

No Pedagogy: The Fourier Transform

Leo B. PeBenito

November 26, 2023

1 Preface

I do not like punctuation in-line with formulas and math expressions when they are given the importance of their own line, this is not standard, periods and commas often appear.

No Pedagogy means do everything! show everything! No claims about what I know, just telling what I think and trying to be as upfront about where my knowledge ends.

By proceeding naively you are incurring more technical debt than you may realize and only be able to implement the Fourier transform once in a specific setting but ultimately have to start over when applying the Fourier transform in a new context.

What is up with Finite Element Methods (Gilbert Strang's first book is on this) and the Fourier transform.

This document is meant to ... you don't have to sit through the sound of a chalk board and try to take notes from illegible penmanship (Osgood lectures). Math equations should never be interrupted by a figure (p.24[5]). I personally hate when periods to sentences terminate equations. I hate how Briggs and Henson reverse the order of frequency and position in the FT when referring to indices in the DFT, ie:

$$\sum e^{-i2\pi\omega_k x_n} = \sum e^{-i2\pi kn/N}$$

Prioritize consistency over aesthetics, ie: writing the Fourier transform as

$$F_k = \int f(x) e^{-i2\pi\omega x} dx$$

and writing the DFT as

$$F_k = \sum_n f_n e^{-i2\pi nk/N}$$

swapping the order of the variables ω and x when referring to the respective indices for frequency and position in the discrete Fourier transform, k and n .

Will attempt to keep things "self-contained", will not extend things over multiple chapters, giving too little information too soon only to reveal an anticlimactic punchline.

This repo will live and die based on the value of its documentation - that is to say, I am no expert! My goal is that someone with the patience to read the documentation will be able to proceed with the code with confidence, and maybe even be curious enough to look at a reference.

No pedagogy: Nothing is being intentionally withheld. The order in which topics are introduced should be logical enough to make transitions seem obvious. Concepts are built up incrementally without skipping steps (if something is missing it's because I don't understand it, yet!) - let the reader skip ahead if she is impatient to see the significant results.

The lead of to one of the most important formulas in the book (eqn. 6.6, p.191 [5]) is “a short calculation (problem 125) reveals that the Fourier coefficients of f on $[-pA/2, pA/2]$ in the Appendix shows that the pN -point DFT is given by” This sentence lacks clarity and defers the derivation/explanation to the reader (as a problem).

2 Fourier Series

Must make executive decision: must get away from Osgood notation. Must get away from using k to index spatial modes. This leads to conflict with other notation which uses k to index frequency, especially when introducing the DFT!... This Tag is placed everywhere where the notation has not been updated to use n in place of k But before you go down this path... think, k is being used to index frequency/wavelength here. Later n will be used to index space/time in the DFT.

The derivation of the Fourier transform begins with Fourier series. Dirichlet's Theorem[1] stated by Bakefi and Barret: Any periodic, single valued function $f(x)$ defined in the interval $[-\pi, \pi]$ and having a periodicty of 2π can be represented by the trigonometric series:

$$f(x) = \frac{A_0}{2} + \sum_{k=1}^{\infty} (A_k \cos(kx) + B_k \sin(kx)) \quad (1)$$

where x is dimensionless, and $k \in \mathbb{Z}$.

Should I change k to n ?... k is used lated to denote something completely different.

Fourier analysis is concerned with finding the coefficients associated with the terms in the series. Consider the first term $\frac{A_0}{2}$, which amounts to a vertical shift, and corresponds to the average value of the function being represented by the Fourier series. To obtain A_0 , and to understand the reason of the factor of $\frac{1}{2}$, integrate the function $f(x)$

$$\int_{-\pi}^{\pi} f(x) dx$$

All the sine and cosine terms will cancel because they are integrated over the full domain of their periodicity, leaving an integral over a complete period of A_0 , which produces a factor of 2π

$$= \int_{-\pi}^{\pi} \frac{A_0}{2} dx = \frac{A_0}{2} 2\pi = A_0\pi$$

Fourier's trick, as it's referred to, amounts to integrating $f(x)$ and dividing by π to get the Fourier coefficient, here A_0

$$A_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx$$

This procedural structure is maintained when obtaining the amplitudes for the sine and cosine terms. For the sine terms first.

$$\int_{-\pi}^{\pi} f(x) \sin(kx) dx = \int_{-\pi}^{\pi} \sin^2(kx) dx$$

Note:

$$\int_{-\pi}^{\pi} \sin(kx) \sin(mx) dx = \delta_{km}$$

Recall:

$$\begin{aligned} e^{i\theta} e^{i\theta} &= e^{i2\theta} = \cos(2\theta) + i \sin(2\theta) \\ &= (\cos(\theta) + i \sin(\theta)) (\cos(\theta) + i \sin(\theta)) = \cos^2(\theta) - \sin^2(\theta) + i2 \cos(\theta) \sin(\theta) \end{aligned}$$

Combining real parts, and using $1 = \sin^2(\theta) + \cos^2(\theta)$

$$\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta) = \cos^2(\theta) - (1 - \cos^2(\theta))$$

So,

$$\cos^2(\theta) = \left(\frac{1 + \cos(2\theta)}{2} \right)$$

By the same token,

$$\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta) = (1 - \sin^2(\theta)) - \sin^2(\theta)$$

$$\sin^2(\theta) = \left(\frac{1 - \cos(2\theta)}{2} \right)$$

Now apply this to Fourier's trick

$$\int_{-\pi}^{\pi} dx f(x) \sin(kx) = \int_{-\pi}^{\pi} dx \sin^2(kx) = \int_{-\pi}^{\pi} dx \left(\frac{1 - \cos(2kx)}{2} \right)$$

The cos term vanishes, being integrated over its domain of periodicity. What remains is

$$\frac{1}{2} \int_{-\pi}^{\pi} dx = \frac{2\pi}{2} = \pi$$

In the same manner,

$$\int_{-\pi}^{\pi} dx \cos^2(kx) = \int_{-\pi}^{\pi} dx \left(\frac{1 + \cos(2kx)}{2} \right)$$

The cos term vanishes, being integrated over its domain of periodicity, leaving

$$\frac{1}{2} \int_{-\pi}^{\pi} dx = \frac{2\pi}{2} = \pi$$

SIDE: Perhaps it is worth noting the factor of $\frac{1}{2}$ here as it helps remember the factor of $\frac{1}{2}$ of B_0 , but in this case the $\frac{1}{2}$ comes from the definition of sin and cos .

The amplitudes of the individual terms can be gotten by

$$A_m = \frac{1}{\pi} \int_{-\pi}^{\pi} dx f(x) \cos(mx) \quad B_m = \frac{1}{\pi} \int_{-\pi}^{\pi} dx f(x) \sin(mx)$$

2.1 Fourier Series Using Complex Exponentials

It is preferable to express the infinite series in terms of complex exponentials rather than sines and cosines, given

$$\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2} \quad \sin(\theta) = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

Then $f(x)$ becomes,

$$\begin{aligned} f(x) &= \frac{A_0}{2} + \sum_{k=1}^{\infty} \left(A_k \frac{e^{ikx} + e^{-ikx}}{2} + B_k \frac{e^{ikx} - e^{-ikx}}{2i} \right) \\ &= \frac{A_0}{2} + \sum_{k=1}^{\infty} \left[\left(\frac{A_k}{2} + \frac{B_k}{2i} \right) e^{ikx} + \left(\frac{A_k}{2} - \frac{B_k}{2i} \right) e^{-ikx} \right] \end{aligned}$$

Simplify the coefficients, given $i^2 = -1$, and $-i^2 = 1$, by writing

$$\frac{1}{i} = \frac{-i^2}{i} = -i \quad , \text{ and } \quad \frac{-1}{i} = \frac{i^2}{i} = i$$

$$= \frac{A_0}{2} + \frac{1}{2} \sum_{k=1}^{\infty} \left(A_k - iB_k \right) e^{ikx} + \frac{1}{2} \sum_{k=1}^{\infty} \left(A_k + iB_k \right) e^{-ikx}$$

Inverting the index of the latter sum to go from -1 to $-\infty$

$$= \frac{A_0}{2} + \frac{1}{2} \sum_{k=1}^{\infty} \left(A_k - iB_k \right) e^{ikx} + \frac{1}{2} \sum_{k=-1}^{-\infty} \left(A_{-k} + iB_{-k} \right) e^{ikx}$$

allows the two sums to be combined into a single sum over their combined indices. Written in this way it is easy to see the familiar form of the Fourier series written as a sum from $-\infty$ to ∞ of complex exponentials

$$= \sum_{k=-\infty}^{\infty} C_k e^{ikx} \quad , \text{ where } \quad \begin{cases} C_0 = A_0/2 & , \text{ if } k = 0 \\ C_k = \frac{1}{2}(A_k - iB_k) & , \text{ if } k > 0 \\ C_k = \frac{1}{2}(A_{-k} + iB_{-k}) & , \text{ if } k < 0 \end{cases}$$

Clearly, $C_k = \overline{C_{-k}}$ (where the bar notation indicates complex conjugate). While this formula is more compact, **keep in mind both positive and negative indices, k and $-k$ are needed to describe a given (the k -th) mode.**

2.2 Fourier Coefficients (scaling the Fourier series)

It is natural to consider the interval $[-\pi, \pi]$ from a mathematical perspective. A shift in the domain to $[0, 2\pi]$ has no affect on the above treatment. From a physical perspective it is more natural to consider the interval from 0 to λ (the implicit wavelenth associated with the spatial extent), or from 0 to L , the length of time sampled. The variable must then be scaled by $2\pi/\lambda$ such that

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kx/\lambda}$$

Applying Fourier's trick now involves an integral over the wavelength

$$\begin{aligned} \int_0^\lambda dx f(x) \exp\left(-\frac{i2\pi mx}{\lambda}\right) &= \int_0^\lambda dx \sum_{k=-\infty}^{\infty} c_k \exp\left(\frac{i2\pi kx}{\lambda}\right) \exp\left(-\frac{i2\pi mx}{\lambda}\right) \\ &= \int_0^\lambda dx c_k \exp\left(\frac{i2\pi(k-m)x}{\lambda}\right) \\ &= \begin{cases} 0 & \text{if } k \neq m \\ \int_0^\lambda c_k dx = c_k(\lambda) & \text{if } k = m \end{cases} \end{aligned}$$

For $k \neq m$ the above equation equals 0. For $k = m$ the above equation is just an integral over the wavelength. The Fourier coefficient for each index k is obtained by

$$c_k = \frac{1}{\lambda} \int_0^\lambda dx f(x) \exp\left(-\frac{i2\pi kx}{\lambda}\right)$$

We can just as easily consider the symmetric interval over the wavelength

$$c_k = \frac{1}{\lambda} \int_{-\lambda/2}^{\lambda/2} dx f(x) \exp\left(-\frac{i2\pi kx}{\lambda}\right)$$

More generally, simply in terms of some length (L), ie: of the sample duration

$$c_k = \frac{1}{L} \int_L dx f(x) \exp\left(-\frac{i2\pi kx}{L}\right)$$

3 Derivation of the Fourier Transform

The derivation of the Fourier transform is based on extending the boundaries over which the Fourier series is defined, and thereby the boundaries of the Fourier analysis integral, to $\pm\infty$. The trick to performing this is to consider the interval that separates the modes, ie: define

$$\omega = \frac{2\pi k}{\lambda} \quad \longrightarrow \quad \Delta\omega = \frac{2\pi}{\lambda}$$

$$c_k = \frac{\Delta\omega}{2\pi} \int_{-\pi/\Delta\omega}^{\pi/\Delta\omega} dx f(x) \exp(-i\omega x)$$

The Fourier series expansion can now be written as

$$f(x) = \sum_{\omega=-\infty}^{\infty} \left(\frac{\Delta\omega}{2\pi} \int_{-\pi/\Delta\omega}^{\pi/\Delta\omega} dx f(x) \exp(-i\omega x) \right) \exp(i\omega x)$$

Taking the limit

$$= \lim_{\substack{\Delta\omega \rightarrow 0 \\ (\lambda \rightarrow \infty)}} \sum_{\omega=-\infty}^{\infty} \left(\frac{\Delta\omega}{2\pi} \int_{-\pi/\Delta\omega}^{\pi/\Delta\omega} dx f(x) \exp(-i\omega x) \right) \exp(i\omega x)$$

At this stage the factor of $1/2\pi$ can be distributed

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp(-i\omega x) dx \right)}_{F(\omega)} \exp(i\omega x) d\omega$$

The Fourier transform of $f(x)$ and inverse Fourier transform are

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp(-i\omega x) dx \quad (2a)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega x) d\omega \quad (2b)$$

It is also possible to not distribute the factor of $1/2\pi$, such that sometimes the Fourier transform and inverse Fourier transform are written as

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) \exp(-i\omega x) dx \quad (3a)$$

$$f(x) = \int_{-\infty}^{\infty} F(\omega) \exp(i\omega x) d\omega \quad (3b)$$

A way to remember which expression gets the minus sign in the exponential is based on which expression is intended to model the original function (synthesis) and which is intended to obtain the Fourier coefficients (analysis). The expression for getting the Fourier coefficients has the minus sign in the exponential. That is, the way to get the Fourier coefficients involves multiplying by the complex conjugate.

