Lipschitz - f is locally Lipschistz at α with exponent μ if

$$|f(\alpha) - f(t)| \le K|\alpha - t|^\mu$$

  where μ indicates/measures how smooth f is at α

- Can we deduce μ without looking directly at $f(t)$ and $t \in \mathbb{R}$

- We proved last lecture that:

> **Theorem 10.1.** If $\widehat{f}$ satisfies
>
> $$\int_{-\infty}^{\infty} |\widehat{f}(\omega)|(1 + |\omega|^\mu)d\omega < \infty$$
>
> then f is globally Lipschitz with exponent μ.

- As a corollary, $f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\widehat{f}(\omega)d\omega$ implies that

$$|f(t)| \le \frac{1}{2\pi}\int_{-\infty}^{\infty} |\widehat{f}(\omega)|d\omega < \infty$$

  So the 1+"something" means f is a well-defined behavior. However, this theorem only tells us the **global** μ, but some parts of f might do better than this global μ. This means we have to analyze the behaviors of f more locally.

> **Question 10.2.** Can we find a better way to identify μ and hopefully how smooth this is at α?

Consider the Lipschitz condition,

$$|f(\alpha) - f(t)| \le K|\alpha - t|^\mu$$

Let's write $t = \alpha - as$ for some $a > 0$ and s, then we have that

$$|f(\alpha) - f(\alpha - as)| \le Ka^\mu|s|^\mu$$

Suppose we scale $a \mapsto a/2$, then the $Ka^\mu|s|^\mu$ becomes $Ka^\mu|s|^\mu/2^\mu$. This means that we can interpret Lipschitz as a scaling relation where μ measures the relative change on f when the arguments get closer (recall $0 < \mu < 1$).

This suggests that by varying "a", we should be able to detect the value of μ when we examine the change in f.

Let's recall the continuous wavelet transform:

$$(W_h f)(\alpha, a) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)h\left(\frac{\alpha - t}{a}\right)dt$$

We can interpret a as a "zoom" or "scale" factor, and $W_h f$ is the response to the scaling. We claim it is plausible that we can use the CWT to identity μ!

We will now make this assertion more precise:

$$(W_h f)(\alpha, a) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(s)h\left(\frac{\alpha - s}{a}\right)ds$$

$$= \sqrt{a}\int_{-\infty}^{\infty} f(\alpha - at)h(t)dt \qquad\qquad \text{Let } t = \frac{\alpha - s}{a}$$

When we change $a \mapsto a/2$, is there a μ-dependent response above? There's no apparent behavior that depends on μ, the scalar doesn't matter, $f(\alpha - at)$ converges to $f(\alpha)$, but this has nothing to do with μ.

But, we recall in the Lipschitz condition, we are really comparing $|f(\alpha) - f(\alpha - at)|$, can we introduce a $f(\alpha)$ in here? Indeed, let we assume that $\int_{-\infty}^{\infty} h(t)dt = 0$, then

$$(W_h f)(\alpha, a) = \sqrt{a} \int_{-\infty}^{\infty} [f(\alpha - at) - f(\alpha) + f(\alpha)]h(t)dt = \sqrt{a} \int_{-\infty}^{\infty} [f(\alpha - at) - f(\alpha)]h(t)dt + 0$$

Estimate using Lipschitz,

$$|W_h f(\alpha, a)| \leq \sqrt{a} \int_{-\infty}^{\infty} |f(\alpha - at) - f(\alpha)||h(t)|dt$$

$$\leq \sqrt{a} \int_{-\infty}^{\infty} K|at|^{\mu}|h(t)| \, dt \qquad \qquad \text{Lipschitz Continuity}$$

$$= a^{1/2+\mu} K \int_{-\infty}^{\infty} |t|^{\mu}|h(t)| \, dt$$

This suggests that we can detect μ locally by seeing how the Wavelet transform behaves with "a"!

Example 10.3. Consider the function $f(t) = |t|$, this is not smooth at $t = 0$. We see that

$$|f(t) - f(s)| = ||t| - |s|| \leq |t - s|^{\mu}$$

The highest value of μ should be 1! Because this is just the reverse triangle inequality. Let's consider

1. $h(t) = e^{-t^2/2}$, then

$$(W_h f)(0, a) = \sqrt{a} \int_{-\infty}^{\infty} |\alpha - at|e^{-t^2/2}dt$$

   There's no μ-dependent response here... This is no good because $h(t)$ is not an analyzing wavelet! ($\int_{-\infty}^{\infty} h(t)dt \neq 0$)

2. Let $h(t)$ be the Haar wavelet:
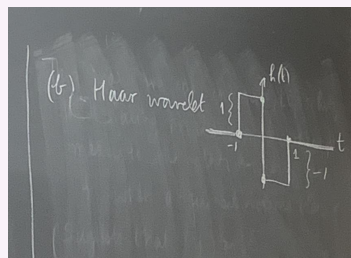
   and we have that

$$(W_h f)(0, a) = \sqrt{a} \int_{-\infty}^{\infty} h(t)|at|dt$$

$$= a^{3/2} \int_{-\infty}^{\infty} |t|h(t)dt$$

   This varies as $a^{3/2} = a^{1/2+1}$, BUT $\int_{-\infty}^{\infty} |t|h(t)dt = 0$, so this actually doesn't change as you change a. Hence this is not very useful. But we can still look at $(W_h f)(\alpha, a)$ with $\alpha \approx 0$.

3. Derivative Gaussian Wavelet - this is $h(t) = -\frac{d}{dt}e^{-t^2/2} = te^{-t^2/2}$, which graphically:

Unfortunately, this has the same problem as

$$(W_h f)(0, a) = a^{3/2} \cdot 0 = 0$$
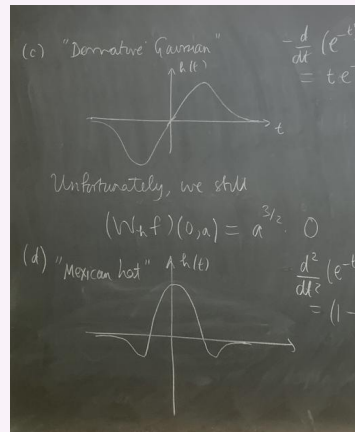
4. Let's consider $h(t) = \frac{d^2}{dt^2}(e^{-t^2/2}) = (1 - t^2)e^{-t^2/2}$, this is called the "Mexican Hat Wavelet", whose picture is given above. We have that

$$(W_h f)(0, a) = \sqrt{a} \int_{-\infty}^{\infty} |t| h(t) dt = -2a^{3/2}$$

THis does respond to $a^{1/2+\mu}$! We also observe that

$$(W_h f)(\alpha, a) = -2a^{3/2} e^{-\alpha^2/(2a^2)}$$

Plotting α against $a > 0$, recall that $e^{-\alpha^2/2a^2} \approx 0$ when $|\alpha/\sqrt{2}a| \geq 6$
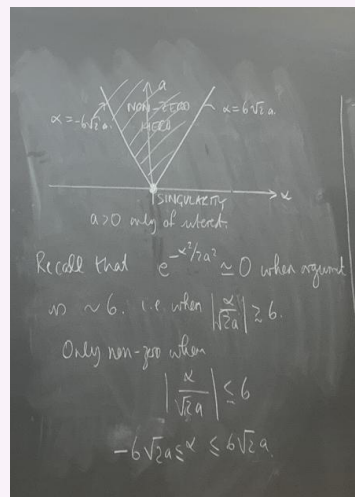
This points to where the singularity is, the rate of increase around the singularity is how bad the singularity is.

11 Lecture 11 - 2/17/2023

11.1 Higher Moments

We recall that last lecture, we obtained the following two equivalent formulation of the Continuous Wavelet Transform:

$$(W_h f)(\alpha, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) h(\frac{\alpha - t}{a}) dt$$

$$= \sqrt{a} \int_{-\infty}^{\infty} f(\alpha - as) h(s) ds \qquad \text{Let } s = \frac{\alpha - t}{a}$$

Question 11.1. What key property of the Mexican hat makes it a better filter for $f(s) = |s|$?

Indeed, let $h(t) = (1 - t^2) e^{-t^2/2}$, we first note

$$\widehat{h}(0) = \int_{-\infty}^{\infty} h(s) ds = 0$$

Thus, we have that

$$(W_h f)(\alpha, s) = \sqrt{a} \int_{-\infty}^{\infty} f(\alpha - as) h(s) ds$$

$$= \sqrt{a} \int_{-\infty}^{\infty} [f(\alpha - as) - f(\alpha)] h(s) ds$$

This allows us to use the Lipschitz property!

Question 11.2. Can we look at higher degrees of smoothness? (ie. when the derivatives are also smooth)

Indeed, we can think of the condition $\int_{-\infty}^{\infty} h(s) ds = 0$ as the fact that the zeroth moment of $h(s)$ is 0. If $h(s)$ vanishes on higher moments too, we can write

$$(W_h f)(\alpha, a) = \sqrt{a} \int_{-\infty}^{\infty} [f(a - \alpha s) - \text{linear polynomial}] h(s) ds$$

What is the term $f(a - \alpha s) - \text{linear polynomial}$? Well, we can write

$$f(s) = f(0) + s f'(0) + ...$$

using Taylor expansions. For the zero-th moment, we have that

$$f(s) - f(0) = s f'(0) + ...$$

In this case, we have filtered out the constant term (ie. the DC component). For higher order moments, we want to filter out more terms. To do this, we want the first n orders of $h(s)$ to be 0, that is

$$\int_{-\infty}^{\infty} s^k h(s) ds = 0, 0 \le k \le n$$

Example 11.3. If we look back at the Mexican Hat example again, we observe that

$$\int_{-\infty}^{\infty} s(1 - s^2) e^{-s^2/2} ds = 0$$

Hence, the first order moment of the filter is also 0.

This means that $(W_h f)(\alpha, a) \approx 0$ at non-zero values too!

If we recall the examples we used last time, the Gaussian filter $h(t) = e^{-t^2/2}$ fails because its zeroth moment is non-zero. The Haar wavelet fails because its first moment is non-zero, hence it is sensitive to derivatives. The first moment of the derivative Gaussian is similarly also non-zero.

11.2   Criterion for Selecting a Wavelet

In general we want out filter $h(s)$ to satisfy:

- $\int_{-\infty}^{\infty} s^m h(s) ds = 0$ for $m = 0, 1, ..., M$ where M is taken to be necessary. This implies that $h(s)$ should also be oscillating.

- $h(s)$ has rapid decay or compactly supported.