The treatment by Brad Osgood (Jerri also puts 2π in the exponential) in terms of time t is nice too, and **the factor of 2π stays in the exponential**. It simplifies construction of the Fourier series by considering the signal over a period of length 1, such that

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kt}$$

The Fourier coefficients are got somewhat more directly without the need for any prefactor as

$$c_k = \int_0^1 f(t) e^{-i2\pi kt} dt$$

To get to the Fourier transform consider a period T that encompasses the signal.

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kt/T}$$

$$c_k = \frac{1}{T} \int_0^T f(t) e^{-i2\pi kt/T} dt$$

The latter can be rewritten over the symmetric interval

$$c_k = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-i2\pi kt/T} dt$$

Similarly as before, define

$$s = \frac{k}{T} \rightarrow \Delta s = \frac{1}{T}$$

Reasoning similarly as before

$$f(t) = \lim_{\substack{\Delta s \rightarrow 0 \\ (T \rightarrow \infty)}} \sum_{s=-\infty}^{\infty} \left(\Delta s \int_{-1/\Delta s}^{1/\Delta s} dt f(t) e^{-i2\pi st} \right) e^{i2\pi st}$$

$$= \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dt f(t) e^{-i2\pi st} e^{i2\pi st} ds}_{F(s)}$$

The Fourier transform analyzes a signal into its component parts

$$(\mathcal{F}f)(s) = \mathcal{F}f = F(s) = \hat{f}(s) = \int_{-\infty}^{\infty} dt f(t) e^{-i2\pi st}$$

and is often thought of in the frequency domain. In general it is necessary to understand the convergence of the integral. The set of $s \in \mathbb{R}$ for which $\mathcal{F}(f(t))$ is defined is called the spectrum. The inverse Fourier transform synthesizes the function from its constituent parts

$$\mathcal{F}^{-1}F = f(t) = \int_{-\infty}^{\infty} ds F(s) e^{i2\pi st}$$

and is often thought of in the time domain. However, the Fourier transform is a sufficiently flexible tool that it is often advantageous to abandon such references.

Notice the relationship between the Fourier transform and the Fourier coefficients using scaled Fourier series for f defined over the region L

$$\frac{F(s_k)}{L} = \frac{\hat{f}(s_k)}{L} = c_k = \frac{1}{L} \int_L dx f(x) e^{-i2\pi s_k x}$$

This relationship is a little subtle (can be confusing, ie: if you thought the Fourier transform gave the Fourier coefficients directly, given that the Fourier coefficients furnished the starting point and a fairly direct route for the derivation of the Fourier transform; but the relationship is clear just by comparing the formulas directly), and reoccurs with the discrete Fourier transform (DFT) and essentially shows that the DFT is actually a direct approximation to the Fourier coefficients (despite the nomenclature).

3.1 The Many Defintions of the Fourier Transform

For time, (t) , in seconds and frequency, (f) , in cycles per second (Hertz) (p. 601, ref. Numerical Methods) the Fourier transform equations are

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{i2\pi ft} dt$$

$$h(t) = \int_{-\infty}^{\infty} H(f) e^{-i2\pi ft} df$$

Using angular frequency, $\omega = 2\pi f$, in radians per second [29]

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

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

But this form is not preferred because of the need to keep track of the factor of $(2\pi)^{-1}$.

3.2 Basic Properties of the Fourier Transform

The Fourier transform is a linear operator.

Prove these relationships between the function and its Fourier transform. These symmetries can be used to increase computational efficiency[29].

$h(t)$ real	$H(-f) = H(f) ^*$
$h(t)$ imaginary	$H(-f) = - H(f) ^*$
$h(t)$ even	$H(f)$ even
$h(t)$ odd	$H(f)$ odd
$h(t)$ real even	$H(f)$ real even
$h(t)$ real odd	$H(f)$ imaginary odd
$h(t)$ imaginary even	$H(f)$ imaginary even
$h(t)$ imaginary odd	$H(f)$ real odd

3.3 Fourier Inversion and the Dirac δ

Recall: the integral representation of the Dirac δ -function using complex exponentials

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\pm i k x} dx$$

Fourier inversion refers to the process by which the inverse Fourier transform of the Fourier transform, and the Fourier transform of the inverse Fourier transform both return the function

$$\begin{aligned} \mathcal{F}^{-1}(\mathcal{F}(f(t))) &= f(t) = \mathcal{F}(\mathcal{F}^{-1}(f(t))) \\ &= \int_{-\infty}^{\infty} e^{-i2\pi st} \int_{-\infty}^{\infty} e^{+i2\pi st'} f(t') dt' ds \\ &= \int_{-\infty}^{\infty} dt' f(t') \int_{-\infty}^{\infty} ds e^{+i2\pi s(t'-t)} \end{aligned}$$

The right integral expression is a complex exponential form (/integral representation) of the **Dirac δ -function** (which is of course symmetric and the argument variables can swap position).

$$= \int_{-\infty}^{\infty} dt' f(t') \delta(t' - t) = f(t)$$

4 Properties of the Fourier Transform

4.1 Duality

$$\boxed{\mathcal{F}(f^-) = \mathcal{F}(f(-t)) = (\mathcal{F}f)^- = (\mathcal{F}f)(-s) = \mathcal{F}^{-1}f}$$

The *reverse* signal defined as

$$f^{reversed} = f^- == f^{(-)} = f(-t)$$

provides a way to express if a function is even or odd

$$\begin{aligned} f &= f^{(-)} && \text{“even”} \\ -f &= f^{(-)} && \text{“odd”} \end{aligned}$$

In practice this amounts to processing the signal in reverse.

The reverse of the Fourier transform is the inverse Fourier transform

$$(\mathcal{F}f)^- = F(-s) = \int_{-\infty}^{\infty} f(t)e^{-i2\pi(-s)t} dt = \int_{-\infty}^{\infty} f(t)e^{+i2\pi st} dt = \mathcal{F}^{-1}f$$

, which is why \mathcal{F} , and \mathcal{F}^{-1} , are sometimes referred to as the forward, and reverse Fourier transform, respectively.

Associating the minus sign introduced by $(-s)$ into the argument of the exponential with (t)

$$= \int_{-\infty}^{\infty} f(t)e^{-i2\pi s(-t)} dt$$

and performing the u -substitution, $\begin{cases} u &= -t \\ du &= -dt \end{cases}$

$$\begin{aligned} &= - \int_{\infty}^{-\infty} f(-u)e^{-i2\pi su} du = \int_{-\infty}^{\infty} f(-u)e^{-i2\pi su} du \\ &= \mathcal{F}(f(-u)) = \mathcal{F}(f(-t)) \quad (\text{since } u \text{ is just a dummy variable}) \end{aligned}$$

shows the Fourier transform of the reverse signal is equal to the inverse Fourier transform.

The Fourier transform of the Fourier transform returns the reverse signal

$$\boxed{\mathcal{F}\mathcal{F}f = f^-}$$

$$\begin{aligned} \mathcal{F}(\mathcal{F}(f(t))) &= \int_{-\infty}^{\infty} e^{-2\pi ist} dt \int_{-\infty}^{\infty} e^{-2\pi ist'} f(t') dt' \\ &= \int_{-\infty}^{\infty} f(t') dt' \int_{-\infty}^{\infty} e^{+2\pi is(-t-t')} dt \\ &= \int_{-\infty}^{\infty} f(t') \delta(-t-t') dt' = f(-t) \end{aligned}$$

This makes it simple to solve some otherwise difficult Fourier transforms.

4.2 The Shift (Delay) Theorem:

Do this for frequency shift as well

$$h(t)e^{-i2\pi f_0 t} \Longleftrightarrow H(f - f_0)$$

$$\boxed{f(t \pm b) \xleftrightarrow[\mathcal{F}^{-1}]{\mathcal{F}} e^{\pm i2\pi sb} F(s)}$$

The Fourier transform of a delayed signal is

$$\mathcal{F}(f(t-b)) = \int_{-\infty}^{\infty} e^{-i2\pi st} f(t-b) dt$$

Perform u -substitution, $u = t - b$, $du = dt$, $t = u + b$

$$\begin{aligned} &= \int_{-\infty}^{\infty} e^{-i2\pi s(u+b)} f(u) du \\ &= e^{-2\pi isb} \int_{-\infty}^{\infty} e^{-i2\pi su} f(u) du \\ &= e^{-i2\pi sb} F(s) \end{aligned}$$

The physical interpretation of this is a shift in time corresponds to a phase shift in frequency. Recall, the Fourier transform is a complex number with a magnitude and a phase $F(s) = |F(s)|e^{i2\pi\Theta(s)}$, so $e^{-2\pi isb} F(s) = |F(s)|e^{i2\pi(\Theta(s)-sb)}$.

4.3 The Scaling (Stretch/Similarity) Theorem

Do this for frequency as well

$$\frac{1}{|b|} h\left(\frac{t}{b}\right) \iff H(bf)$$

$$\boxed{f(\alpha t) \xleftrightarrow[\mathcal{F}^{-1}]{\mathcal{F}} \frac{1}{|\alpha|} F\left(\frac{s}{\alpha}\right)}$$

The Fourier transform of the scaled signal is

$$\mathcal{F}(f(\alpha t)) = \int_{-\infty}^{\infty} e^{-i2\pi st} f(\alpha t) dt$$

For $\alpha > 0$, the u -substitution, $u = \alpha t$, gives

$$= \int_{-\infty}^{\infty} \exp\left(-i2\pi s\left(\frac{u}{\alpha}\right)\right) f(u) \frac{du}{\alpha} = \frac{1}{\alpha} \int_{-\infty}^{\infty} \exp\left(-i2\pi\left(\frac{s}{\alpha}\right)u\right) f(u) du$$

For $\alpha < 0$, this u -substitution effectively swaps the limits of integration

$$= \frac{1}{\alpha} \int_{+\infty}^{-\infty} e^{-i2\pi(s/\alpha)u} f(u) du$$

To recover the original limits of integration introduces a minus sign. But α is negative in this case, so

$$= \frac{-1}{\alpha} \int_{-\infty}^{\infty} e^{-i2\pi(s/\alpha)u} f(u) du = \frac{1}{|\alpha|} F(s/\alpha)$$

The physical interpretation of this is, if $|\alpha| > 1$ then f gets squeezed, and $|F(s)|$ gets stretched horizontally and squashed vertically. Whereas, if $|\alpha| < 1$ then f gets stretched, and $|F(s)|$ gets squeezed horizontally and stretched vertically. This reciprocal relationship shows that *a signal cannot be simultaneously localized in time and in frequency*.

DO THIS: What happens when you combine shifts and stretches?

4.4 The Convolution Theorem

$\mathcal{F}(f) \cdot \mathcal{F}(g) = \mathcal{F}(f * g)$
$\mathcal{F}(f) * \mathcal{F}(g) = \mathcal{F}(f \cdot g)$
$\mathcal{F}^{-1}(f) \cdot \mathcal{F}^{-1}(g) = \mathcal{F}^{-1}(f * g)$
$\mathcal{F}^{-1}(f) * \mathcal{F}^{-1}(g) = \mathcal{F}^{-1}(f \cdot g)$

Table 1: Summary of important results. The main take away is, colloquially, the Fourier transform (and inverse Fourier transform) “swaps convolution and multiplication”.

The linearity of the Fourier transform

$$\begin{aligned}\mathcal{F}(f + g) &= \mathcal{F}(f) + \mathcal{F}(g) \\ \mathcal{F}(\alpha f) &= \alpha \mathcal{F}(f)\end{aligned}$$

shows how addition and scalar multiplication of signals influences the Fourier transform. The product of two signals has a more dramatic impact on the Fourier transform.

The best known and most usefull expression involving multiplication of signals and the Fourier transform is given by the Fourier transform of the convolution

$F(s)G(s) = \mathcal{F}(f * g)$

which is the product of the Fourier transforms. This expression describes application of a filter according to frequency.