Note that sometimes we do not want to make M too big. For example,

$$\frac{d^{1066}}{dt^{1066}}(e^{-t^2/2})$$

is a good filter moment wise, but it is really complicated to work with because we will be dealing with $e^{-t^2/2} \times$ polynomial of degree 1066.
Thus, we have that

$$(W_h f)(\alpha, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) h(\frac{t - \alpha}{a}) dt$$

Generally, this integral is not directly solvable, so we would want to obtain a numerical approximation, for fixed α, a, we have

$$(W_h f)(\alpha, a) \approx \frac{1}{\sqrt{a}} \sum_{n \in \mathbb{Z}} f(n) h(\frac{n - \alpha}{a})$$

where $f(n)$ is thought of as the "digital representation" of f. This is because computers can't represent functions analytically, so we instead take at each digit.

Definition 11.4. Let $\{f(t) : t \in \mathbb{R}\}$ be a signal, we call

$$\{f(n) : n \in \mathbb{Z}\}$$

the digital representation of the signal.

12   Lecture 12 - 2/22/2023

12.1   Band-Limited Signals and Shannon Sampling Theorem

Consider the following problem:

> **Question 12.1.** We have a digital phone exchange where calls are transmitted as digital signals. $\{f(t) : t \in \mathbb{R}\}$ is thought of as an analog signal, no information is really lost here, but now we only have packets of the signal (thought of as digital), so we only know the signal at integer values. In other words, we only have the information $\{f(m) : m \in \mathbb{Z}\}$.
>
> The question is - how often do we need to sample the signal in order to get a perfect reconstruction at the other end? (... why can we even consider getting a perfect reconstruction? If there's no constraint on the function other than continuity, you can find infinity many functions fitting through the digital signal.)

At a mathematical level, the question makes no sense (because frequency can be arbitrarily high). However, in practice, we do have a constraint on the signal because we are thinking about this in the context of a "phone exchange", so we are only interested in certain frequency ranges. It is then natural to consider signals such that the frequency is limited:

$$\widehat{f}(\omega) = 0 \text{ if } |\omega| > \Omega$$

for some $\Omega \geq 0$.

Let's now reformulate our question:

> **Question 12.2** (Shannon's Problem). Suppose f is band-limited by Ω, how often do we need to sample f in order to reconstruct f from the digital samples?

12.2   Poisson Summation Formula

In order to answer this question, we derive a useful mathematical tool - the Poisson Summation Formula.

Suppose that $\widehat{f}(\omega)$ satisfies a decay condition (this holds for all band-limited signals but we are considering a more general case here):

$$|\widehat{f}(\omega)| \leq C(1 + |\omega|)^{-\alpha} \text{ for some } \alpha > 1 \ (*)$$

(Note that the condition for Lipschitz α we proved implies this, but the converse is not true).

To see that band-limited signals satisfy this, we can choose $C = (1 + |\Omega|)^{\alpha} \cdot \max_{|\xi| \leq \Omega} |\widehat{f}(\xi)|$, hence

$$|\widehat{f}(\omega)| \leq \frac{(1 + |\Omega|)^{\alpha}}{(1 + |\omega|)^{\alpha}} \max_{|\xi| \leq \Omega} |\widehat{f}(\xi)|$$

In particular, $(*)$ means that

$$\int_{-\infty}^{\infty} |\widehat{f}(\omega)| d\omega \leq C \int_{-\infty}^{\infty} \frac{d\omega}{(1 + |\omega|)^{\alpha}} < \infty$$

Hence we have that

$$|\int_{-\infty}^{\infty} e^{i\omega t} \widehat{f}(\omega) d\omega| \leq \int_{-\infty}^{\infty} |\widehat{f}(\omega)| d\omega < \infty$$

Thus, Fourier series is well-defined and f is well-defined.

Definition 12.3. Let $p > 0$ be given, and define a periodisation of $\widehat{f}$ as follows:

$$\widehat{f}_{per}(\omega) = \sum_{k \in \mathbb{Z}} \widehat{f}(w + kp)$$

Then, $\widehat{f}_{per}(\omega)$ is clearly periodic with period p.

Proposition 12.4. $\widehat{f}_{per}(\omega)$ is well-defined for all $\omega \in \mathbb{R}$.

Proof. To see that the sum converges, we only need to consider $\widehat{f}_{per}(\omega)$ for $\omega \in [0, p]$ and note that

$$
\begin{aligned}
\int_0^p |\widehat{f}_{per}(\omega)| d\omega &= \int_0^p |\sum_{k \in \mathbb{Z}} \widehat{f}(\omega + kp)| d\omega \\
&\leq \sum_{k \in \mathbb{Z}} \int_0^p |\widehat{f}(\omega + kp)| d\omega \\
&= \sum_{k \in \mathbb{Z}} \int_{kp}^{kp+p} |\widehat{f}(\xi)| d\xi \qquad\qquad \text{Let } \xi = \omega + kp \\
&= \int_{-\infty}^{\infty} |\widehat{f}(\xi)| d\xi \\
&< \infty
\end{aligned}
$$

$\blacksquare$

In particular, $\widehat{f}_{per}(\omega)$ has a discrete Fourier series.

$$\widehat{f}_{per}(\omega) = \sum_{n \in \mathbb{Z}} c_n e^{-2\pi i \omega n / p}, \ \omega \in [0, p]$$

where

$$c_n = \int_0^p \frac{1}{p} \widehat{f}_{per}(\omega) e^{2\pi \omega n / p} d\omega$$

Computing c_n explicitly:

$$
\begin{aligned}
c_n &= \int_0^p \frac{1}{p} \widehat{f}_{per}(\omega) e^{2\pi\omega n/p} d\omega \\
&= \frac{1}{p} \sum_{k\in\mathbb{Z}} \int_0^p \widehat{f}(\omega+kp) e^{2\pi i\omega n/p} d\omega \\
&= \frac{1}{p} \sum_{k\in\mathbb{Z}} \int_{kp}^{kp+p} \widehat{f}(\xi) e^{2\pi i(\xi-kp)n/p} d\xi && \text{Let } \xi = \omega + kp \\
&= \frac{1}{p} \sum_{k\in\mathbb{Z}} \int_{kp}^{kp+p} \widehat{f}(\xi) e^{2\pi i(\xi)n/p} \cdot 1 d\xi \\
&= \frac{1}{p} \sum_{k\in\mathbb{Z}} \int_{kp}^{kp+p} \widehat{f}(\xi) e^{2\pi i(\omega)n/p} d\omega \\
&= \frac{1}{p} \int_{-\infty}^{\infty} \widehat{f}(\omega) e^{2\pi i\omega n/p} d\omega \\
&= \left(\frac{2\pi}{p}\right)\left(\frac{1}{2\pi}\right) \int_{-\infty}^{\infty} \widehat{f}(\omega) e^{2\pi i\omega n/p} d\omega \\
&= \frac{2\pi}{p} f\left(\frac{2\pi n}{p}\right) && \text{By Fourier Inversion Formula}
\end{aligned}
$$

Hence, we obtain the Possion Summation Formula:

Theorem 12.5. Let $p > 0$ be fixed and $\widehat{f}$ be a function that satisfies then aforementioned decay condition, then

$$
\widehat{f}_{per}(\omega) = \frac{2\pi}{p} \sum_{n\in\mathbb{Z}} f\left(\frac{2\pi n}{p}\right) e^{-2\pi i\omega n/p}
$$

for all $\omega \in [0, p]$

This is really useful for proving various identities:

Corollary 12.6 (Basel Problem). $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$

Proof. Exercise!   ∎

Remark 12.7. The closed form value of $\sum_{n=1}^{\infty} \frac{1}{n^3}$ is not known! It's interesting to investigate why you can't use Possion Summation Formula to solve this.

13 Lecture 13 - 2/24/2023

13.1 Proof of the Shannon Sampling Theorem

Recall in the last lecture, we have written out the **Poisson Summation Formula**:

$$\sum_{k\in\mathbb{Z}}\widehat{f}(\omega+pk)=\frac{2\pi}{p}\sum_{n\in\mathbb{Z}}f(\frac{2\pi n}{p})e^{-2\pi in\omega/p}$$

We'll use this to answer the sampling question, but before that - let's rethink about what we want to do:

Question 13.1. Suppose we have f sampled at discrete points $f(0), f(\pm h), f(\pm 2h), ...$, and that f is band limited with bandwidth Ω. How small should h be (depending on Ω) such that we can reconstruct f exactly, ie. we want a formula of the form

$$f(t)=\sum_{n\in\mathbb{Z}}f(nh)\cdot[\text{some correcting term here}]$$

In particular, our correcting term should probably depend on t, n, and h, hence

$$f(t)=\sum_{n\in\mathbb{Z}}f(nh)\phi(t;n,h)$$

Then our question really is - how should we construct $\phi(t;n,h)$?

Here are some properties of ϕ that we would expect:

- $\phi(mh;m,h)=1$ and $\phi(mh;n,h)=0$ if $m\neq n$

- What is $\phi(t-h;m,h)$? Well,

$$f(t-h)=\sum_{n}f(n,h)\phi(t-h;n,h)=\sum_{n}f(nh+h)\phi(t;n,h)$$

  They really have the same coefficients shifted by a spot, so we have that

$$\phi(t-h;n,h)=\phi(t;n+1,h)$$

  More generally,

$$\phi(t-mh;n,h)=\phi(t;n+m,h)$$

  This means that the structure of ϕ is very simple. Choose $n=0$, then

$$\phi(t-mh;0,h)=\phi(t;m,h)\forall t$$

  Let's call $\varphi_0(t,h)=\phi(t;0,h)$, then ϕ is really just determined by what φ_0 is!

- We conclude that $\phi(t;m,h)$ is a translate of $\varphi_0(t;h)$

- $\varphi_0(t;h)$ has to be an even function with respect to t which decays as $|t|\to\infty$

- What's the last variable we haven't messed with? It's h! We also expect a "scaling property" in terms of how we'd change h,

$$\varphi_0(t;h)\propto\varphi_0(\frac{t}{h})$$

A pretty good suspect of a function that satisfies this is the **sink function**!

Theorem 13.2 (Shannon Sampling Theorem). Suppose $\widehat{f}(\omega) = 0$ for all $|\omega| > \Omega$ with $f \in L_2(\mathbb{R})$. If $0 < a \le \frac{\pi}{\Omega}$, then $f(t)$ is given by

$$f(t) = \sum_{k \in \mathbb{Z}} f(ka)\phi_s\left(\frac{t}{a} - k\right)$$

where $\phi_s(t) = \frac{\sin \pi t}{\pi t}$ is the "Shannon Sampling Function".