Proof:

Differentiate t variables of the Fourier transforms

$$\begin{aligned}(F(s))(G(s)) &= \mathcal{F}(f)\mathcal{F}(g) \\ &= \left(\int_{-\infty}^{\infty} e^{-2\pi i s t_1} f(t_1) dt_1 \right) \left(\int_{-\infty}^{\infty} e^{-2\pi i s t_2} g(t_2) dx_2 \right)\end{aligned}$$

There are two *decoupled* variables so the integrals can be combined and the components rearranged

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i s (t_1 + t_2)} f(t_1)g(t_2) dt_1 dt_2$$

Introducing a change of variables $t = t_1 + t_2$, and $t_2 = t - t_1$

$$\det(J) = \left| \frac{\partial(t_1, t_2)}{\partial(t_1, t)} \right| = \left| \begin{array}{cc} \frac{\partial t_1}{\partial t_1} & \frac{\partial t_1}{\partial t} \\ \frac{\partial t_2}{\partial t_1} & \frac{\partial t_2}{\partial t} \end{array} \right| = \left| \begin{array}{cc} 1 & 0 \\ -1 & 1 \end{array} \right| = 1$$

The Jacobian of the transformation is 1, and the exponential is separable, such that the integral can be rewritten as

$$= \int_{-\infty}^{\infty} e^{-2\pi i s t} \underbrace{\left(\int_{-\infty}^{\infty} f(t_1) g(t - t_1) dt_1 \right)}_{(f * g)} dt$$

The integral in parentheses is the convolution of f and g , $(f * g)$

$$= \int_{-\infty}^{\infty} e^{-2\pi i s t} (f * g) dt = \mathcal{F}(f * g)$$

(Note the limits of integration range from $(-\infty, \infty)$ unlike the convolution as written in its familiar form in connection with the Laplace transform, which has limits of integration from 0 to t .)

As a means of self check, given the result

$$F(s) G(s) = \mathcal{F}(f * g)$$

the signal can be recovered by Fourier inversion

$$\boxed{\mathcal{F}^{-1}((F(s))(G(s))) = (f * g)}$$

as follows

$$\mathcal{F}^{-1}((F(s))(G(s))) = \mathcal{F}^{-1} \mathcal{F}(f * g) = (f * g)$$

To demonstrate this, write the convolution as an explicit function of t

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

$$\mathcal{F}^{-1} \mathcal{F}(f * g) = \int_{-\infty}^{\infty} e^{i2\pi s t} \int_{-\infty}^{\infty} e^{-i2\pi s t} (f * g) ds dt$$

Distinguish the t -variable of the Fourier transform as t' (time in one expression need not be correlated with time in another expression)

$$\begin{aligned} &= \int_{-\infty}^{\infty} (f * g) \int_{-\infty}^{\infty} e^{i2\pi s(t-t')} ds dt' \\ &= \int_{-\infty}^{\infty} (f * g) \delta(t - t') dt' = (f * g)(t) \end{aligned}$$

The convolution of Fourier transforms is

$$\boxed{\mathcal{F}(f) * \mathcal{F}(g) = \mathcal{F}(f \cdot g)}$$

To see this write Fourier transforms in terms of the (s) variables

$$\mathcal{F}(f) * \mathcal{F}(g) = F(s) * G(s)$$

which by definition is

$$\begin{aligned} &= \int_{-\infty}^{\infty} F(s') G(s - s') ds' \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-i2\pi s' t} f(t) dt \right) \left(\int_{-\infty}^{\infty} e^{-i2\pi(s-s')t} g(t) dt \right) ds' \end{aligned}$$

Use Fubini to rearrange the order of integration as needed. Use one integration over (t) to define the delta δ -function

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i2\pi s' t} f(t) g(t) \delta(s - s') ds' dt$$

Use integration over (s') to apply the delta function. What remains is the Fourier transform

$$= \int_{-\infty}^{\infty} e^{-i2\pi s t} f(t) g(t) dt = \mathcal{F}(f \cdot g)$$

Colloquially, the Fourier transform (as well as the inverse Fourier transform) swap multiplication for convolution.

A similar set of statements holds for the inverse Fourier transform. The inverse Fourier transform of the convolution

$$\boxed{\mathcal{F}^{-1}(f) \cdot \mathcal{F}^{-1}(g) = \mathcal{F}^{-1}(f * g)}$$

is the product of the inverse Fourier transforms.

$$\begin{aligned} \mathcal{F}^{-1}(f) \cdot \mathcal{F}^{-1}(g) &= \int_{-\infty}^{\infty} e^{+i2\pi s_1 t} f(s_1) ds_1 \int_{-\infty}^{\infty} e^{+i2\pi s_2 t} g(s_2) ds_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{+i2\pi(s_1+s_2)t} f(s_1) g(s_2) ds_1 ds_2 \end{aligned}$$

Use the substitution $s = s_1 + s_2$, again the determinant of the Jacobian is 1

$$= \int_{-\infty}^{\infty} e^{+i2\pi s t} \left(\int_{-\infty}^{\infty} f(s_1) g(s - s_1) ds_1 \right) ds = \mathcal{F}^{-1}(f * g)$$

The *convolution of the inverse Fourier transform*

$$\boxed{\mathcal{F}^{-1}(f) * \mathcal{F}^{-1}(g) = \mathcal{F}^{-1}(f \cdot g)}$$

is the inverse Fourier transform of the product.

$$\begin{aligned}
\mathcal{F}^{-1}(f) * \mathcal{F}^{-1}(g) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{i2\pi st'} f(s) ds \right) \left(\int_{-\infty}^{\infty} e^{i2\pi s(t-t')} g(s) ds \right) dt' \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i2\pi st'} f(s) g(s) \delta(t - t') ds dt' \\
&= \int_{-\infty}^{\infty} e^{i2\pi st} f(s) g(s) ds = \mathcal{F}^{-1}(f \cdot g)
\end{aligned}$$

4.5 Filtering

It is clear from the convolution theorem that multiplying the spectrum by a filter function, $H(s) = \mathcal{F}(h)$, in the frequency domain corresponds to convolution of $h(t)$ with the signal in the time domain.

A filter is a system that convolves an input (or set of inputs) with a fixed function/signal (known as the "impulse response").

In the time domain

$$\underset{\text{output}}{g} = \underset{\text{input signal}}{f} * \underset{\text{impulse response}}{h}$$

In the frequency domain, via the Fourier transform

$$G(s) = \underset{\text{input spectra}}{F(s)} \underset{\text{transfer function}}{H(s)}$$

To design a filter involves specifying $H(s)$. Recall, mathematically there are positive and negative frequencies to consider, hence filters are defined symmetrically about the vertical axis. A low-pass filter is a rectangle function with desired width and height 1 (idealization, because of the sharp cut-offs)

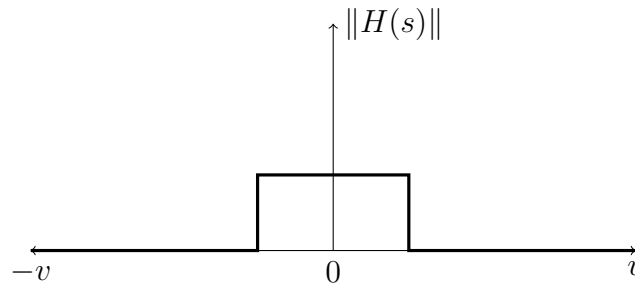


Figure 1: The low-pass filter.

A high-pass filter (allow high frequencies and eliminate low frequencies, useful for edge detection) is a square well.

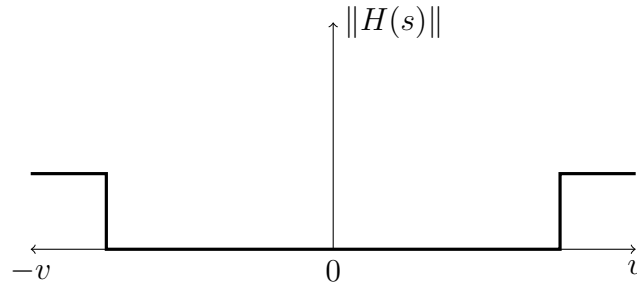


Figure 2: The high-pass filter.

A band-pass filter consists of rectangle functions at the frequencies of interest.

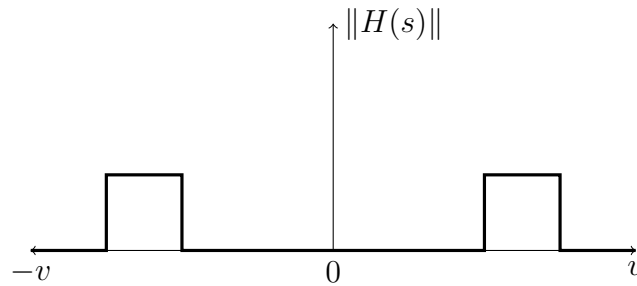


Figure 3: The band-pass filter.

It's comparatively easy to intuit filtering in the frequency domain compared to the time domain... To visualize the effect of filtering in time one must visualize $f * h$ (not worth the effort). An interpretation that is as flexible as its application is "the convolution is what the convolution does" (Brad Osgood). In many contexts, although not universally, convolution is associated with smoothing or averaging, ie: the low-pass filter smooths (but the high-pass filter does not).

Example (DO THIS):

$$\Pi * \Pi = \Lambda$$

Example of smoothing because the triangle function is continuous whereas the rectangle function is not. Furthermore, this example provides a consistency check for the Fourier transform of the triangle function is sinc^2

$$\mathcal{F}(\Pi * \Pi) = (\mathcal{F}\Pi)(\mathcal{F}\Pi) = \text{sinc}^2$$

(FIND EXAMPLE:) An important aspect of convolution is conferring properties of one function to another, ie: turning an aperiodic function into a periodic one through convolution with a periodic function.

Example: If f is differentiable and g is not, then $(f * g)$ is differentiable and

$$(f * g)' = f' * g$$

Similarly with higher order derivatives.

4.6 The Derivative Theorem for Fourier Transforms

$$\mathcal{F}(f') = (i2\pi s)\mathcal{F}f$$

Similarly for higher derivatives

$$\mathcal{F}(f^{(n)}) = (i2\pi s)^n \mathcal{F}f$$

Aphorism: the Fourier transform turns differentiation into multiplication. Fundamental property of F.T. that makes it so useful in applications. The derivation for the case when $f(t) \rightarrow 0$ as $t \rightarrow \pm\infty$ is

$$\mathcal{F}f' = \int_{-\infty}^{\infty} e^{-i2\pi st} f'(t) dt \quad (4)$$

Integrate by parts using

$$\begin{cases} u = e^{-i2\pi st} \\ dv = f'(t) dt \end{cases}$$

$$= \underbrace{\left[f(t)e^{-i2\pi st} \right]_{-\infty}^{\infty}}_{=0} - \int_{-\infty}^{\infty} f(t)(-i2\pi se^{-i2\pi st}) dt \quad (5)$$

$$= (i2\pi s) \int_{-\infty}^{\infty} e^{-i2\pi st} f(t) dt \quad (6)$$

$$= (i2\pi s) \mathcal{F}f \quad (7)$$

$$(8)$$

Do quintessential application of F.T.: The Heat Equation (end of lect. 9) -For the ring: -For the infinite rod:

4.7 The Central Limit Theorem

(Lect 10: Convolution and the Central Limit Theorem) (Lect 11: fix a mistake made at the end of Lect. 10 for expression for the exponential deduced from the truncated Taylor series expansion) - maybe combine this with the section above on the convolution?...

Recall: CLT explains the universal appearance of the Gaussian in probability.

Suppose x_1 and x_2 are independent identically distributed random variables (iid) with distributions $p(x_1) = p(x_2) = p(x)$, respectively. Question: How is the sum of the random variables distributed? Answer: by the convolution $p_1 * p_2$

Proof:

For any value t

$$P(x_1 + x_2 \leq t) = \int \int p(x_1)p(x_2) dx_1 dx_2$$

4.8 The Correlation Theorem

Correlation is defined as

$$\text{Corr}(g, h) \equiv \int_{-\infty}^{\infty} g(\tau + t)h(\tau) d\tau$$

where t is called “the lag”.

The correlation theorem is

$$\text{Corr}(g, h) \iff G(s)H^*(s)$$

when g and h are real.

And, more generally

$$\text{Corr}(g, h) \iff G(s)H(-s)$$

The autocorrelation $\text{corr}(g, g)$, which leads to the Wiener-Khinchin theorem

$$\text{Corr}(g, g) \iff |G(s)|^2$$

4.9 Parseval's Theorem

$$\text{total power} \equiv \int_{-\infty}^{\infty} |h(t)|^2 dt = \int_{-\infty}^{\infty} |H(s)|^2 ds$$

Usually the entity of interest is the power within a frequency interval s to $s + ds$. Here it is common to ignore negative frequencies. Then the “one-sided power spectral density” (PSD) of the function f is

$$P_f(s) \equiv |F(s)|^2 + |F(-s)|^2 \quad 0 \leq s < \infty$$

Numerical Recipes p.603 ... what is the difference between the one-sided and two-sided spectral density?