Proof. Let's start with the Poisson Summation Formula:

$$\sum_{k \in \mathbb{Z}} \widehat{f}(\omega + pk) = \frac{2\pi}{p} \sum_{n \in \mathbb{Z}} f\left(\frac{2\pi n}{p}\right) e^{-2\pi i n \omega/p}$$

Let's multiply both sides by

$$\sum_{k \in \mathbb{Z}} \widehat{f}(\omega + pk) e^{2\pi i \omega t/p} = \frac{2\pi}{p} \sum_{n \in \mathbb{Z}} f\left(\frac{2\pi n}{p}\right) e^{2\pi i \omega(t-n)/p}$$

Let's integrate this expression on both sides with respect to ω:

$$\sum_{k \in \mathbb{Z}} \int_{-p/2}^{p/2} \widehat{f}(\omega + pk) e^{2\pi i \omega t/p} d\omega = \frac{2\pi}{p} \sum_{n \in \mathbb{Z}} f\left(\frac{2\pi n}{p}\right) \int_{-p/2}^{p/2} e^{2\pi i \omega(t-n)/p} d\omega$$

The integral on the RHS gives us

$$
\begin{aligned}
\int_{-p/2}^{p/2} e^{2\pi i \omega(t-n)/p} d\omega &= \frac{p}{2\pi i(t-n)} \left[e^{2\pi i \omega(t-n)/p}\right]_{-p/2}^{p/2} \\
&= \frac{p}{2\pi i(t-n)} \left[\frac{2i \sin(\pi(t-n))}{1}\right] \\
&= p\frac{\sin \pi(t-n)}{\pi(t-n)} \\
&= p\phi_s(t-n)
\end{aligned}
$$

The integral on the LHS should give $f(t)$! But we need $p/2$ to be sufficiently large so that $p/2 > \Omega$ so that we can get the entire integral on the LHS (since $\widehat{f}(\omega)$ is band-limited). To do this, we choose $a = \frac{2\pi}{p}$. The condition that $p/2 > \Omega$ is then equivalent to $\frac{\pi}{a} > \Omega$.

$$
\begin{aligned}
LHS &= \sum_{k \in \mathbb{Z}} \int_{-p/2}^{p/2} \widehat{f}(w + kp) e^{2\pi i \omega t/p} d\omega \\
&= \int_{-p/2}^{p/2} \widehat{f}(\omega) e^{2\pi i \omega t/p} && \text{If } p/2 > \Omega, \text{ this is because all other values of } k \text{ have } \widehat{f}(w + kp) = 0 \\
&= \int_{-\infty}^{\infty} \widehat{f}(\omega) e^{2\pi i \omega t/p} d\omega && \text{Since } p/2 > \Omega, \text{ integrating over the support} \\
&= 2\pi f\left(\frac{2\pi t}{p}\right)
\end{aligned}
$$

Hence our equation becomes

$$2\pi f\left(\frac{2\pi t}{p}\right) = 2\pi \sum_{n \in \mathbb{Z}} f\left(\frac{2\pi n}{p}\right) \phi_s(t-n)$$

for $p > 2\Omega$.

So we have that

$$f(\frac{2\pi t}{p}) = \sum_{n \in \mathbb{Z}} f(\frac{2\pi n}{p})\phi_s(t - n)$$

Write $\frac{2\pi}{p} = a$, $p > 2\Omega$ is the same as $a < \frac{\pi}{\Omega}$, hence we have that

$$f(at) = \sum_{n \in \mathbb{Z}} f(an)\phi_s(t - n)$$

A change of variables gives the desired equation given by the Theorem and finishes the proof.   ∎

14   Lecture 14 - 2/27/2023

14.1   Oversampling

Consider the Shannon-Whittaker Formula, we have

$$f(t) = \sum_{n \in \mathbb{Z}} f(an)\phi_s\left(\frac{t}{a} - n\right)$$

provided that:

- f has bandwidth Ω and is in $L^2(\mathbb{R})$

- $a \leq \pi/\Omega$

- $\phi_s(t) = \frac{\sin \pi t}{\pi t}$ is the "Shannon Sampling Function"

There are some problems with this in practical applications:

- This formula is an infinite sum

- If we want to truncate the equation and approximate f, we get a slowly convergent series because

$$\phi_s(t) \sim \frac{1}{t} \text{ as } t \to \infty$$

Example 14.1. Consider the sum

$$S = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1}$$

How many terms are needed to get an accuracy 10^{-3}? You need about 1000 terms! However, we note that

$$S = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \dots$$

So in general we have that

$$\frac{1}{2r-1} - \frac{1}{2r} = \frac{+1}{2r(2r-1)}$$

So we have that

$$S = \sum_{r=1}^{\infty} \frac{1}{2r(2r-1)}$$

How many terms do we need to get to 10^{-3}? It's roughly $\sqrt{1000} \approx 30$! There are a lot of series where you can tweak to make it converge faster.

Question 14.2. Can we rewrite the Shannon-Whittaker Formula to make the sum converge faster? In other words, we seek a formula of the same form

$$f(t) = \sum_{n \in \mathbb{Z}} f(an)\phi(t/a - n), \ a \leq \pi/\Omega \ (\dagger)$$

that is exact when f has bandwidth Ω and such that $\phi(t)$ decays more rapidly as $t \to \infty$ than $\phi_s(t)$.

Let's start by trying to construct a counter-example:

> **Example 14.3.** Let $\Omega > 0$ be given, then choose
>
> $$f(t) = \phi_s(t/a) = \frac{\sin \pi t/a}{\pi t/a}$$
>
> for some fixed a. Then f has bandwidth π/a. We want to choose a such that $\pi/a \le \Omega$ for sanity.
>
> We also need $a \le \pi/\Omega$ because that's an additional constraint in our original question.
>
> Why don't we just choose $a = \pi/\Omega$, so f has bandwidth Ω. Plugging this into (†) gives us
>
> $$\phi_s(t/a) = \sum_{n \in \mathbb{Z}} \phi_s(an)\phi(t/a - n) = \sum_{n \in \mathbb{Z}} \delta_{n,0}\phi(t/a - n) = \phi(t/a)$$
>
> where $\delta_{n,0}$ is the Kronecker delta symbol that is 0 on all $n \ne 0$ and 1 on $n = 0$.
>
> Hence, we have that $\phi = \phi_s$, so we have a counter-example?

But ... there's some problem with the counter-example we just gave. If we choose $a = \pi/\Omega$, we always get our conclusion above. But what if we choose $a < \pi/\Omega$? Then the example above is no longer a counter-example.

If $a < \pi/\Omega$, then we have actually over-parametrized the system, there are redundancies in the formula because we have more data than we need. If $a = \pi/\Omega$, the formula has no redundancy! The extra redundancy in $a < \pi/\Omega$ let's us exploit new linear dependencies to achieve a faster decay.

The process of $a < \pi/\Omega$ is known as **oversampling**. The hope is that - by oversampling, we could get a formula that converges faster.

Let's then choose $a < \pi/\Omega$ (π/Ω is sometimes called the **Nyquist Rate**), so there is flexibility in the choice of ϕ.

Let's rethink about our proof of Shannon's Sampling Theorem, let's consider the Poisson SUmmation Formula:

$$\sum_{k \in \mathbb{Z}} \widehat{f}(w + kp) = \frac{2\pi}{p} \sum_{n \in \mathbb{Z}} f(2\pi n/p)e^{2\pi i n\omega/p}$$

Now let's suppose that f has bandwidth Ω, we should consider $\omega \in (-\Omega, \Omega)$ as this is periodic.

If $k \ge 1$, then $w + kp \ge w + p \ge -\Omega + p$. If p is large enough, we can choose $p \ge 2\Omega$ such that

$$w + kp \ge -\Omega + p \ge \Omega$$

Thus, if $p \ge 2\Omega$ and $k \ge 1$, then

$$\widehat{f}(w + kp) \equiv 0$$

On the other hand, if $k \le -1$, then $w + kp \le w - p \le \Omega - p \le -\Omega$ if $p \ge 2\Omega$.

Hence, if $p \ge 2\Omega$, the Poisson Summation Formula becomes

$$\widehat{f}(w) = \frac{2\pi}{p} \sum_{n \in \mathbb{Z}} f(\frac{2\pi n}{p})e^{-2\pi i n\omega/p}$$

for $\omega \in (-\Omega, \Omega)$.

Let's write $2\pi/a = p \ge 2\Omega$, then our condition is that $a \le \pi/\Omega$.

Let's choose $\widehat{g}$ such that $\widehat{g}(\omega) = 1$ for all $|\omega| \leq \Omega$, then we have that

$$\widehat{f}(\omega) = \widehat{f}(\omega)\widehat{g}(\omega) = a \sum_{n \in \mathbb{Z}} f(a_n)e^{-ia\omega n}\widehat{g}(\omega)$$

Taking the Inverse Fourier Transform on both sides gives

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

$$= a \sum_{n \in \mathbb{Z}} f(an)[\frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{g}(\omega)e^{i\omega(t-an)}d\omega]$$

$$= a \sum_{n \in \mathbb{Z}} f(an)g(t - an)$$

for all f with bandwidth less than Ω and that $a \leq \pi/\Omega$.

Thus, we can choose $\phi(t/a) = g(t)$ in †.

Remark 14.4. Note that $\widehat{g}$ can essentially be any function outside the interval $[-\Omega, \Omega]$, so the only constraint is that is has to be constant within the interval.

But if we are not over-sampling, ie. when $a = \pi/\Omega$, our previous counter-example kicks in again.

15   Lecture 15 - 3/1/2023

Recall last lecture, we have proven our new theorem:

Theorem 15.1. Let $\Omega > 0$,

$$\widehat{g}(\omega) = \begin{cases} 1, \ |\omega| \leq \Omega \\ \text{anything}, |\omega| > \Omega \end{cases}$$

If f has bandiwdth $\leq \Omega$, then

$$f(t) = a \sum_{n \in \mathbb{Z}} f(an)g(t - an)$$

for all $a \leq \pi/\Omega$.

This gives flexibility for us in the choice of g. We will list some notable applications:

Example 15.2 (Recovering one Lost Signal)**.** Choose

$$\widehat{g}(\omega) = \begin{cases} 1, \ |\omega| \leq \Omega \\ 0, \ |\omega| > \Omega \end{cases}$$

Then we have that

$$g(t) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} e^{i\omega t} d\omega = \frac{\sin \Omega t}{\pi t}$$

Then we have that

$$ag(t) = \frac{a\Omega}{\pi} \frac{\sin \Omega t}{\Omega t} = \frac{a\Omega}{\pi} \Phi_s(\frac{\Omega}{\pi} t)$$

where $\Phi_s(t) = \frac{\sin \pi t}{\pi t}$ is the usual Shannon Sampling Function. Applying the Theorem previously, we have that

$$f(t) = \frac{a\Omega}{\pi} \sum_{n \in \mathbb{Z}} f(an)\Phi_s(\frac{\Omega}{\pi}(t - an)), \ a \leq \pi/\Omega$$

When $a\Omega/\pi = 1$, then this reduces to $\frac{\Omega}{\pi} = 1/a$, which returns the Shannon-Whittaker formula. This is what we expect, and we get nothing new by choosing $a = \frac{\Omega}{\pi}$. This is because we are not oversampling.