If the function/signal is infinite then in general so is the total power and power spectral density. Then of interest is the (1 or 2 sided) power spectral density per unit time, ie: select a portion of the signal over a long but finite duration, compute the PSD, and divide by the duration/distance.

5 The Fourier Transform of Basic Signals

5.1 The Box Function, Π

The square pulse (aka box, characteristic, indicator) function is

$$\Pi(t) = \begin{cases} 1 & \text{if } |t| < 1/2 \\ 0 & \text{if } |t| \geq 1/2 \end{cases}$$

The Fourier transform of the box function is

$$\begin{aligned}
\mathcal{F}(\Pi(t)) &= \int_{-\infty}^{\infty} dt \Pi(t) \exp(-2\pi i s t) = \int_{-1/2}^{1/2} dt \exp(-2\pi i s t) \\
&= \left[\frac{-1}{2\pi i s} \exp(-2\pi i s t) \right]_{t=-1/2}^{t=1/2} \\
&= \left(\frac{-1}{2\pi i s} \exp(-\pi i s) \right) - \left(\frac{-1}{2\pi i s} \exp(\pi i s) \right) \\
&= \frac{1}{\pi s} \left(\frac{\exp(\pi i s) - \exp(-\pi i s)}{2i} \right) \\
&= \frac{\sin(\pi s)}{\pi s}
\end{aligned}$$

$$\boxed{\mathcal{F}\Pi = \text{sinc}(s)}$$

Here, the (*normalized*) $\text{sinc}()$ function is defined as $\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$, which is suited for interpolation of bandlimited signals (cite: wikipedia, Sinc function, 10/26/2022). (it is common in Calculus text books for $\text{sinc}(x)$ to be defined as $\text{sinc}(x) = \frac{\sin(x)}{x}$ (cite Strang, Calculus Vol.1, Ch.4.8, p.454)).

The boxcar (aka rectangle) function (Osgood unabashedly calls this the “stretched rect” p.126 [4])

$$\Pi_b(t) = \begin{cases} 1 & \text{if } |t| < b/2 \\ 0 & \text{if } |t| \geq b/2 \end{cases}$$

The Fourier transform of the boxcar function is

$$\begin{aligned}
\mathcal{F}\Pi_b(t) &= \int_{-\infty}^{\infty} dt e^{-i2\pi s t} \Pi_b(t) = \int_{-b/2}^{b/2} dt e^{-i2\pi s t} \\
&= \left[\frac{e^{-i2\pi s t}}{(-i2\pi s)} \right]_{t=-b/2}^{t=b/2} = \frac{1}{i2\pi s} \left(-e^{-i\pi s b} - (-e^{i\pi s b}) \right) \\
&= \frac{1}{\pi s} \left(\frac{e^{i\pi s b} - e^{-i\pi s b}}{i2} \right) = \frac{\sin(b\pi s)}{\pi s}
\end{aligned}$$

Technically, the $\text{sinc}(x)$ function is $\sin(\pi x)/(\pi x)$ - *the denominator is the same as the argument to $\sin()$* (this is a subtle but important point, vital to applications, which you should check using Julia), such that multiplication by a factor of $1 = b/b$ is necessary to construct the sinc function

$$= \frac{b \sin(b\pi s)}{b\pi s}$$

$$\boxed{\mathcal{F}\Pi_b = b \text{sinc}(bs)}$$

This comes up in describing the aperture function in diffraction... This also comes up in sampling when performing interpolation...

The Fourier transform of the (normalized) boxcar function of width b and area one, $\frac{1}{b}\Pi_b(t)$, is

$$\begin{aligned}\mathcal{F}\left(\frac{\Pi_b(t)}{b}\right) &= \frac{1}{b} \int_{-\infty}^{\infty} dt \Pi_b(t) e^{-i2\pi st} = \frac{1}{b} \int_{-b/2}^{b/2} dt e^{-i2\pi st} \\ &= \left[-\frac{e^{-i2\pi st}}{i2\pi sb} \right]_{t=-b/2}^{t=b/2} = \frac{1}{i2\pi sb} \left(-e^{-i\pi sb} - (-e^{i\pi sb}) \right) \\ &= \frac{1}{\pi sb} \left(\frac{e^{i\pi sb} - e^{-i\pi sb}}{i2} \right) = \frac{\sin(b\pi s)}{(b\pi s)}\end{aligned}$$

$$\mathcal{F}\left(\frac{\Pi_b}{b}\right) = \text{sinc}(bs)$$

5.2 The Hat Function, $\Lambda(t)$

The triangle (aka hat, tent) function is

$$\Lambda(t) = \begin{cases} 1 - |t| & \text{if } |t| < 1 \\ 0 & \text{if } |t| \geq 1 \end{cases}$$

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

$$\mathcal{F}(\Lambda(t)) = \int_{-\infty}^{\infty} dt \Lambda(t) \exp(-2\pi ist)$$

Break the integral up into two sections

$$= \int_{-1}^0 dt (1+t) \exp(-2\pi ist) + \int_0^1 dt (1-t) \exp(-2\pi ist)$$

Solve each section using integration by parts.

Making the substitutions for u and v

$$= \int_{-1}^0 \underbrace{(1+t)}_u \underbrace{\exp(-2\pi ist)}_{dv} dt$$

gives $du = dt$, and $v = \frac{-1}{2\pi is} \exp(-2\pi ist)$

$$\begin{aligned}&= \left[(1+t) \left(\frac{-1}{2\pi is} \exp(-2\pi ist) \right) \right]_{t=-1}^{t=0} - \int_{-1}^0 dt \left(\frac{-1}{2\pi is} \right) \exp(-2\pi ist) \\ &= \left(\frac{-1}{2\pi is} \right) + \left[\frac{-1}{(2\pi is)^2} \exp(-2\pi ist) \right]_{t=-1}^{t=0} \\ &= \frac{-1}{2\pi is} + \left(\frac{-1}{(2\pi is)^2} + \frac{\exp(-2\pi ist)}{(2\pi is)^2} \right)\end{aligned}$$

The second section ...

$$\begin{aligned}
&= \int_0^1 \underbrace{(1-t)}_u \underbrace{\exp(-2\pi i s t)}_{dv} dt \\
&= \left[(1-t) \left(\frac{-1}{2\pi i s} \exp(-2\pi i s t) \right) \right]_{t=0}^{t=1} - \int_{-1}^0 (-dt) \left(\frac{-1}{2\pi i s} \right) \exp(-2\pi i s t) \\
&= \frac{-1}{2\pi i s} + \left(-\frac{\exp(-2\pi i s)}{(2\pi i s)^2} + \frac{1}{(2\pi i s)^2} \right)
\end{aligned}$$

Combining the results of both sections gives

$$\begin{aligned}
&= \frac{1}{(\pi s)^2} \frac{\exp(2\pi i s) - \exp(-2\pi i s)}{(2i)^2} \\
&= \frac{\sin^2(\pi s)}{(\pi s)^2}
\end{aligned}$$

$\mathcal{F}\Lambda = \text{sinc}^2(s)$

5.3 The Gaussian, Φ

The Gaussian in its simplest form is

$$\Phi(t) = e^{-\pi t^2}$$

, which when integrated gives

$$\int_{-\infty}^{\infty} dt e^{-\pi t^2} = 1$$

The trick for evaluating its Fourier transform

$$F(s) = \int_{-\infty}^{\infty} e^{-2\pi i s t} e^{-\pi t^2} dt$$

is to notice

$$F'(s) = \frac{d}{ds} \left(\int_{-\infty}^{\infty} e^{-i2\pi s t} e^{-\pi t^2} dt \right) = \int_{-\infty}^{\infty} \frac{d}{ds} \left(e^{-i2\pi s t} \right) e^{-\pi t^2} dt$$

, the derivative with respect to s , is

$$= \int_{-\infty}^{\infty} (-i2\pi t) e^{-i2\pi s t} e^{-\pi t^2} dt$$

, which when written as

$$= (i) \int_{-\infty}^{\infty} \underbrace{e^{-i2\pi s t}}_u \underbrace{((-2\pi t) e^{-\pi t^2})}_{dv} dt$$

suggests integration by parts, with the substitutions

$$\begin{aligned} u &= e^{-i2\pi st} \quad , \quad dv = (-2\pi t) e^{-\pi t^2} dt \\ v &= e^{-\pi t^2} \quad , \quad du = (-i2\pi s) e^{-i2\pi st} dt \end{aligned}$$

Integration by parts $\left(\int uv' = \int (uv)' - \int vu' \right)$ gives

$$\begin{aligned} &= (i) \left\{ \underbrace{\left[e^{-\pi t^2} e^{-i2\pi st} \right]}_{=0} \bigg|_{t=-\infty}^{t=\infty} - \int_{-\infty}^{\infty} e^{-\pi t^2} (-i2\pi s) e^{-i2\pi st} dt \right\} \\ &= (i) \left\{ \left(\underbrace{e^{-\pi(\infty)^2}}_{=0} e^{-i2\pi s(\infty)} - \underbrace{e^{-\pi(-\infty)^2}}_{=0} e^{-i2\pi s(-\infty)} \right) + (i2\pi s) \int_{-\infty}^{\infty} e^{-\pi t^2} e^{-i2\pi st} dt \right\} \\ &= (i^2 2\pi s) \int_{-\infty}^{\infty} e^{-i2\pi st} e^{-\pi t^2} dt \\ &= (-2\pi s) F(s) \end{aligned}$$

The solution that is “screaming out” to solve this differential equation is, $F(s) = e^{-\pi s^2}$, given

$$\frac{d}{ds} \left(e^{-\pi s^2} \right) = (-2\pi s) e^{-\pi s^2}$$

, which in general is subject to some initial condition, but in this case

$$F(0) = \int_{-\infty}^{\infty} e^{-i2\pi(0)t} e^{-\pi t^2} dt = \int_{-\infty}^{\infty} e^{-(\pi)t^2} dt = \sqrt{\frac{\pi}{(\pi)}} = 1$$

and the Fourier transform of a Gaussian *is a Gaussian*

$$\boxed{e^{-\pi t^2} \xleftrightarrow[\mathcal{F}^{-1}]{\mathcal{F}} e^{-\pi s^2}}$$

as is the inverse Fourier transform.

5.4 The Sinc Function, sinc()

This Fourier transform of the sinc() function can be computed without integration using duality

$$\mathcal{F} \text{ sinc} = \mathcal{F} \mathcal{F} \mathbf{\Pi} = \mathbf{\Pi}^- = \mathbf{\Pi}$$

because $\mathbf{\Pi}$ is an even function.

Similarly, for $\text{sinc}^2()$

$$\mathcal{F} \text{ sinc}^2 = \mathcal{F} \mathcal{F} \mathbf{\Lambda} = \mathbf{\Lambda}^{\text{reversed}} = \mathbf{\Lambda}$$

because $\mathbf{\Lambda}$ is an even function.

6 The Modern Fourier Transform

6.1 Issues with the Old Framework

A more robust definition of the Fourier transform is required to handle even the most common signals.

There are two issues

1. Convergence of the integral that defines the F.T.

- The expression for the Fourier transform of even the simplest possible signal $f(t) = 1$

$$\mathcal{F}f = \int_{-\infty}^{\infty} e^{-2\pi i s t} \cdot 1 dt$$

is meaningless.

- Similarly, the F.T. for $\sin(2\pi t)$, and $\cos(2\pi t)$, is not defined (does not converge).

2. Convergence of the integral that defines the inverse F.T.

- The F.T. of the boxcar/rectangle function is

$$\mathcal{F}\Pi = \text{sinc}(s) = \frac{\sin(\pi s)}{\pi s}$$

Duality was used to show the inverse F.T. of the sinc function is the boxcar function. But to actually solve the integral

$$\mathcal{F}^{-1} \text{sinc} = \int_{-\infty}^{\infty} e^{2\pi i s t} \frac{\sin(\pi t)}{\pi t} dt$$

requires special techniques (issue at the end points, $s = \pm 1/2$).

6.2 “Where Do We Go From Here?” (circa 1940)

In response to this ambiguity there are two approaches, 1. ad hoc approaches, or 2. rework the foundations. Taking the latter approach involves...

- Choose the phenomena (a class of functions) upon which to build the theory (that will explain all others).

– Note: \mathcal{S} is a restricted class that serves as the foundation for a broader class.

- Establish the requirements for “the best” class of signals/functions (\mathcal{S}) for performing Fourier transforms

1. If f is in \mathcal{S} then its Fourier transform and inverse Fourier transform are defined and are also in \mathcal{S}

$$f \in \mathcal{S} \Rightarrow \mathcal{F}f \in \mathcal{S}, \text{ and } \mathcal{F}^{-1}f \in \mathcal{S}$$

- Notice: The most fundamental signals, eg: \sin , \cos , 1 , are not in this class \mathcal{S} because the F.T. is not defined.