Suppose we instead over-sample, ie. have that $a < \frac{\pi}{\Omega}$. We define the sampling rate β as

$$\beta := \frac{\pi}{\Omega a} > 1$$

(ie. if $\beta = 2$, we are sampling twice as often, etc.) Using the Theorem as before, formula then becomes,

$$f(t) = \frac{1}{\beta} \sum_{n \in \mathbb{Z}} f(\frac{\pi n}{\beta \Omega})\Phi_s(\frac{\Omega t}{\pi} - \frac{1}{\beta})$$

When $t = 0$, we have

$$f(0) = \frac{1}{\beta} \sum_{n \in \mathbb{Z}} f(\frac{\pi n}{\beta \Omega})\Phi_s(\frac{n}{\beta})$$

Observe that $f(0)$ appears on both sides. Look at the coefficients of $f(0)$ on the RHS, ie.

$$\frac{1}{\beta}\Phi_s(0)f(0) = \frac{1}{\beta}(1)f(0)$$

But since $\beta > 1$. we kmow that $\frac{1}{\beta}f(0) \neq f(0)$, so we have that

$$f(0) = \frac{\beta}{\beta - 1} \sum_{n \in \mathbb{Z} - \{0\}} f(\frac{\pi n}{\beta \Omega}) \Phi_s(\frac{n}{\beta})$$

Suppose we transmit $\{f(\frac{\pi n}{\beta \Omega}) : n \in \mathbb{Z}\}$ but the value at $n = 0$ is lost, we can actually recover it by oversampling.

More generally, if we replace $f(t)$ by $f(t + \frac{\pi m}{\beta \Omega})$, then we get

$$f(\frac{m\pi}{\beta \Omega}) = \frac{\beta}{\beta - 1} \sum_{n \neq 0} f(\frac{\pi}{\beta \Omega}(m + n)) \Phi_s(\frac{n}{\beta})$$

Let $\ell = m + n$, we have this denoising filter

$$f(\frac{m\pi}{\beta n}) = \frac{\beta}{\beta - 1} \sum_{\ell \neq m} f(\frac{\pi \ell}{\beta \Omega}) \Phi_s(\frac{\ell - m}{\beta}), \ m \in \mathbb{Z}$$

These relations are the basis for "digital remastering" or "noise reduction" in signal.

Let's examine another application. Despite being a big deal, these formulae are not usually used in the above form. This is because this is still an infinite sum that doesn't decay really fast.

Let's try to use the flexibility in the choice of $\widehat{g}$ to obtain a new (faster) converging formula.

The inverse Fourier Transform of g needs to decay faster! So we want to make $\widehat{g}$ decay slower (this is what the time-frequency inequality tells us). Let's try to extend the support from $[-\Omega, \Omega]$ to $[-\beta\Omega, \beta\Omega]$ for $\beta > 1$:

$$\widehat{g}(\omega) = \begin{cases} 1, \ |\omega| \leq \Omega \\ \text{linear}, \ \Omega \leq |\omega| \leq |\beta\Omega| \\ 0, \ \text{otherwise} \end{cases}$$

Converting $\widehat{g}$ to g gives:

$$g(t) = \frac{\cos \Omega t - \cos \beta \Omega t}{\pi t^2 (\beta - 1) \Omega}$$

This converges a lot faster, we observe that

- L'Hopital's Rule gives us

$$\lim_{\beta \to 1} g(t) = \frac{\sin \Omega t}{\pi t} \text{ (as expected)}$$

- We also have that

$$g(t) \sim \frac{1}{t^2} \text{ as } |t| \to \infty$$

- $\lim_{t \to 0} g(t) \frac{1}{2}\Omega(\beta + 1) \neq 1$. The fact that this doesn't equal to 1, we can do digital oversampling!

Question 15.3. What exactly do band-limited signals look like? If f is band-limited, how smooth is f? (Find out in Homework)

Question 15.4. What if f is not bandlimited (but smooth)? We can approximate

$$f(t) \approx a \sum_{n \in \mathbb{Z}} f(an) \Phi_s(\frac{t}{a} - n)$$

We expect the approximation to be good as $a \to 0$. Can you think of a bound of error for this? (This week's homework).

16   Lecture 16 - 3/3/2023

Recall in the last lecture, we derived a new sampling formula

$$f(t) = a \sum_{n \in \mathbb{Z}} f(an)g(t - an)$$

where $g(t) = \dots$ simplifies to get

$$f(t) = \sum_{n \in \mathbb{Z}} f(an)\phi(t - an) \ (*)$$

where $\phi(t) = \frac{\cos \Omega t - \cos \beta \Omega t}{\beta(\beta-1)t^2 \Omega^2}$ for $\beta > 1$. We made the observations:

- $\lim_{\beta \to 1} \phi(t) = \frac{\sin \Omega t}{\Omega t} \implies$ Shannon

- By applying L'Hopital's Rule twice, we have that $\lim_{t \to 0} \phi(t) = \frac{1+\beta}{2\beta}$. In particular, $\frac{1+\beta}{2\beta} = 1$ if and only if $\beta = 1$. Thus, if $\beta > 1$, we can obtain a reconstruction formula by solving for $f(0)$ in $(*)$ when $(t = 0)$.

- We can thus use $(*)$ to find that

$$f(0) = \frac{\beta}{2(\beta - 1)} \sum_{n \in \mathbb{Z} - \{0\}} f(\frac{\pi n}{\beta \Omega}) \frac{1}{\pi^2 n^2} [\cos \frac{\pi n}{\beta} - \cos \pi n]$$

  Observe that this formula decays close to $\frac{1}{n^2}$.

- This means that we can recover $f(0)$ without knowing what $f(0)$ is, by using the samples $\{f(\frac{\pi n}{\beta \Omega}) \ : \ n \in \mathbb{Z} - \{0\}\}$.

16.1   Filtering and Denoising

A slightly different question is to ask - suppose we transmit the digital packet

$$\{f(\frac{\pi n}{\beta \Omega}) \ : \ n \in \mathbb{Z}\}$$

but they are subject to random noise, so we have received

$$\{f(\frac{\pi n}{\beta \Omega}) + \epsilon_n \ : \ n \in \mathbb{Z}\}$$

where ϵ_n is random.

> **Question 16.1.** How well can we recover the original signal?

We can use $(*)$ to denoise! Let

$$f(t) = \sum_{n \in \mathbb{Z}} f(an)\phi(t - an)$$

Let $t = am$ for $m \in \mathbb{Z}$, we have that

$$f(am) = \sum_{n \in \mathbb{Z}} f(an)\phi(a(m - n))$$

> **Definition 16.2.** Let f^{noisy} be the function altered by noise, we define,
>
> $$f^{new}(am) := \sum_{n \in \mathbb{Z}} f^{noisy}(an)\phi(a(m - n))$$

We think of $\phi(a(m-n))$ as a form of weighting and $f^{new}(am)$ as the wegithed sum of the noises.

Let's instead consider the packet $\{f^{new}(am) : m \in \mathbb{Z}\}$. We expect the noise of this new function to be smaller (by APMA 1650 - taking an average always reduces the variance). We can refine the function by plugging $f^{new}(am)$ into the denoising function above again, and do it recursively down.

Remark 16.3. Oversampling increases the bandwidth we can collect, but it actually impedes the smoothing process of the signal. So our denoising process really depends on how much of the bandwidth we are asking to sample over. If it's around the bandwidth of the original signal, we have a really good approximation. However, if the original bandwidth is unknown, this is a really difficult question on how to recover the original signal.

Conjecture 16.4. If we choose our filter such that the bandwidth is the exactly the bandwidth of the original signal, we can recover $\{f(\frac{\pi n}{\beta n}) + \epsilon_n : n \in \mathbb{Z}\}$ exactly (most of the time - there are some counter example, for example choose $\epsilon_n = 1$ for all n, this is not actually a noise but the probability of all ϵ_n being the same is really low ... but the question is, on average (the expected probability), you will always get the original signal). This is a research question that even the instructor does not know.

17 Lecture 17 - 3/6/2023

17.1 Signal Spaces

In the previous section, we showed that (under quite general conditions on g), that any band-limited signal can be represented digitally with:
$$f(t) = a \sum_{n \in \mathbb{Z}} f(an)g(t - an)$$

provided that $a \leq \frac{\pi}{\Omega}$ and
$$\widehat{g}(\omega) = \begin{cases} 1, \ |\omega| \leq \Omega \\ \text{anything , otherwise} \end{cases}$$

Definition 17.1. Fix $g(t - an)$ and a, This suggests that
$$V_{g,a} := \overline{\text{span}\{g(t - an) \ : \ n \in \mathbb{Z}\}}$$

is a good space for representing signals digitally. Any signal can be represented by
$$\sum_{n \in \mathbb{Z}} c_n g(t - an)$$

for appropriate digital information $\{c_n : n \in \mathbb{Z}\}$. We call $V_{g,a}$ a **signal space**.

Definition 17.2 (Scalar Product on Signal Space)**.** There's a canonical scalar product on $V_{g,a}$:
$$f, h \in V_{g,a} \mapsto (f, h)_* := \int_{\mathbb{R}} f(t)\overline{h}(t)dt$$

There's a question on whether this value is finite, but as long as we require our signal to have finite energy, this will be well-defined.

Question 17.3. Is the digital representation **stale**? That is, suppose $f \in V_{g,a}$ and we write
$$f(t) = \sum_{n \in \mathbb{Z}} c_n g(t - an)$$

We assume that the c_n are uniquely determined (choose $g(t-an)$ linearly independent). This defines a bijective mapping
$$f \mapsto \{c_n : n \in \mathbb{Z}\}$$

This is a bijective map from $V_{g,a} \to \mathbb{C}[\mathbb{Z}]$ (where $\mathbb{C}[\mathbb{Z}]$ is thought of as the **sequence space** of countable length with coefficients in $\mathbb{C}$).

The question of staleness is - is the map continuous?