– Notice: Π is not in the class \mathcal{S} because the F.T. $\mathcal{F}\Pi$, the sinc function, is not in \mathcal{S} .

2. If f is in \mathcal{S} then Fourier inversion works

$$f \in \mathcal{S} \Rightarrow \mathcal{F}^{-1} \mathcal{F} f = \mathcal{F} \mathcal{F}^{-1} f = f$$

3. * **Parseval's Identity** - this is not a requirement of the class \mathcal{S} but comes up often

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

Interpretation: total spectral power can be computed in frequency as well as in time.

Proof: based on definition of the inner product for functions f and $g \in \mathcal{S}$

$$\int_{-\infty}^{\infty} \mathcal{F} f(s) \overline{\mathcal{F} g(s)} ds = \int_{-\infty}^{\infty} f(t) \overline{g(t)} dt$$

Consider

$$\begin{aligned} g(t) &= \mathcal{F}^{-1}((\mathcal{F} g)(s)) \\ &= \mathcal{F}^{-1} \left(\int_{-\infty}^{\infty} g(t) e^{-2\pi i s t} dt \right) \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(t) e^{-2\pi i s t} dt \right) e^{2\pi i s t} ds \\ \overline{g(t)} &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \overline{g(t)} e^{2\pi i s t} dt \right) e^{-2\pi i s t} ds \end{aligned}$$

Substituting this expression for $\overline{g(t)}$ into the definition of the inner product

$$\begin{aligned} \int_{-\infty}^{\infty} f(t) \overline{g(t)} dt &= \int_{-\infty}^{\infty} f(t) \left(\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \overline{g(t)} e^{2\pi i s t} dt \right) e^{-2\pi i s t} ds \right) dt \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(t) e^{-2\pi i s t} dt \right) \left(\int_{-\infty}^{\infty} \overline{g(t)} e^{2\pi i s t} dt \right) ds \\ &= \int_{-\infty}^{\infty} \mathcal{F} f(s) \overline{\mathcal{F} g(s)} ds \end{aligned}$$

The best class of functions for Fourier transforms, known as the “distributions” (distinct from probability distributions, aka generalized functions), are named in honor of Laurent Schwartz, \mathcal{S} . The insight for \mathcal{S} was inspired by the derivative theorem

$$\mathcal{F}(f^{(m)})(s) = (2\pi i s)^m \mathcal{F} f(s)$$

The class \mathcal{S} is described as the set of rapidly decreasing functions, and is defined by the following properties

1. $f(x)$ is infinitely differentiable.

- The constant functions are not in the class \mathcal{S} .
- The rectangle function $\Pi \notin \mathcal{S}$ because it's not continuous.
- The triangle function $\Lambda \notin \mathcal{S}$ because it's not differentiable.

2. Any derivative of $f(x)$ tends to 0 faster than any power of x . For any $m, n \geq 0$

$$|x|^n \left| \frac{d^m f(x)}{dx^m} \right| \longrightarrow 0 \quad \text{as } x \rightarrow \pm\infty$$

- The smooth functions (infinitely differentiable) that are *equal to zero* outside a finite interval (“functions of compact support”) \mathcal{C} are in the class \mathcal{S} .
- The Gaussian is in the class \mathcal{S} .
- The trig functions are not in the class \mathcal{S} .

For the Schwartz \mathcal{S} functions the following properties hold

1. If f is in the Schwartz class, then so is its Fourier transform

$$f \in \mathcal{S} \implies \mathcal{F}f(s) \in \mathcal{S}$$

2. If f is in the Schwartz class, then Fourier inversion works

$$f \in \mathcal{S} \implies \mathcal{F}^{-1} \mathcal{F}f = f$$

The Schwartz functions are ultimately used to define the **tempered distributions**, \mathcal{T} , a broad class of “generalized functions”. The tempered distributions expands the Schwartz \mathcal{S} class to include sines and cosines, delta δ -functions (nonintegrable functions), Furthermore, the properties

$$f \in \mathcal{S} \implies \mathcal{F}f(s) \in \mathcal{S}$$

$$f \in \mathcal{S} \implies \mathcal{F}^{-1} \mathcal{F}f = f$$

continue to hold for the tempered distributions, \mathcal{T} .

Having restricted the class of functions to \mathcal{S} , two requirements must be demonstrated.

- Nothing must be lost by restricting the class of functions to \mathcal{S} .
- Greater generality must be gained.

6.3 The Dirac δ -function

The Dirac delta function, $\delta(x)$ is the quintessential distribution. The statements typically used to define the Dirac- δ lack rigor

1.

$$\begin{cases} \delta(x) = 0, x \neq 0 \\ \delta(0) = \infty \end{cases}$$

2.

$$\int_{-\infty}^{\infty} \delta(x) dx = 1$$

3.

$$\int_{-\infty}^{\infty} \delta(x) \varphi(x) dx = \varphi(0)$$

The δ is supposed to represent a function that is concentrated at a point, and is commonly defined via a limiting process on a function. For example, consider the one-parameter family of rectangle functions Π_ε

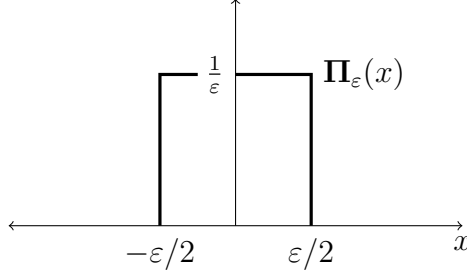


Figure 4: The parent function for the family of rectangle functions of area 1, with width (ε) and height ($1/\varepsilon$).

Note that the limit

$$\lim_{\varepsilon \rightarrow 0} \Pi_\varepsilon(x)$$

on its own makes no sense.

However, operationally, the following does make sense

$$\lim_{\varepsilon \rightarrow 0} \int_{-\infty}^{\infty} \Pi_\varepsilon(x) \varphi(x) dx = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} \varphi(x) dx$$

Assuming $\varphi(x)$ is smooth and continuous, it can be written as a Taylor expansion

$$\begin{aligned} &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} \left(\varphi(0) + \varphi'(0)x + \varphi''(0)\frac{x^2}{2} + \dots \right) dx \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} \varphi(0) dx + \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} \varphi'(0)x dx + \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{-\varepsilon/2}^{\varepsilon/2} \varphi''(0)\frac{x^2}{2} dx + \dots \end{aligned}$$

Beyond the first term is a series of higher order terms in x

$$\begin{aligned} &= \varphi(0) + \lim_{\varepsilon \rightarrow 0} \mathcal{O}(\varepsilon) \\ &= \varphi(0) \end{aligned}$$

Notice, the lim of Π_ε as $\varepsilon \rightarrow 0$ is rigorously defined when *paired with a function and integrated*. This amounts to a fundamental shift in view point, which emphasizes outcomes over the process, ie: extracting the value of a function at a particular point v. concentration via a limiting process.

The Dirac δ -function is endowed with two important properties that are essential for Fourier transforms, the shifting and scaling properties:

1. The shifting property of the δ -function is

$$\varphi * \delta = \int_{-\infty}^{\infty} \varphi(x) \delta(x - b) dx = \varphi(b)$$

, which amounts to convolution with the δ -function (Might have to demonstrate this further...).

2. Somewhat counter intuitively there is the scaling property for the δ -function defined as

$$\delta(ax) = \frac{1}{|a|} \delta(x)$$

For $a > 0$, use the u -substitution $\begin{cases} u = ax \\ du = a dx \end{cases}$

$$\begin{aligned} \int_{-\infty}^{\infty} \delta(ax) \varphi(x) dx &= \frac{1}{a} \int_{-\infty}^{\infty} \delta(u) \varphi\left(\frac{u}{a}\right) du \\ &= \frac{1}{a} \varphi(0) \end{aligned}$$

For $a < 0$ the u -substitution effectively switches the limits of integration because a is negative

$$= \frac{1}{a} \int_{\infty}^{-\infty} \delta(u) \varphi\left(\frac{u}{a}\right) du$$

To recover the original limits of integration introduces a negative sign

$$= -\frac{1}{a} \int_{-\infty}^{\infty} \delta(u) \varphi\left(\frac{u}{a}\right) du$$

But a is negative, so

$$= \frac{1}{|a|} \int_{-\infty}^{\infty} \delta(u) \varphi\left(\frac{u}{a}\right) du$$

, which implies $\delta(ax) = (1/|a|)\delta(x)$. While a scaling factor has been introduced into the argument of the pairing function it is ultimately of no consequence as it will be evaluated at 0 by the delta function. This is distinct from the scaling theorem for the Fourier transform since the argument variable (on the right hand side of the equation) is not also scaled.

7 Distributions

The thermometer analogy...

The definition of distributions/generalized functions

- First specify a class of “test functions” (signals), φ (the best suited for a given problem).

- For the F.T. this is the Schwartz functions.
- Associated with a class of test functions is a class of distributions \mathcal{T} (a distribution is “paired” with a test function).
 - A pairing is defined such that a distribution operates on a test function to produce a number, such that $\mathcal{T}(\varphi) \in \mathbb{C}$.
 - Pairing notation: $\langle \mathcal{T}, \varphi \rangle$ (not an inner product), or (less commonly) $\mathcal{T}(\varphi)$.
 - A physical interpretation might be that φ corresponds to the distribution for a system or of one of its properties, and \mathcal{T} is a measurement.
- \mathcal{T} must be linear.
- \mathcal{T} must be continuous (needed for taking limits)
 - The convergence of a sequence of test functions φ_n converges to a function φ

$$\lim_{n \rightarrow \infty} \{\varphi_n\} \rightarrow \varphi \Rightarrow \mathcal{T}(\varphi_n) \rightarrow \mathcal{T}(\varphi)$$

- ***(This seems like more of a requirement on test functions than on distributions).**
- This is the most problematic aspect of the framework because this is the difficult thing to prove.
- Although each individual function might be infinitely differentiable (as in all the derivatives decay) the more properties conferred upon the component functions the more difficult it becomes to control convergence (foundations for the theory of distributions/generalized functions delve into this).

In sum, a distribution is a continuous linear functional that operates on a set of test function to produce a number.

7.1 The Dirac δ -Function as a Distribution

To recover the Dirac- δ in the context of distributions, consider it operationally as function evaluation (at the origin) and turn that into a definition. For the Dirac- δ the pairing is defined by sheer assertion as

$$\langle \delta, \varphi \rangle = \varphi(0)$$

Check linearity, for the distributive property

$$\langle \delta, \varphi_1 + \varphi_2 \rangle = (\varphi_1 + \varphi_2)(0) = \varphi_1(0) + \varphi_2(0) = \langle \delta, \varphi_1 \rangle + \langle \delta, \varphi_2 \rangle$$

And similarly, for scalar multiplication

$$\langle T, \alpha\varphi \rangle = \alpha\langle T, \varphi \rangle$$

To check continuity, show

$$\varphi_n \rightarrow \varphi \Rightarrow \langle \delta, \varphi_n \rangle \rightarrow \langle \delta, \varphi \rangle$$

By definition

$$\langle \delta, \varphi_n \rangle = \varphi_n(0)$$

If $\varphi_n \rightarrow \varphi \Rightarrow \varphi_n(0) \rightarrow \varphi(0)$, then

$$\langle \delta, \varphi_n \rangle \rightarrow \langle \delta, \varphi \rangle$$

The Dirac- δ is re-couched as the simplest possible distribution. The pairing is defined by sheer assertion. Similarly, the shifted δ -function, δ_a as a distribution is defined by the pairing $\langle \delta_a, \varphi \rangle = \varphi(a)$.

7.2 Distributions Induced by Functions

For distributions “induced” by functions the pairing is defined by integration, which is both linear and *continuous (the latter is difficult to prove, requires limiting theorems for integrals)

$$\langle f, \varphi \rangle = \int_{-\infty}^{\infty} f(x) \varphi(x) dx < \infty$$

, provided the integral converges.

- $f(x) = 1$ as a distribution

$$\langle 1, \varphi \rangle = \int_{-\infty}^{\infty} 1 \cdot \varphi(x) dx$$

- $\Pi(x)$ as a distribution

$$\langle \Pi, \varphi \rangle = \int_{-\infty}^{\infty} \Pi(x) \varphi(x) dx$$

- $f(x) = \exp(-2\pi i a x)$ (which includes sines and cosines)

$$\langle 1, \varphi \rangle = \int_{-\infty}^{\infty} e^{-i2\pi a x} \varphi(x) dx < \infty$$

Notice the flexibility in the definition. There is a trade-off between $\varphi(x)$ and $f(x)$. The more restricted $\varphi(x)$ is, the more freedom $f(x)$ can have. (The space of distributions is the “dual space” of the space of test functions).