How are we going to define a topology on $\mathbb{C}[\mathbb{Z}]$, we could choose the ℓ_∞ norm:
$$||\{c_n : n \in \mathbb{Z}\}|| := \sup_{n \in \mathbb{Z}} |c_n| \sim \ell_\infty$$

But we wouldn't choose this, instead, let's consider the ℓ_2 norm:
$$||\{c_n : n \in \mathbb{Z}\}|| := \left(\sum_{n \in \mathbb{Z}} |c_n|^2\right)^{1/2} \sim \ell_2$$

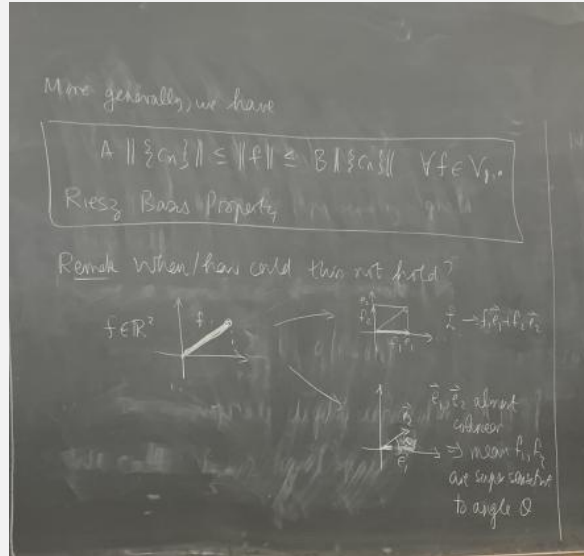To show that the math is continuous, we really want a constant map $A > 0, B > 0$ (finite) such that

$$A||\{c_n\} - \{\tilde{c}_n\}|| \le ||f - \tilde{f}|| \le B||\{c_n\} - \{\tilde{c}_n\}||$$

This is what's typically known as a **norm equivalence**. More generally, we have

$$A||\{c_n\}|| \le ||f|| \le B||c_n|| \; \forall f \in V_{g,a}$$

This is known as the **Riesz Basis Property**

> **Remark 17.4.** When / How could this Riesz Basis Property not hold?
>
>
>
> We call this phenomenon **super sensitivity**. In particular, A and B only depends on the choice of bases $\{g(t - an)\}_{n \in \mathbb{Z}}$.

Ideally, we want to have an *orthonormal basis*, ie.

$$\int_{\mathbb{R}} g(t - an)\overline{g(t - am)}dt = \begin{cases} 1, & \text{if } m = n \\ 0, & \text{if } m \ne n \end{cases}$$

> **Proposition 17.5.** If $\{g(t - an)\}$ is an orthonormal basis, then it satisfies the Riesz Basis Property

Proof. In the case of orthonormal basis, for any

$$f(t) = \sum_{n \in \mathbb{Z}} c_n g(t - an)$$

We have that

$$\begin{aligned}
||f||^2 &= (\sum c_n g(t - an), \sum c_m g(t - am)) \\
&= \sum_m \sum_n c_n \overline{c_m}(g(t - an), g(t - am)) \\
&= \sum_m |c_m|^2 \\
&= ||\{c_n\}||^2
\end{aligned}$$

Choose $A = B = 1$, we have that

$$A||\{c_n\}|| \leq ||f|| \leq B||c_n|| \ \forall f \in V_{g,a}$$

■

Example 17.6. Consider the Shannon-Whittaker Basis, where

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

We have two questions:

1. Is $\{g(t - an)\}$ an Riesz basis?

2. Is $\{g(t - an)\}$ an orthonormal basis?

Let $a = 1$, consider the integral below. We will evaluate this using Parserval's Identity:

$$\int_{\mathbb{R}} \phi_s(t - n)\phi_s(t - m)dt$$

What is the Fourier Transform of $t \mapsto \phi_s(t - m)$? This is the mapping:

$$\omega \mapsto \int_{\mathbb{R}} e^{-i\omega t}\phi_s(t - m)dt = e^{-i\omega m}\int_{-\infty}^{\infty} e^{-i\omega t}\phi_s(t)dt$$

$$= e^{-i\omega m} \cdot \begin{cases} 1, & |\omega| < \pi \\ 0, & \text{otherwise} \end{cases}$$

Let's call $\chi(\omega)$ the function $\begin{cases} 1, & |\omega| < \pi \\ 0, & \text{otherwise} \end{cases}$. Then by Parserval's Identity, we have that

$$\int_{\mathbb{R}} \phi_s(t - n)\phi_s(t - m)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega n}\chi(\omega)e^{i\omega m}\overline{\chi(\omega)}d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega(m-n)}d\omega$$

$$= \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases}$$

Hence, the answer to Question (1) and (2) are both true!

Corollary 17.7. The Shannon-Whittaker Basis is an orthonormal basis, hence is a Riesz Basis with $A = B = 1$.

18   Lecture 18 - 3/8/2023

18.1   Best Approximations and Projections

Suppose we want to represent a signal $f(t) \approx \sum_{n \in \mathbb{Z}} c_n \phi(t - an)$ with $t \in \mathbb{R}$. In other words, we wish to approximate f by an element of the signal space $V_{\phi,a} = Span\{\phi(t - an) : n \in \mathbb{Z}\}$ and ϕ is some appropriate function.

> **Question 18.1.** How can we choose the $\{c_n\}$ to approximate this desired $f(t)$?

There are two options:

1. We can consider *digital sampling*, ie.
$$c_n = f(an), n \in \mathbb{Z}$$

   then we have that
$$f(t) \approx (V_a^\phi f)(t) := \sum_{n \in \mathbb{Z}} f(an)\phi(t - an)$$

   Is this a good idea? Well, the approximation above is exact when ϕ is a Shannon function and f is band-limited by Ω, and $a \le \frac{\pi}{\Omega}$.

   However, this may not be a good idea if any of these hypothesis are not satisfied.

2. *Best possible choice*:
   The idea here is to choose $\{c_n\}$ such that we minimize the error

$$\varepsilon(f; \{c_n\}) = ||f(t) - \sum_{n \in \mathbb{Z}} c_n \phi(t - an)||^2$$

   Let's differentiate this term with respect to c_j (let's assuming f, c_n, ϕ all real), hence

$$0 = \frac{1}{2} \frac{\partial \varepsilon}{\partial c_j}$$
$$= \langle f(t) - \sum c_n \phi(t - an), \phi(t - aj) \rangle \qquad \text{The inner product}$$
$$\iff \langle f(t), \phi(t - aj) \rangle = \sum_n c_n \langle \phi(t - an), \phi(t - aj) \rangle \; \forall j \in \mathbb{Z}$$
$$\iff \sum_n M_{jn} c_n = f_j \; \forall j \in \mathbb{Z}$$

   f_j is defined as $\langle f(t), \phi(t - aj) \rangle$ and $M_{jn} := \langle \phi(t - an), \phi(t - aj) \rangle$.

   This is equivalent to the problem
$$M\vec{c} = \vec{f}$$

   where $M = [M_{jn}]$ and $\vec{f} = [f_j]$ and $\vec{c} = [c_n]$, where we ended up with an infinite matrix equation.

   Let's suppose that we have an orthonormal basis, then M_{jn} is really the infinite dimensional identity matrix. In this case, we have that
$$c_n = f_n = \langle f(t), \phi(t - an) \rangle \; \forall n \in \mathbb{Z}$$

   One should also use the second derivative test to check this is really a minimum, that's left as an exercise.

This gives us the following proposition:

Proposition 18.2. The best approximation (optimal c's) given by

$$(L_a f)(t) := \sum_{n \in \mathbb{Z}} c_n \phi(t - an)$$

where either the $\{c_n\}$ determined by either

- If $M\vec{c} = \vec{f}$ $(\star)$

- Furthermore, if $\{\varphi(t - an) : n \in \mathbb{Z}\}$ is orthonormal, we have that

$$c_n = \langle f(t), \phi(t - an) \rangle$$

Remark 18.3. Suppose f is band-limited again by Ω and $a \leq \frac{\pi}{\Omega}$ with ϕ chosen to be Shannon Sampling function (or similar) again. What can we say about c_n given by $(\star)$?

They should actually coincide with $f(an)$ because the minimum of the optimization in Best Approximation is unique. We know $V_a^\phi f = f$ in this case and $||f - L_a f|| \leq ||f - V_a^\phi f|| = 0$, hence $L_a f = f - V_a^\phi f$.

This leads us to suspect that if we choose φ such that $\hat{\varphi}$ looks like the step function from $-\Omega$ to Ω, ie. if

- $\hat{\varphi}(0) = 1$ and $\hat{\varphi}'(0) = \hat{\varphi}''(0) = ... = \hat{\varphi}^{(m)}(0) = 0$ for m as large as possible

then φ should be a good sampling function.

$$\hat{\varphi}(\omega) \approx 1 \text{ for } |\omega| \text{ in some range}$$

These are the conditions for a **Butterworth filter**, and, in this case, the best approximation $L^a f(t)$ is about the same as $(V_a^\phi f)(t)$. Hence, we can use Digital Sampling as a "good enough" approximation.

Theorem 18.4. Suppose that

- (a) $\{\phi(t - an) : n \in \mathbb{Z}\}$ is orthonormal

- (b) $\hat{\phi}(\omega) \approx 1$ for $|\omega| \leq \Omega$

Then for f bandlimited by Ω, we have that

$$V_a^\phi f \approx L_a f \text{ for } a \leq \pi/\Omega$$

Proof. By Parserval's Identity,

$$c_k^\star := \int_{-\infty}^{\infty} f(t)\phi(t - ak)\, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(\omega) e^{i\omega k a} \overline{\hat{\phi}(\omega)} d\omega$$

$$= \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \widehat{f}(\omega) e^{i\omega k a} \overline{\hat{\phi}(\omega)} d\omega$$

$$\approx \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \widehat{f}(\omega) e^{i\omega k a} d\omega \qquad\qquad \text{Since } \hat{\phi}(\omega) \approx 1$$

$$= f(ka) \qquad\qquad\qquad \text{Definition of Inverse Fourier Transform}$$

$\blacksquare$

19   Lecture 19 - 3/10/2023

19.1   Scaling and Two Scale Relation

We have proposed to use spaces of form

$$V_{\phi,a} = \text{span}\{\phi(t - an) : n \in \mathbb{Z}\}$$

to represent signals in the form

$$f(t) \approx \sum_{n \in \mathbb{Z}} c_n \phi(t - an)$$

provides that $\hat{\phi}$ satisfies same assumptions. If $a \to 0$, we expect the approximations to improve. This is because the Bandwidth Ω can grows as π/a, so we would get more information about the signal.

We have also shown that choosing $c_n = f(an)$ is a near optimal choice (under reasonable circumstances on f), a question we are now interested in is

> **Question 19.1.** How to choose a? When do we know that a is sufficiently small?

We know that we have a good approximation when $\Omega \leq \frac{\pi}{a}$! But what if f has a wider bandwidth? In the previous lectures, we have doubled the sampling rate by changing $a \mapsto a/2$ so we can represent the bandwidth up to $\Omega \leq \frac{2\pi}{a}$. In general, we can keep doubling the sampling rate by sending $a \mapsto a/2^n$ to find a sufficiently high bound on the bandwidth.