7.3 The Fourier Transform of a Distribution

(...from “classical” F.T.s to “generalized” F.T.s) The test functions are the Schwartz functions \mathcal{S} , and the class of distributions is called the class of *tempered* distributions. If T is a tempered distribution, then its F.T. is a tempered distribution. The pairing is defined by integration as

$$\begin{aligned} \langle \mathcal{F}T, \varphi \rangle &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-2\pi i x y} T(y) dy \right) \varphi(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i x y} T(y) \varphi(x) dy dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-2\pi i x y} \varphi(x) dx \right) T(y) dy \\ &= \int_{-\infty}^{\infty} \mathcal{F}\varphi(y) T(y) dy \\ &= \langle T, \mathcal{F}\varphi \rangle \end{aligned}$$

Turn this into a definition. Define $\mathcal{F}T$ by

$$\boxed{\langle \mathcal{F}T, \varphi \rangle = \langle T, \mathcal{F}\varphi \rangle}$$

Likewise, the inverse F.T. of a distribution is

$$\boxed{\langle \mathcal{F}^{-1}T, \varphi \rangle = \langle T, \mathcal{F}^{-1}\varphi \rangle}$$

Importantly, Fourier inversion works for tempered distributions

$$\mathcal{F}^{-1}\mathcal{F}T = \mathcal{F}\mathcal{F}^{-1}T = T$$

$$\begin{aligned} \langle \mathcal{F}^{-1}\mathcal{F}T, \varphi \rangle &= \langle \mathcal{F}T, \mathcal{F}^{-1}\varphi \rangle \\ &= \langle T, \mathcal{F}\mathcal{F}^{-1}\varphi \rangle \\ &= \langle T, \varphi \rangle \end{aligned}$$

because Fourier inversion works for Schwartz functions.

The Fourier transform of the Dirac δ -function is

$$\begin{aligned} \langle \mathcal{F}\delta, \varphi \rangle &= \langle \delta, \mathcal{F}\varphi \rangle \\ &= \mathcal{F}\varphi(0) \end{aligned}$$

Now right the classical definition of the F.T. and evaluate it at zero

$$\begin{aligned} &= \int_{-\infty}^{\infty} e^{-2\pi i 0 x} \varphi(x) dx \\ &= \int_{-\infty}^{\infty} 1 \cdot \varphi(x) dx \\ &= \langle 1, \varphi \rangle \Rightarrow \mathcal{F}\delta = 1 \end{aligned}$$

Notice, for the distribution that is infinitely concentrated its F.T. is uniformly spread out.

The Fourier transform of the shifted delta function, δ_a , is

$$\begin{aligned} \langle \mathcal{F}\delta_a, \varphi \rangle &= \langle \delta_a, \mathcal{F}\varphi \rangle \\ &= \mathcal{F}\varphi(a) \\ &= \int_{-\infty}^{\infty} e^{-2\pi i a x} \varphi(x) dx \\ &= \langle e^{-2\pi i a x}, \varphi \rangle \implies \mathcal{F}\delta_a = e^{-2\pi i a x} \end{aligned}$$

The Fourier transform of the complex exponential is

$$\begin{aligned} \langle \mathcal{F}e^{2\pi i a x}, \varphi \rangle &= \langle e^{2\pi i a x}, \mathcal{F}\varphi \rangle \\ &= \int_{-\infty}^{\infty} e^{2\pi i a x} \mathcal{F}\varphi(x) dx \end{aligned}$$

This is the inverse F.T. of the classical F.T. evaluated at a

$$\begin{aligned} &= \varphi(a) \\ &= \langle \delta_a, \varphi \rangle \implies \mathcal{F} e^{2\pi i a x} = \delta_a \end{aligned}$$

Note that for $a = 1 \longrightarrow \mathcal{F}(1) = \delta$.

The Fourier transform of cosine is even and real

$$\mathcal{F}(\cos(2\pi a x)) = \mathcal{F}\left(\frac{1}{2}(e^{2\pi i a x} + e^{-2\pi i a x})\right) = \frac{1}{2}(\delta_a + \delta_{-a})$$

The Fourier transform of sine is odd and imaginary

$$\mathcal{F}(\sin(2\pi a x)) = \mathcal{F}\left(\frac{1}{2i}(e^{2\pi i a x} - e^{-2\pi i a x})\right) = \frac{1}{2i}(\delta_a - \delta_{-a})$$

7.4 The Derivative of a Distribution

The derivative of a distribution

$$\langle T', \varphi \rangle$$

is defined via the pairing. If T' is given by a function then the pairing is defined by integration

$$\langle T', \varphi \rangle = \int_{-\infty}^{\infty} T'(x) \varphi(x) dx$$

Integration by parts, and using properties of Schwartz functions, gives

$$\begin{aligned} &= \underbrace{T(x)\varphi(x)}_{=0, \varphi(\pm\infty)=0} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} T(x)\varphi'(x) dx \\ &= - \int_{-\infty}^{\infty} T(x)\varphi'(x) dx \\ &= -\langle T, \varphi' \rangle \end{aligned}$$

Note the requirement on Schwartz functions, $\varphi \in \mathcal{S} \Rightarrow \varphi' \in \mathcal{S}$.

Consider the derivative of a function that has no business having a derivative, which is just a jump discontinuity

$$u(x) = \begin{cases} 1 & , x > 0 \\ 0 & , 0 \leq 0 \end{cases}$$

, the Heaviside (unit step) function. The Heaviside function induces a distribution if the integral exists

$$\langle u, \varphi \rangle = \int_{-\infty}^{\infty} u(x) \varphi(x) dx$$

, “and therefore has a derivative because all distributions have derivatives” (Osgood).

So, the **derivative of the Heaviside function**, u' , exists as a distribution.

$$\begin{aligned} \langle u, \varphi \rangle &= -\langle u, \varphi' \rangle = - \int_{-\infty}^{\infty} u(x) \varphi'(x) dx \\ &= - \int_0^{\infty} \varphi'(x) dx \\ &= - \left[\varphi(x) \right]_0^{\infty} = -(\varphi(\infty) - \varphi(0)) \\ &= \varphi(0) = \langle \delta, \varphi \rangle \implies u' = \delta \end{aligned}$$

To take the Heaviside one step further in complexity, consider

$$\text{sgn}(x) = \begin{cases} 1 & , x > 0 \\ 0 & , 0 = 0 \\ -1 & , 0 < 0 \end{cases}$$

, the signum function.

The **derivative of the signum (aka sign) function** is

$$\begin{aligned} \langle \text{sgn}', \varphi \rangle &= -\langle \text{sgn}, \varphi' \rangle = - \left(\int_{-\infty}^0 (-1) \varphi'(x) dx + \int_0^{\infty} (+1) \varphi'(x) dx \right) \\ &= - \left((-\varphi(0) - (-\varphi(-\infty))) + (\varphi(\infty) - \varphi(0)) \right) \\ &= \varphi(0) + \varphi(0) = 2\varphi(0) \\ &= \langle 2\delta, \varphi \rangle \implies \text{sgn}'(x) = 2\delta \end{aligned}$$

Applications to F.T.s

Derivative Theorem Pt. I:

$$\boxed{\mathcal{F}(T') = (2\pi i s) \mathcal{F}T}$$

$$\begin{aligned}
\langle \mathcal{F}T', \varphi \rangle &= \int_{-\infty}^{\infty} \mathcal{F}T' \varphi(t) dt \\
&= \int_{-\infty}^{\infty} \underbrace{\left(\int_{-\infty}^{\infty} T' e^{-2\pi i s t} dt \right)}_{\text{By Parts}} \varphi(t) dt \\
&= \int_{-\infty}^{\infty} \left(T e^{-2\pi i s t} \Big|_{-\infty}^{\infty} - (-2\pi i s) \int_{-\infty}^{\infty} T e^{-2\pi i s t} dt \right) \varphi(t) dt \\
&= \int_{-\infty}^{\infty} \left(\underbrace{T e^{-2\pi i s t} \varphi(t) \Big|_{-\infty}^{\infty}}_{=0} + (2\pi i s) \int_{-\infty}^{\infty} T e^{-2\pi i s t} \varphi(t) dt \right) dt \\
&= (2\pi i s) \int_{-\infty}^{\infty} \mathcal{F}T \varphi(t) dt = (2\pi i s) \langle \mathcal{F}T, \varphi \rangle
\end{aligned}$$

Derivative Theorem Pt. II:

$$\boxed{(\mathcal{F}T)' = \mathcal{F}(-2\pi i s)T}$$

$$\begin{aligned}
\langle (\mathcal{F}T)', \varphi \rangle &= -\langle \mathcal{F}T, \varphi' \rangle \\
&= - \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} T e^{-2\pi i s t} dt \right] \varphi'(t) dt \\
&= - \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \varphi'(t) e^{-2\pi i s t} dt \right) T dt \\
&= - \int_{-\infty}^{\infty} \left(\left[e^{-2\pi i s t} \varphi(t) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-2\pi i s) \varphi(t) e^{-2\pi i s t} dt \right) T dt \\
&= - \int_{-\infty}^{\infty} (2\pi i s) \int_{-\infty}^{\infty} \varphi(t) e^{-2\pi i s t} dt T dt \\
&= -(2\pi i s) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T e^{-2\pi i s t} dt \varphi(t) dt \\
&= -2\pi i s \langle \mathcal{F}T, \varphi \rangle
\end{aligned}$$

Use the derivative theorem pt. 1 to find the Fourier transform of the sgn function

Based on the formula for $\text{sgn}' = 2\delta$

$$\mathcal{F}(\text{sgn}') = \mathcal{F}(2\delta) = 2$$

Using the derivative theorem

$$\mathcal{F}(\text{sgn}') = (2\pi i s) \mathcal{F}(\text{sgn})$$

Combining these two expressions

$$\mathcal{F} \operatorname{sgn}(s) = \frac{1}{\pi i s}$$

This derivation requires further justification (the function $\frac{1}{s}$ has a singularity, to show this is a distribution requires a special definition for the pairing, the principal value distribution, refer to text).

The Fourier transform of the signum function provides a simple way to obtain the **Fourier transform of the Heaviside function**.

Writing the Heaviside function in terms of the signum function, allowing the value at the origin to be $1/2$, as

$$u(x) = \frac{1}{2}(1 + \operatorname{sgn}(x))$$

then its Fourier transform is given by

$$\begin{aligned} \mathcal{F}(u(x)) &= \frac{1}{2} \mathcal{F}(1 + \operatorname{sgn}) \\ &= \frac{1}{2} \left(\delta + \frac{1}{\pi i s} \right) \end{aligned}$$

7.5 The Shift Theorem for Distributions

$$\langle \tau_b T, \varphi \rangle = \langle T, \tau_{-b} \varphi \rangle$$

Introduce the shift operator, τ_b

$$(\tau_b \varphi)(x) = \varphi(x - b)$$

Assume the tempered distribution, T , comes from a function $f(x - b)$ and the pairing is defined by integration

$$\begin{aligned} \langle \tau_b f, \varphi \rangle &= \langle f(x - b), \varphi \rangle = \\ &= \int_{-\infty}^{\infty} (\tau_b f)(x) \varphi(x) dx = \int_{-\infty}^{\infty} f(x - b) \varphi(x) dx \end{aligned}$$

Use the u -substitution $u = x - b$

$$\begin{aligned} &= \int_{-\infty}^{\infty} f(u) \varphi(u + b) du \\ &= \int_{-\infty}^{\infty} f(x) (\tau_{-b} \varphi)(x) dx = \langle f, \tau_{-b} \varphi \rangle \end{aligned}$$

The switch from x to u introduces ambiguity that the “variable-free” angle-bracket notation avoids.

The shift theorem for the Fourier transform of a tempered distribution, T , is

$$\mathcal{F}(\tau_b T) = e^{-2\pi i b x} \mathcal{F} T$$

$$\langle \mathcal{F}(\tau_b T), \varphi \rangle = \langle \tau_b T, \mathcal{F} \varphi \rangle$$

$$\begin{aligned}
&= \langle T, \tau_{-b} \mathcal{F} \varphi \rangle \\
\tau_{-b}(\mathcal{F} \varphi)(x) &= \mathcal{F} \varphi(s+b) \\
&= \int_{-\infty}^{\infty} e^{-2\pi i(s+b)x} \varphi(x) dx \\
&= \int_{-\infty}^{\infty} e^{-2\pi i s x} e^{-2\pi i b x} \varphi(x) dx \\
&= \mathcal{F}(e^{-2\pi i b x} \varphi)(s)
\end{aligned}$$

Plugging in

$$\begin{aligned}
&\langle T, \tau_{-b} \mathcal{F} \varphi \rangle \\
&= \langle T, \mathcal{F}(e^{-2\pi i b x} \varphi) \rangle \\
&= \langle \mathcal{F} T, e^{-2\pi i b x} \varphi \rangle \\
&= \langle e^{-2\pi i b x} \mathcal{F} T, \varphi \rangle
\end{aligned}$$

It is instructive to examine the effect on the Dirac δ -function... The shifted δ -function is

$$\begin{aligned}
\langle \tau_a \delta, \varphi \rangle &= \langle \delta, \tau_{-a} \varphi \rangle \\
&= (\tau_{-a} \varphi)(0) \\
&= \varphi(0+a) = \langle \delta_a, \varphi \rangle
\end{aligned}$$

So,

$$\tau_a \delta = \delta_a$$

The Fourier transform of the shifted δ -function is

$$\mathcal{F}(\tau_b \delta) = e^{-2\pi i b s} \mathcal{F} \delta = e^{-2\pi i b s}$$

...as expected.