To adjust the bound on Ω, we are in effect using a family of signal spaces of the form ($a = 1$, for convenience):

$$V_n = V_{\phi,2^{-n}} = \text{span}\{\phi(2^n t - k) \mid k \in \mathbb{Z}\}$$

for $n \in \mathbb{Z}$. (Note that we can rewrite $\phi(t - an)$ as $\phi(\frac{t}{a} - n)$ up to some reparametrization).

The issue is that we not have an infinite signal spaces.

> **Remark 19.2.** Suppose ϕ is real-valued with center $t_* = 0$ and spread Δ_ϕ, then the map $t \mapsto \phi(2^n t - k)$ would satisfy:
>
> - For energy, we have that $\int_{-\infty}^{\infty} |\phi(2^n t - k)|^2 dt = 2^{-n} \int_{-\infty}^{\infty} |\phi(s)|^2 ds$ by consider the substitution $s = 2^n t - k$.
>
> - For center, we have that
>
> $$\int_{-\infty}^{\infty} t|\phi(2^n t - k)|^2 dt = 2^{-n} \int_{-\infty}^{\infty} |\phi(s)|^2 2^{-n}(k + s) ds \qquad \text{Same Substituion}$$
>
> $$= 2^{-2n} k \int_{-\infty}^{\infty} |\phi(s)|^2 ds \qquad \text{Since } t^* = 0$$
>
> $$= 2^{-n} k \int_{-\infty}^{\infty} |\phi(2^n t - k)|^2 dt$$
>
> Hence, $t \to \phi(2^n t - k)$ has center at $2^{-n}k$.
>
> - For radius, we have that
>
> $$\int_{-\infty}^{\infty} (t - 2^{-n}k)^2 |\phi(2^n t - k)|^2 dt = \dots$$
>
> Since this is a Friday morning, the instructor decided to not calculate the radius but heuristically guess what the radius is. But we note that the multiplication by 2^n in $\phi(2^n t - k)$ would make the signal a lot

tighter, so its radius should be smaller. We eventually find that

$$t \mapsto \phi(2^n t - k) \text{ has radius } 2^{-n}\Delta_\phi$$

In Summary, we compare the two signals $\phi(t)$ and $\phi(2^n t - k)$ as follows:
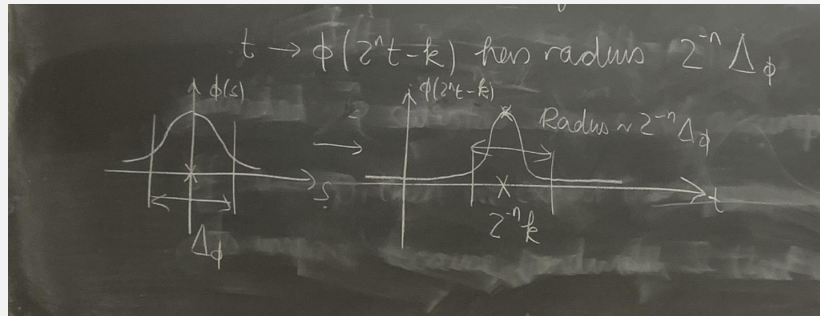
This means that if we have an expansion of the form

$$V_n \ni f(t) = \sum_{k \in \mathbb{Z}} c_{n,k} \phi(2^n t - k)$$

then $c_{n,k}$ is gicing information about f in the neighborhood of the point $t = 2^{-n}k$ and a region of size $\sim 2^{-n}\Delta_\phi$. This gives us the expectation that

$$c_{n,k} \sim f(2^{-n}k) = f(ak)$$

where $a = 2^{-n}$ and $2^{-n}k = ak$. Hence we really do great better resolution by considering the family $\{V_n\}_{n \in \mathbb{Z}}$
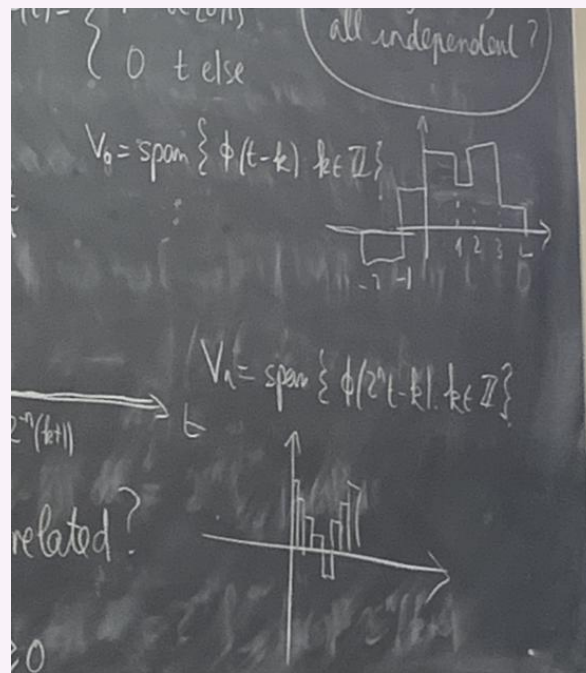
This is all very well, but we still need an infinite number of spaces! But here's a thought - are the $\{V_n\}_{n \in \mathbb{Z}}$ really all independent?

Example 19.3. Let's choose $\phi(t) = \begin{cases} 1, & t \in [0,1] \\ 0, & \text{otherwise} \end{cases}$, then $\phi(t)$ and $\phi(2^n t - k)$ looks like:
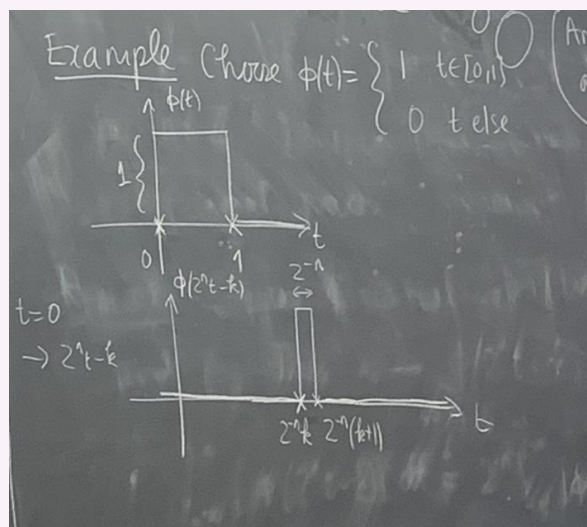
We observe that $V_0 = \text{span}\{\phi(t-k) : k \in \mathbb{Z}\}$ and $V_n = \text{span}\{\phi(2^n t - k) : k \in \mathbb{Z}\}$ A function in V_0 and V_n

look like

Judging from the image, it seems plausible that $V_0 \subset V_n$ for $n \geq 0$.

More generally, iterating this relation, we have the following claim

Proposition 19.4. When $\phi(t) = \begin{cases} 1, & t \in [0, 1] \\ 0, & \text{otherwise} \end{cases}$, the family $\{V_n : n \in \mathbb{Z}\}$ is nested:

$$... V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset ...$$

The proposition above is not true in general! Hence, there's something special about this choice of $\phi(t)$ that makes it true. What's special about $\phi(t)$ is that $\phi(t)$ is a linear combination of $\phi(2t - k)$ for $k \in \mathbb{Z}$:

$$\phi(t) = \phi(2t - 1) + \phi(2t)$$

This is what's called a **Two-Scale Relation**.

Definition 19.5. More generally, if we can find scalars $\{p_j\}$ such that

$$\phi(t) = \sum_{j \in \mathbb{Z}} p_j \phi(2t - j)$$

Then we say that ϕ satisfies a **Two- Scale Relation**.

Theorem 19.6. If $\phi(t)$ satisfies a 2-scale relation, then $\{V_n\}$ are nested.

We will prove this theorem next time.

20   Lecture 20 - 3/13/2023

Last lecture, we stated the following theorem:

> **Theorem 20.1.** If $\phi(t)$ satisfies a 2-scale relation, that is, we can find scalars $\{p_j\}$ such that
>
> $$\phi(t) = \sum_{j \in \mathbb{Z}} p_j \phi(2t - j)$$
>
> Then the associated
>
> $$\{V_n^\phi := \mathrm{span}\{\phi(2^n t - k) \mid k \in \mathbb{Z}\}$$
>
> are nested.

Proof. For any $n \in \mathbb{Z}$, we will show that $V_n^\phi \subset V_{n+1}^\phi$. Indeed, let $u \in V_n^\phi$, we can write

$$u = \sum_{k \in \mathbb{Z}} c_{n,k} \phi(2^n t - k)$$

for suitable scalars $\{c_{n,k} \;:\; k \in \mathbb{Z}\}$. By the Two Scale-Relation, we can write $\phi(2^n t - k)$ as

$$\phi(2^n t - k) = \sum_{j \in \mathbb{Z}} p_j \phi(2[2^n t - k] - j) = \sum_{j \in \mathbb{Z}} p_j \phi(2^{n+1} t - 2k - j)$$

Substituting this back into the summation for u:

$$
\begin{aligned}
u &= \sum_{k \in \mathbb{Z}} c_{n,k} \phi(2^n t - k) \\
&= \sum_{k \in \mathbb{Z}} c_{n,k} \sum_{j \in \mathbb{Z}} p_j \phi(2^{n+1} t - 2k - j) \\
&= \sum_{k \in \mathbb{Z}} c_{n,k} \sum_{i \in \mathbb{Z}} p_{i-2k} \phi(2^{n+1} t - i) &&\text{Let } i = 2k + j \\
&= \sum_{i \in \mathbb{Z}} [\sum_{k \in \mathbb{Z}} p_{i-2k} c_{n,k}] \phi(2^{n+1} t - i)
\end{aligned}
$$

Note that $[\sum_{k \in \mathbb{Z}} p_{i-2k} c_{n,k}]$ is a suitable scalar, and by definition we have that $u \in \mathrm{span}\{\phi(2^{n+1} t - i) \mid i \in \mathbb{Z}\} = V_{n+1}^\phi$.

Furthermore, this equation shows how we would transform/rewrite a function in V_n^ϕ as a function in V_{n+1}^ϕ. ∎

> **Question 20.2.** What happens as $n \to \infty$? Is there a limit of V_n^ϕ as $n \to \infty$?

Let's be a bit more precise about V_n^ϕ to have a convergent behavior:

- We obviously want to consider functions that do converge, hence

$$V_n^\phi = \{\sum_{k \in \mathbb{Z}} c_k \phi(2^n t - k) \;:\; \{c_k\} \in \ell_2\}$$

where $\{c_k\} \in \ell_2$ means that

$$||\{c_k\}||^2 := \sum_k |c_k|^2 < \infty$$

Hence we really want our coefficients to be in ℓ_2 for a nice convergence property.