7.6 The Scaling Theorem for Distributions

$$\boxed{\langle \varsigma_a f, \varphi \rangle = \left\langle f, \frac{\varsigma_{1/a}}{|a|} \varphi \right\rangle}$$

Assume the tempered distribution comes from a function $f(ax)$ and the pairing is defined by integration

$$\langle f(ax), \varphi \rangle = \int_{-\infty}^{\infty} f(ax) \varphi(x) dx$$

For $a > 0$, the u -substitution $u = ax$ gives

$$= \frac{1}{a} \int_{-\infty}^{\infty} f(u) \varphi\left(\frac{u}{a}\right) du$$

For $a < 0$, the u -substitution $u = ax$ gives

$$\begin{aligned} &= \frac{1}{a} \int_{\infty}^{-\infty} f(u) \varphi\left(\frac{u}{a}\right) du \\ &= \frac{-1}{a} \int_{-\infty}^{\infty} f(u) \varphi\left(\frac{u}{a}\right) du \end{aligned}$$

But a is negative, so

$$= \frac{1}{|a|} \int_{-\infty}^{\infty} f(u) \varphi\left(\frac{u}{a}\right) du$$

To avoid ambiguity in notation caused by conferring the factor of $\frac{1}{a}$ to the argument of $\varphi(x)$ introduce the **scaling operator**, ς , which scales *the argument* of the function

$$\varsigma_a \varphi(x) = \varphi(ax)$$

Combining the results

$$\begin{aligned} \langle \varsigma_a f, \varphi \rangle &= \int_{-\infty}^{\infty} (\varsigma_a f(x)) \varphi(x) dx \\ &= \frac{1}{|a|} \int_{-\infty}^{\infty} f(x) (\varsigma_{1/a} \varphi(x)) dx \\ &= \left\langle f, \frac{\varsigma_{1/a}}{|a|} \varphi \right\rangle \end{aligned}$$

7.7 The Scaling Theorem for the Fourier Transform

$$\boxed{\mathcal{F}(\varsigma_a T) = \frac{1}{|a|} \varsigma_{1/a}(\mathcal{F}T)}$$

(but what does this actually mean? How to implement this in practice?) Using the definition of the Fourier transform

$$\langle \mathcal{F}(\varsigma_a T), \varphi \rangle = \langle \varsigma_a T, \mathcal{F} \varphi \rangle$$

Applying the scaling theorem

$$= \left\langle T, \frac{\varsigma_{1/a}}{|a|} \mathcal{F} \varphi \right\rangle$$

Applying the scaling theorem again (“Any property or operation on test functions may be extended to distributions” (p.12-13, [20]))

$$= \langle T, \mathcal{F}(\varsigma_a \varphi) \rangle$$

Using the definition of the Fourier transform again

$$= \langle \mathcal{F}T, \varsigma_a \varphi \rangle$$

Applying the scaling theorem yet again

$$= \left\langle \frac{\varsigma_{1/a}}{|a|} \mathcal{F}T, \varphi \right\rangle$$

Honestly WFT, I hate this proof...

NOW:... The scaling property of the delta function defined via the scaling operator What happens when the scaling operator is applied to the shifted delta function?...is this essentially what is happening in the diffraction example.

7.8 Multiplication and Convolution of Distributions

Not all operations on functions carry over to distributions. It is not generally true that multiplication of distributions makes sense. If \mathcal{S}, \mathcal{T} are distributions, then $\mathcal{S}\mathcal{T}$ is not necessarily defined. Rather, a function times a distribution, $f\mathcal{T}$, using integration to specify the pairing

$$\begin{aligned} \langle f\mathcal{T}, \varphi \rangle &= \int_{-\infty}^{\infty} f(x)\mathcal{T}(x)\varphi(x) dx \\ &= \int_{-\infty}^{\infty} \mathcal{T}(x)f(x)\varphi(x) dx \\ &= \langle \mathcal{T}, f\varphi \rangle \end{aligned}$$

is more likley to be defined. In general, define $f\mathcal{T}$ as

$$\langle f\mathcal{T}, \varphi \rangle = \langle \mathcal{T}, f\varphi \rangle$$

provided $f\varphi$ has the properties that define the class, is a test function. (this is actually necessary to derive the derivative theorem pt. 1)

A special case of multiplying a function times a distribution is, $f\delta$

$$\begin{aligned} \langle f\delta, \varphi \rangle &= \langle \delta, f\varphi \rangle \\ &= f(0)\varphi(0) \\ &= \langle f(0)\delta, \varphi \rangle \\ &\Rightarrow \boxed{f\delta = f(0)\delta} \end{aligned}$$

More generally,

$$\boxed{f\delta_a = f(a)\delta_a}$$

This is the sampling property of the delta δ -function.

Similarly, if \mathcal{S} and \mathcal{T} are distributions the convolution, $\mathcal{S} * \mathcal{T}$, is not always defined. There are some additional restrictions on \mathcal{S} and \mathcal{T} in order to define the pairing for the convolution of distributions. There are many cases where it is defined but what is most often considered is a function convolved with a distribution, $f * \mathcal{T}$. Importantly, the convolution theorem holds, that is $\mathcal{F}(f * \mathcal{T}) = (\mathcal{F}f)(\mathcal{F}\mathcal{T})$.

A special case, which is particularly important is when $\mathcal{T} = \delta$

$$\boxed{f * \delta = f}$$

(the δ -function serves as the identity element for convolution). More generally, the shifting property of the δ -function is based on convolution of a function with the shifted δ_a -function

$$(f * \delta_a)(x) = f(x - a)$$

which produces a shifted version of the function.

A case when the convolution of two distributions does makes sense is

$$\delta_a * \delta_b = \delta_{a+b}$$

the convolution of two δ -functions. Note that $(f * \delta_a) * \delta_b = f(x - a) * \delta_b = f(x - a - b)$. Furthermore,

$$\begin{aligned} f * \delta_{a+b} &= f(x - (a + b)) \\ &= f(x - a - b) \end{aligned}$$

, so it is consistent.

7.9 Application of the Fourier Transform to Diffraction

source: monochromatic light source at large distance from the aperture plane such that waves impinging on the aperture plane can be regarded as plane waves. aperture plane: apertures on a flat surface generate waves of uniform strength (E_0), frequency, and (initially) phase - Huygens' principle (each aperture generates a new source). Image plane: diffraction pattern observed, based on field strength E . Near-field (Fresnel) v. far-field (Fraunhofer) diffraction: based on the distance between the image and the source relative to the wavelength. (Another thing to consider is the aperture size relative to the wavelength.) Light is represented by electro(magnetic) oscillations $Ee^{2\pi i \nu t}$, where E is the magnitude of the electric field.

Fraunhofer Diffraction In Fraunhofer diffraction the light source is at a distance from the aperture plane large relative to the wave length, such that impinging wave fronts can be regarded as plane waves. Assume constant strength of the electric field, and uniform phase, at the aperture plane. Reduce the problem to a one dimensional problem by looking at the aperture plane sideways. Regarding a vertical distance, dx , as a small distance along the aperture, the real part of

$$E_0 e^{2\pi i \nu t} dx$$

represents the electric part of the electromagnetic field generated along the aperture.

Phase change associated with separate paths to a point on the image plane is

$$\frac{2\pi r}{\lambda}$$

, where r is the distance from the aperture plane to the image plane. The magnitude of the electric field at a distance r is

$$dE = E_0 e^{2\pi i \nu t} e^{-2\pi i r/\lambda} dx$$

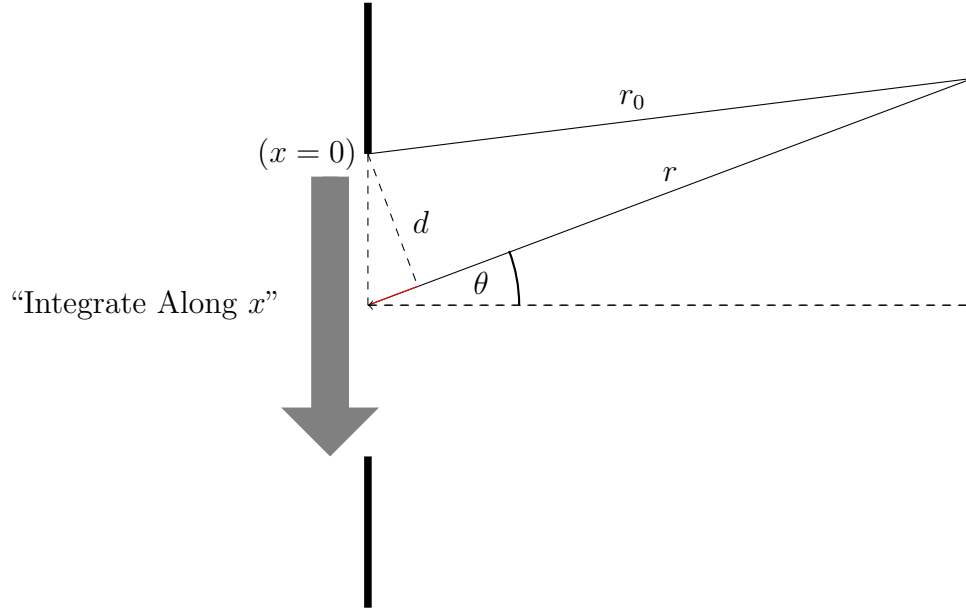
Integrating over (x) the aperture plane

$$E = \int_{\text{aperture}} E_0 e^{2\pi i \nu t} e^{-2\pi i r/\lambda} dx$$

One observes and experiments measure the magnitude of the electric field $|E|$, the total field strength at the image plane, so the time varying part becomes 1 and can be ignored

$$= \int_{\text{aperture}} e^{-2\pi i r/\lambda} dx$$

However, this expression is too complicated to be useful ... The Fraunhofer approximation to the diffraction experiment where the distance between the aperture plane and the image plane is large relative to the aperture size. Here, the difference in path length between points along the aperture is is approximately $\approx x \sin \theta$. Where, d is the perpendicular from the longer path to the aperture of the shorter path. And, θ is the angle between the perpendicular (d) and the aperture plane



$$\begin{aligned} E &= \int_{\text{aperture}} e^{-2\pi i(r_0 + x \sin \theta)/\lambda} dx \\ &= e^{-2\pi i r_0/\lambda} \int_{\text{aperture}} e^{-2\pi i x \sin \theta/\lambda} dx \end{aligned}$$

Again, the complex exponential factor can be ignored.

Introduce $p = \sin \theta/\lambda$

$$= \int_{\text{apertures}} e^{-2\pi i x p} dx$$

Introduce the “aperture function”

$$A(x) = \begin{cases} 1 & x \in \text{aperture} \\ 0 & \text{otherwise} \end{cases}$$

, and integrating over the entire aperture plane

$$E = \int_{-\infty}^{\infty} e^{-2\pi i x p} A(x) dx$$

, gives the Fourier transform of the aperture function, $\mathcal{F}A(p)$. For far-field diffraction under the Fraunhofer approximation, the diffraction pattern (strength of the field, intensity of light) at the image plane is the magnitude of the Fourier transform of the aperture function.

Single Slit Diffraction: The aperture function for a slit of width a is

$$\Pi_a(x)$$

The Fourier transform is

$$a \operatorname{sinc}(ap)$$

where $p = \sin \theta / \lambda$. The absolute value (magnitude/intensity) is observed at the image plane.

Diffraction from a Point Source: The aperture function for a point source is

$$A(x) = \delta$$

The Fourier transform is

$$\mathcal{F}\delta = 1$$

The image plane is uniformly illuminated.

Young's Double Slit Experiment: The aperture function is

$$A(x) = \Pi_a(x - b/2) + \Pi_a(x + b/2)$$

where the slits of width a are separated by a distance b . The Fourier transform (using the modulation theorem) is

$$a \operatorname{sinc}(ap) 2 \cos(\pi bp)$$

where $p = \sin \theta / \lambda$.