- We also want $f = \sum_{k \in \mathbb{Z}} c_k \phi(2^n t - k) \in V_n^\phi$ to satisfy the Resiz Basis Property, so

$$A||\{c_k\}|| \leq ||f|| \leq B||\{c_k\}||$$

Note that this condition forces $||f|| < \infty$.

We can derive the following properties based on these two assumptions.

> **Proposition 20.3.** V_n^ϕ is a subset of $L^2(\mathbb{R})$.

Proof. For any $u \in V_n^\phi$:

- We have that, by the Riesz Basis Property

$$\int_{-\infty}^{\infty} |u(t)|^2 dt = ||u||^2 \leq B||\{c_k\}|| < \infty$$

- If $u \in L_2$, then by the Riesz Basis Property

$$A||\{c_k\}||^2 \leq ||u||^2 < \infty$$

♣♣♣ Mattie: [I am uncertain why the second part is needed.]

Hence we have that $V_n^\phi \subset L^2(\mathbb{R})$. ∎

> **Question 20.4.** Suppose we have the nested V_n^ϕ again such that
>
> $$...V_n^\phi \subsetneq V_{n+1}^\phi \subsetneq ...$$
>
> Is the union of this increasing sequence $L^2(\mathbb{R})$? To formulate this question more rigorously, suppose $f \in L^2(\mathbb{R})$ and $q > 0$, for any $\epsilon > 0$, does there exist some $n \in \mathbb{Z}$ and $f_n \in V_n^\phi$ such that $||f - f_n|| < \epsilon$?

> **Example 20.5.** Suppose $\phi(t)$ is the indicator function on $[0, 1]$. Let $f \in L^2(\mathbb{R})$, can we find some $n \in \mathbb{Z}$ and $f_n \in V_n^\phi$ such that $||f - f_n|| < \epsilon$?
>
> Well, we can always write
>
> $$f_n(t) = \sum_{k \in \mathbb{Z}} c_k \phi(2^n t - k)$$
>
> So this really amounts to choosing suitable coefficients c_k instead. Let $h = 2^{-n}$, then consider
>
> $$c_k = \frac{1}{h} \int_{-\infty}^{\infty} f(t) \phi(\frac{t}{h} - k) dt$$
>
> Since $\phi(t)$ is the indicator function given as before, we note that
>
> $$\phi(\frac{t}{h} - k) \neq 0 \iff 0 \leq \frac{t}{h} - k \leq 1 \iff kh \leq t \leq (k+1)h$$
>
> Hence we have that
>
> $$\begin{aligned} c_k &= \frac{1}{h} \int_{kh}^{kh+h} f(t) \phi(\frac{t}{h} - k) dt \\ &= \frac{1}{h} \int_{kh}^{kh+h} f(t)(1) dt && \phi \text{ is an indicator function} \\ &= \frac{1}{h} \int_{kh}^{kh+h} f(t) dt \end{aligned}$$

We observe that, if $t \in [kh, kh + h]$, then $f_n(t) = c_k$, hence

$$f(t) - f_n(t) = f(t) - c_k$$

21 Lecture 21 - 3/15/2023

21.1 Multi-Resolution Analysis

Lat lecture we proved that

> **Theorem 21.1.** Suppose ϕ satisfies the Two-Scale relation, ie.
>
> $$\phi(t) = \sum_{j \in \mathbb{Z}} p_j \phi(2t - j)$$
>
> Then the spaces
>
> $$V_n^{\phi} := \text{span}\{\phi(2^n t - k) \mid k \in \mathbb{Z}\}$$
>
> are nested
>
> $$\subset V_0^{\phi} \subset V_1^{\phi} \subset \dots$$

We then asked if the sequence of spaces have any convergent behavior, so we added the constraints that

$$V_n^{\phi} := \{\sum_{k \in \mathbb{Z}} c_{n,k} \phi(2^n t - k) \; : \; ||\{c_{n,k}\}||_{\ell^2} < \infty\}$$

and letting the spaces satisfy the Riestz Basis Property. Then we showed that the sequences are effectively all in $L^2(\mathbb{R})$.

> **Question 21.2.** Suppose $f \in L_2(\mathbb{R})$ and $\epsilon > 0$, does there exists $n \in \mathbb{Z}$ and an $f_n \in V_n^{\phi}$ such that
>
> $$||f - f_n|| < \epsilon$$
>
> The fancy way of saying this is that these spaces are dense in $L^2(\mathbb{R})$.

We investigate the particular case when $\phi(t)$ is the indicator function on $[0, 1]$. The idea is to let $h = 2^{-n}$ with n to be determined, and choose

$$f_n(t) = \sum_{k \in \mathbb{Z}} c_k \phi(2^n t - k)$$

The function $\phi(2^{-n} t - k)$ is supported on $[2^{-n}k, 2^{-n}(k+1)] = [kh, kh + h]$ and we take

$$c_k = \frac{1}{h} \int_{kh}^{kh+h} f(t) dt$$

to be the average value on the interval supported.

Then for $kh \le t \le kh + h$, we have the error

$$f(t) - f_n(t) = f(t) - c_k$$

Then we observe that

$$
\begin{aligned}
||f - f_n||^2_{[kh, kh+h]} &= \int_{kh}^{kh+h} [f(t)^2 - 2f(t)c_k + c_k^2] dt \\
&= \int_{kh}^{kh+h} (f(t)^2 - c_k^2) dt - 2c_k \int_{kh}^{kh+h} (f(t) - c_k) dt \\
&= \int_{kh}^{kh+h} (f(t)^2 - c_k^2) dt \qquad\qquad \text{By definition of } c_k \\
&= -hc_k^2 + \int_{kh}^{kh+h} [f(t)^2] dt
\end{aligned}
$$

Summing over all $k \in \mathbb{Z}$, we have that

$$
||f - f_n||^2 = \int_{-\infty}^{\infty} f(t)^2 dt - h \sum_{k \in \mathbb{Z}} c_k^2
$$

But we note that the $h \sum_{k \in \mathbb{Z}} c_k^2$ is the Riemann Sum for $||f||^2$, so as $n \to \infty$, h would tend to 0, so

$$
\lim_{n \to \infty} ||f - f_n||^2 = 0
$$

To summarize:

> **Proposition 21.3.** We have that $\phi(t)$ is the indicator function of $[0, 1]$
>
> - $\{\phi(t - k) \ : \ k \in \mathbb{Z}\}$ is a Riesz basis (because they are orthonormal)
>
> - $\phi(t)$ satisfies a Two Scale Relation:
>
> $$
> \phi(t) = \phi(2t) + \phi(2t - 1)
> $$
>
> - V_n^ϕ satisfies
>
> $$
> ... \subset V_n^\phi \subset V_{n+1}^\phi \subset ... \to L^2(\mathbb{R})
> $$
>
>   ie. V_n^ϕ are fense in $L_2(\mathbb{R})$.
>
> Another way of saying this is - $\phi(t)$ generates a "**Multi-Resolution Analysis**" (MRA) of L_2.

> **Remark 21.4.** Historically, S. Mallot invented the term Multi-Resolution Analysis and won the Abel Prize for work in MRA.

> **Question 21.5.** Are there any other ϕ that generates an MRA of L_2?

How about the Shannon Sampling Function $\phi(t) = \frac{\sin \pi t}{\pi t}$?

It is certainly orthonormal, so it is a Riesz basis, but does it satisfy the Two Scale relation? We first note that

> **Definition 21.6.** The equivalent definition of Two-Scale Relation in Frequency Domain:
>
> $$
> \widehat{\phi}(2\omega) = P(e^{-i\omega}) \widehat{\phi}(\omega)
> $$
>
> where $P(z)$ is some power series.

> **Proposition 21.7.** The Shannon Sampling function satisfies the two scale relation.

Proof. We wish to write $\phi(t) = \sum_{j \in \mathbb{Z}} p_j \phi(2t - j)$.

Let's replace t with $t/2$, then

$$\phi(2t) = \sum_{j \in \mathbb{Z}} p_j \phi(t - j)$$

Let's take the Fourier Transform on both sides, then

$$2\widehat{\phi}(2\omega) = \sum_{j \in \mathbb{Z}} p_j e^{-i\omega j} \widehat{\phi}(\omega)$$

$$= [\sum_{j \in \mathbb{Z}} p_j e^{-i\omega j}] \widehat{\phi}(\omega)$$

Hence we have that

$$\widehat{\phi}(2\omega) = [\frac{1}{2} \sum_{j \in \mathbb{Z}} p_j e^{-i\omega j}] \widehat{\phi}(\omega)$$

Then sum $[\sum_{j \in \mathbb{Z}} p_j e^{-i\omega j}] = P(e^{-i\omega})$ is some power series in $e^{-i\omega}$ where

$$P(z) = \frac{1}{z} \sum_{j \in \mathbb{Z}} p_j z^j$$

It remains for us to find an appropriate polynomial. We know that $\widehat{\phi}(t)$ is the indicator function on $(-\pi, +\pi)$. A necessary condition on

$$\widehat{\phi}(2\omega) = P(e^{-i\omega}) \widehat{\phi}(\omega) \; (*)$$

is that for $|\omega| < \pi$,

$$\widehat{\phi}(2\omega) = \sum_{j \in \mathbb{Z}} [\frac{1}{2} p_j] e^{-i\omega j}$$

The function $\widehat{\phi}(2\omega)$ is the indicator function on $(-\pi/2, +\pi/2)$. How do we write this as a sum of complex exponential, well - we need a Fourier Series expansion of $\widehat{\phi}(2\omega)$. Hence, the Fourier Coefficients tell us

$$\frac{1}{2} p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\omega} \widehat{\phi}(2\omega) d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} e^{ik\omega} d\omega$$

$$= \begin{cases} \frac{1}{2}, \; k = 0 \\ \frac{(-1)^{(k+1)/2}}{k\pi}, \; k \text{ is odd} \\ 0, \text{ otherwise} \end{cases}$$

Replacing k with $2k + 1$ (since non-zero even terms disappear), we have that

$$P(z) = \frac{1}{2}[1 + \sum_{k \in \mathbb{Z}} \frac{(-1)^k}{\pi} \frac{2}{2k + 1} z^k]$$

Therefore, the Shannon sampling satisfies a Two-Scale relation (transforming this back into time domain)

$$\phi(t) = \phi(2t) + \sum_{k \in \mathbb{Z}} \frac{(-1)^k}{\pi} \frac{2}{2k + 1} \phi(2t - 2k - 1)$$

■

22 Lecture 22 - 3/17/2023

Let's recall the question we had last time -

> **Question 22.1.** Does the Shannon Sampling Function $\phi(s) = \frac{\sin \pi s}{\pi s}$ generate a MRA of $L^2(\mathbb{R})$?

We already know that

- $\{\phi(s - j)\}_{j \in \mathbb{Z}}$ are orthonormal - Riesz basis with $A = B = 1$

- ϕ satisfies a two-scale relation (shown last lecture)

- What happens when $V_n \to L^2(\mathbb{R})$?

- For all $f \in L^2(\mathbb{R})$, does there exist $f_n \in V_n$ such that

$$f_n \to f \text{ as } n \to \infty$$

In other words,
$$||f - f_n|| = 0 \text{ as } n \to \infty$$

> **Question 22.2.** How can we construct such f_n?

We want to use Parserval's Identity to construct on $\widehat{f}_n$ instead, then we know that

$$||f - f_n|| = \frac{1}{2\pi} ||\widehat{f} - \widehat{f}_n||$$

Hence, it suffices for us to construct a sequence of a functions in the frequency domain, this will create a series of functions in the time domain that approximates f equally well.

> **Theorem 22.3.** The Shannon Sampling Function $\phi(s) = \frac{\sin \pi s}{\pi s}$ generates a MRA of $L^2(\mathbb{R})$.

Proof. Let's choose $\widehat{f}_n(\omega)$ as follows:

$$\widehat{f}_n(\omega) = \begin{cases} \widehat{f}(\omega), & |\omega| < \Omega_n \\ 0, & \text{otherwise} \end{cases}$$

We postpone the construction of Ω_n and will choose them appropriately later on in this proof. We intuitively want to choose Ω_n such that $\Omega_n \to \infty$ as $n \to \infty$, then

$$||\widehat{f} - \widehat{f}_n||^2 = \int_{|\omega| > \Omega_n} |\widehat{f}(\omega)|^2 d\omega \to 0 \text{ as } \Omega_n \to \infty$$

Then, by the Shannon Sampling Theorem, we have that

$$f_n(t) = \sum_{k \in \mathbb{Z}} f_n(ka) \phi\left(\frac{t}{a} - k\right) \ (\dagger)$$

provided that $a = \frac{\pi}{\Omega_n}$. Our goal now is to choose Ω_n such that $f_n \in V_n$, ie. we want

$$f_n \in \text{span}\{\phi(2^n t - k) \ : \ k \in \mathbb{Z}\}$$

Indeed, if we choose $\Omega_n = 2^n \pi$, then $a = \frac{\pi}{\Omega_n} = 2^{-n}$, substituting back into $(\dagger)$ gives

$$f_n(t) = \sum_{k \in \mathbb{Z}} f_n(ka) \phi(2^n t - k)$$

Thus, we have that $f_n \in V_n$. $\blacksquare$

Question 22.4. How do we use Multi-Resolution Analysis in practice?

$$\{0\}... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset ... L^2(\mathbb{R})$$

Note that as we go to $V_{-\infty}$, we only have the constant zero function left. In principle we can pick any $f \in L^2(\mathbb{R})$ and compute the $\{f_n \in V_n\}$ sequence of approximation of f for signal analysis, but it seems somewhat wasteful since we can retrieve some information about $f_{n-1} \in V_{n-1}$ just based on $f_n \in V_n$, so there's inefficiency in storage and duplication of information.

Now consider $f_n \in V_n \subset V_{n+1} \ni f_{n+1}$. We can write $f_{n+1} = f_n + (f_{n+1} - f_n)$, ie. we work with the sequence of differences instead $\{f_{n+1} - f_n\}_{n \in \mathbb{Z}}$, observe that

- As $f_n \to f$, $f_{n+1} - f_n \to 0$, so the difference is getting smaller.

- This leads us to suspect tht this sequence is simpler than saving $\{f_n\}$.

- Ex. We don't lost any information, if we have f_0, and we want f_4, then

$$f_4 = (f_4 - f_3) + (f_3 - f_2) + (f_2 - f_1) + (f_1 - f_0) + f_0$$

Question 22.5. Can we quantify in what sense $f_{n+1} - f_n$ is "simpler"?

Example 22.6. Let $\phi(t)$ be the inidcator function on $[0, 1]$, then we still have that $f_{n+1} - f_n \in V_n$ so it doesn't seem simpler, but let's try to investigate this in detail. What does $f_{n+1} - f_n$ look like? Well, let's first examine f_n and f_{n+1}:
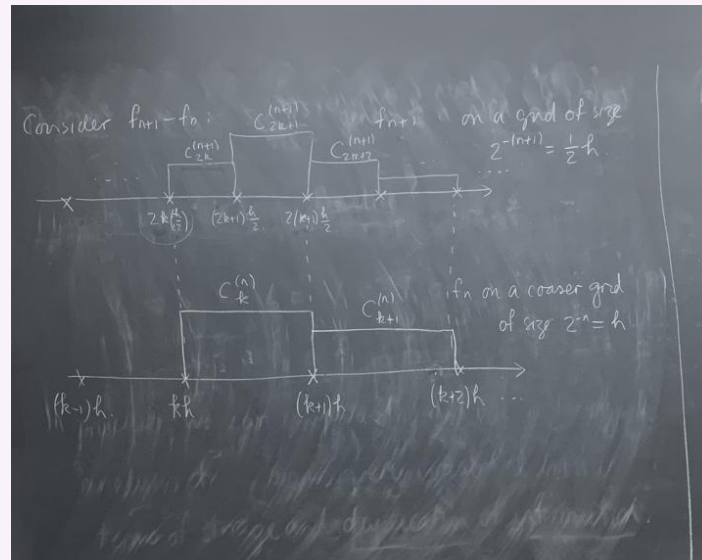
Suppose we are given f_{n+1} and we want to downgrade to f_n. Let's choose

$$c_k^{(n)} = \frac{1}{2}[c_{2k}^{(n+1)} + c_{2k+1}^{(n+1)}] \text{ (Best Choice)}$$

Now, for $kh \leq t \leq (k+1)h$, we have that

$$f_{n+1}(t) = c_{2k}^{(n+1)}\phi(\frac{2t}{h} - 2k) + c_{2k+1}^{(n+1)}\phi(\frac{2t}{h} - 2k - 1)$$

$$f_n(t) = c_k^{(n)}\phi(\frac{t}{h} - k)$$

Using the 2-Scale Relation, we can write

$$\phi(\frac{t}{h} - k) = \phi(\frac{2t}{h} - 2k) + \phi(\frac{2t}{h} - 2k - 1)$$

Hence we write

$$f_n(t) = c_k^{(n)}[\phi(\frac{2t}{h} - 2k) + \phi(\frac{2t}{h} - 2k - 1)]$$

Hence the diffeence is

$$f_{n+1}(t) - f_n(t) = d_L\phi(\frac{2t}{h} - 2k) + d_R\phi(\frac{2t}{k} - 2k - 1) \qquad d_L = c_{2k}^{(n+1)} - c_k^{(n)} \text{ and } d_R = c_{2k+1}^{(n+1)} - c_k^{(n)}$$

We observe, however, that $-d_L = d_R$. Hence, we have that

$$f_{n+1}(t) - f_n(t) = d_k^{(n)}[\phi(\frac{2t}{h} - 2k) - \phi(\frac{2t}{h} - 2k - 1)]$$

where $d_k^{(n)} = d_L$.

Let's define

$$\psi(s) = \phi(2s) - \phi(2s - 1)$$

Then we have that

$$\psi(\frac{t}{h} - k) = \phi(\frac{2t}{h} - 2k) - \phi(\frac{2t}{h} - 2k - 1)$$

Hence we have that

$$f_{n+1}(t) - f_n(t) = d_k^{(n)}\psi(\frac{t}{h} - k), \ kh \le t \le (k+1)h$$

The same arguments gives that, for a general t, for $h = 2^{-n}$
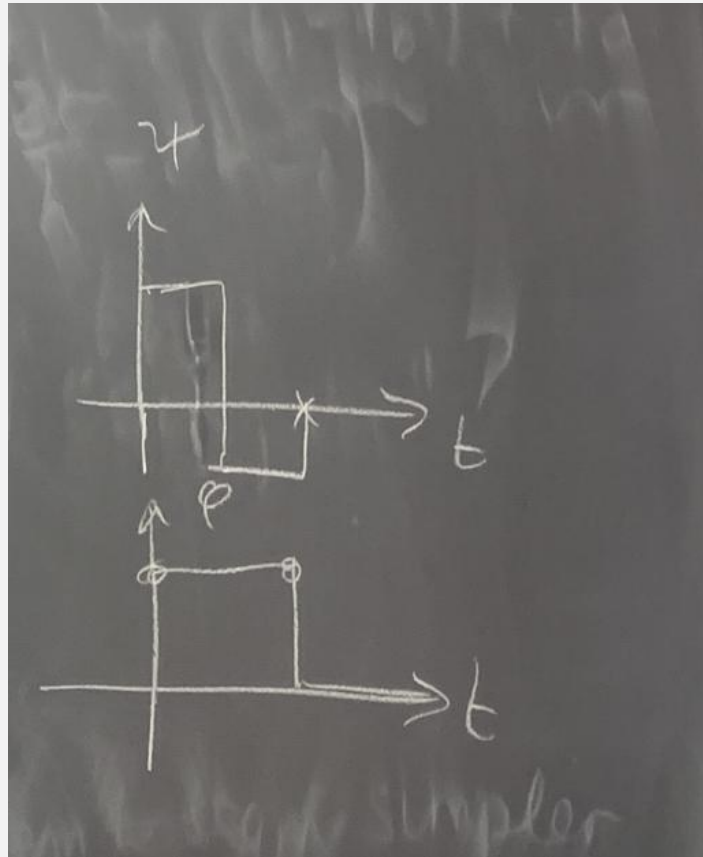
$$f_{n+1}(t) - f_n(t) = \sum_{k \in \mathbb{Z}} d_k^{(n)}\psi(\frac{t}{h} - k) = \sum_{k \in \mathbb{Z}} d_k^{(n)}\psi(2^n t - k)$$

Hence we have that

$$f_{n+1} - f_n \in \text{span}\{\psi(2^n t - k) \ : \ k \in \mathbb{Z}\}$$

Why don't we call $\text{span}\{\psi(2^n t - k) \ : \ k \in \mathbb{Z}\}$ as W_n? Hence, we have that $f_{n+1} - f_n \in W_n$ for all n.

Remark 22.7. As a final remark, what does the function $\psi(s)$ look like:

This is the Haar Wavelet again!

Remark 22.8. Why do you think we are calling these spaces W_n? Wavelets!