X-ray Crystallography: a sampling primer

Treat the 1-D crystal as a periodized version of the electron density of a single atom $\rho(x)$

$$P_d(x) = \sum_{n=-\infty}^{\infty} \rho_d = \sum_{n=-\infty}^{\infty} \rho(x - nd)$$

as an infinite array of identical atoms evenly spaced with spacing (d) along a line. The shifted electron density can be written as the convolution of the electron density with a shifted δ -function

$$\rho_d = \rho(x - nd) = \rho(x) * \delta(x - nd)$$

and write the periodization

$$\begin{aligned} P_d(x) &= \sum_{n=-\infty}^{\infty} \rho(x - nd) \\ &= \sum_{n=-\infty}^{\infty} \rho(x) * \delta(x - nd) \\ &= \rho(x) * \left(\sum_{n=-\infty}^{\infty} \delta(x - nd) \right) \end{aligned}$$

Introduce the Shah function/distribution (Bracewell 1999) of spacing, d

$$= \rho(x) * \text{III}_d(x)$$

The diffraction pattern is determined by the Fourier transform of P_d . By the convolution theorem

$$\mathcal{F}(P_d) = \mathcal{F}(\rho(x) * \text{III}_d) = (\mathcal{F}\rho)(\mathcal{F}\text{III}_d)$$

Now the problem reduces to, “what is $\mathcal{F}\text{III}$?” The Shaw function makes sense as a distribution

$$\text{III}(x) = \sum_{n=-\infty}^{\infty} \delta(x - n)$$

in that it operates on a test function as

$$\langle \text{III}, \varphi \rangle = \int_{-\infty}^{\infty} \text{III} \varphi(x) dx = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - n) \varphi(x) dx = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x - n) \varphi(x) dx = \sum_{n=-\infty}^{\infty} \varphi(n)$$

, and the sum converges (because φ is rapidly decreasing), so the pairing makes sense. Because the III -function makes sense as a distribution so does its Fourier transform

$$\langle \mathcal{F}\text{III}, \varphi \rangle = \langle \text{III}, \mathcal{F}\varphi \rangle = \sum_{n=-\infty}^{\infty} \mathcal{F}\varphi(n)$$

Note, classically this series does not converge

$$\mathcal{F}\text{III} = \sum_{n=-\infty}^{\infty} \mathcal{F}(\delta(x - n)) = \sum_{n=-\infty}^{\infty} e^{-i2\pi sn}$$

although it makes sense as a distribution.

The modern approach involves **the Poisson summation formula** (one of the deep insights regarding the integers). If φ is a rapidly decreasing function (there are other classes for which this applies as well), then

$$\boxed{\sum_{n=-\infty}^{\infty} \varphi(n) = \sum_{n=-\infty}^{\infty} \mathcal{F}\varphi(n)}$$

Proof:

The periodized version[20] of φ with period 1 is (why is it periodized? in what sense is it periodized? In the case of describing the crystal it was clear what periodization was for. But why is it necessary to replicate just a function? I don't know if this answer is really enough, but the periodically replicated signal definitely has a Fourier transform!)

$$\phi(x) = \sum_{n=-\infty}^{\infty} \varphi(x - n)$$

Expanding ϕ in a Fourier series

$$= \sum_{n=-\infty}^{\infty} \hat{\phi}(n) e^{i2\pi sn}$$

where the n^{th} Fourier coefficient is $\hat{\phi}(n) = \mathcal{F}\varphi(n)$

$$= \sum_{n=-\infty}^{\infty} \mathcal{F}\varphi(n) e^{i2\pi sn}$$

Evaluating the two expressions for ϕ at x equals zero

$$\phi(0) = \sum_{n=-\infty}^{\infty} \varphi(-n) = \sum_{n=-\infty}^{\infty} \varphi(n) \quad \text{and} \quad \phi(0) = \sum_{n=-\infty}^{\infty} \mathcal{F}\varphi(n)$$

allows them to be equated, and gives the Poisson summation formula

$$\Rightarrow \sum_{n=-\infty}^{\infty} \varphi(n) = \sum_{n=-\infty}^{\infty} \mathcal{F}\varphi(n)$$

Invoking the Poisson summation formula, the **Fourier transform of the Shah III-function (as a distribution) is the Shah III-function**

$$\langle \mathcal{F}\text{III}, \varphi \rangle = \langle \text{III}, \mathcal{F}\varphi \rangle = \sum_{n=-\infty}^{\infty} \mathcal{F}\varphi(n) = \sum_{n=-\infty}^{\infty} \varphi(n) = \langle \text{III}, \varphi \rangle$$

$$\boxed{\mathcal{F}\text{III} = \text{III}}$$

The **Fourier transform of the Shah function with spacing (d)**

$$\boxed{\mathcal{F}\text{III}_d = \frac{1}{d}\text{III}_{1/d}}$$

Scaling the shifting parameter of the δ -function by (d)

$$\text{III}_d(x) = \sum_{n=-\infty}^{\infty} \delta(x - nd)$$

Factoring the scaling factor

$$= \sum_{n=-\infty}^{\infty} \delta\left(d\left(\frac{x}{d} - n\right)\right)$$

By the scaling property of the δ -function ($\delta(ax) = (1/|a|)\delta(x)$)

$$= \frac{1}{d} \sum_{n=-\infty}^{\infty} \delta\left(\frac{x}{d} - n\right)$$

$$= \frac{1}{d} \text{III}\left(\frac{x}{d}\right)$$

Then, the Fourier transform is

$$\mathcal{F}(\text{III}_d(x)) = \frac{1}{d} \mathcal{F}\left(\text{III}\left(\frac{x}{d}\right)\right)$$

By the scaling theorem for Fourier transforms ($\mathcal{F}(\varkappa_a f) = (1/|a|)(\mathcal{F}(\varkappa_{1/a} f))$, where \varkappa , scales *the argument* of the function it is applied to (**UPDATE notation from ς to \varkappa in previous work on scaling theorems and distributions**))

$$\begin{aligned}
&= \frac{1}{d} \left((d) \varkappa_d (\mathcal{F} \mathbb{I}) \right) \\
&= \mathbb{I}(dx) = \sum_{n=-\infty}^{\infty} \delta(dx - n) \\
&= \sum_{n=-\infty}^{\infty} \delta \left(d \left(x - \frac{n}{d} \right) \right) \\
&= \frac{1}{d} \sum_{n=-\infty}^{\infty} \delta \left(x - \frac{n}{d} \right) = \frac{1}{d} \mathbb{I}_{1/d}(x)
\end{aligned}$$

Notice the reciprocal relationship between the original domain and the Fourier transform domain. The spacing *and the amplitude* are scaled by $(1/d)$.

The crystal is represented by the periodized version of the electron density distribution for the atom with spacing (d)

$$P_d(x) = \rho(x) * \mathbb{I}_d(x)$$

The observation of the diffraction experiment produces the Fourier transform

$$\begin{aligned}
\mathcal{F}(\rho * \mathbb{I}_d) &= (\mathcal{F} \rho)(\mathcal{F} \mathbb{I}_d) \\
&= (\mathcal{F} \rho) \left(\frac{1}{d} \mathbb{I}_{1/d} \right) \\
&= \frac{1}{d} (\mathcal{F} \rho(x)) \sum_{n=-\infty}^{\infty} \delta \left(x - \frac{n}{d} \right) \\
&= \frac{1}{d} \sum_{n=-\infty}^{\infty} \mathcal{F} \rho(x) \delta \left(x - \frac{n}{d} \right) \\
&= \frac{1}{d} \sum_{n=-\infty}^{\infty} \mathcal{F} \rho \left(\frac{n}{d} \right) \delta \left(x - \frac{n}{d} \right)
\end{aligned}$$

The spacing between the spots on the diffraction pattern is proportional to *the inverse* of the spacing between the atoms. In general, the diffraction pattern is given by the Fourier transform of the aperture function. The representation for the crystal lattice for X-ray diffraction should be based on the electron density, which is comprised mostly of empty space in comparison to the nuclei, which demarcate the aperture boundary. (While seemingly banal, this is an extremely important point based on the structure of the atom, without which leads to ambiguity when interpreting the diffraction pattern.)

8 Sampling and Interpolation

*** note: “interpolation” in the context of the Fourier transform amounts to reconstructing the signal from the set of samples - the data points - but specifically from the frequency content extracted from the sampling process. ***

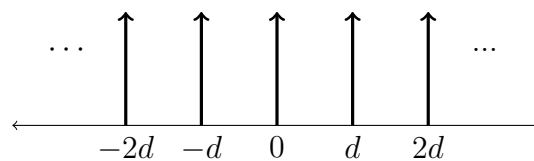
The aperture can be understood as a sampling using a density to describe the measuring device. The density shifted by a distance d can be written as the convolution with a shifted δ -function

$$\rho(x - d) = \rho(x) * \delta(x - d)$$

The periodized version aka periodization is (**UPDATE?: does periodized mean "replicated"?)

$$\begin{aligned}
 P_d(x) &= \sum_{n=-\infty}^{\infty} \rho(x - nd) \\
 &= \sum_{n=-\infty}^{\infty} \rho(x) * \delta(x - nd) \\
 &= \rho(x) * \sum_{n=-\infty}^{\infty} \delta(x - nd) \\
 &= \rho(x) * \text{III}_d(x)
 \end{aligned}$$

The Shah function/distribution III (aka Dirac comb, pulse-train) is

$$\text{III}_d(x) = \sum_{n=-\infty}^{\infty} \delta(x - nd)$$


Three important properties of the III -function are critical for solving the interpolation problem

1. Sampling

$$f(x) \text{III}_d(x) = \sum_{n=-\infty}^{\infty} f(nd) \delta(x - nd)$$

2. Periodization

$$f(x) * \text{III}_d(x) = \sum_{n=-\infty}^{\infty} f(x - nd)$$

3. Fourier transform

$$\mathcal{F}(\text{III}_d) = \frac{1}{d} \text{III}_{1/d} = \mathcal{F}^{-1} \text{III}_d$$

The Interpolation Problem: *Sampling a process evolving in time at equal intervals, can one interpolate all values of a signal from a discrete set of measurements?* ie: can one write the function $f(t) \forall t$ based on a discrete set of samples, $f(t_n)$ at t_n ? An important issue to consider is, the more rapidly a signal oscillates the more uncertainty there is in interpolating the values between sample points.

8.1 Bandlimited Signals

***** ** Basis for a Medium Article: Are there two definitions for the 'bandwidth' in use? ** ***** **The definition of the bandwidth here is...not standard? wrong? domain specific? ... And not what is used in the treatment of the DFT...?** ref. (Osgood, The FT and Applications, CH 5.6, Sampling and Interpolation for Bandlimited Signals, p.225). **However, it is important to point out that in Osgood's definition, the reciprocal of the bandwidth is the sample spacing. Which is an important mental check...**

The wikipedia definition of bandwidth (signal processing) is "the difference between the upper and lower frequencies in a continuous band of frequencies", which sounds like it at least agrees with Osgood's definition.

For bandlimited signals the interpolation problem can be solved exactly. A function $f(t)$ is "bandlimited" if

$$(\mathcal{F}f)(s) \equiv 0 \quad \text{for } |s| \geq \Omega/2$$

where the smallest such (Ω) is the "bandwidth".

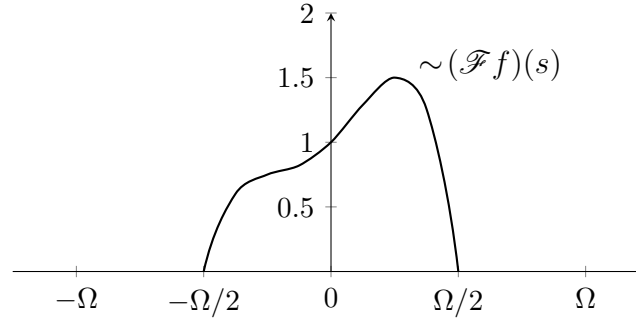


Figure 5: Conceptual representation of the Fourier transform of f (this image is not to be taken literally, because the Fourier transform is in general complex). For Real signals, the Fourier transform in the frequency domain is symmetric.

To derive the expression for the signal reconstruction start with the periodization of the Fourier transform by convolution with III_Ω .

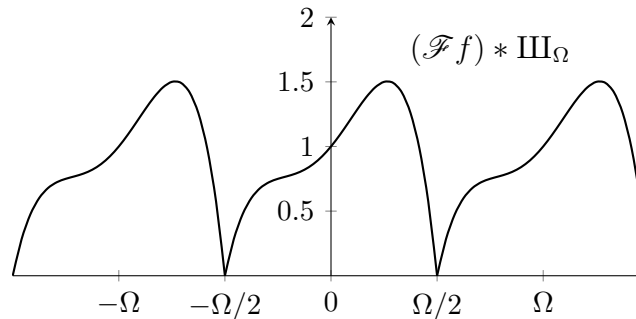


Figure 6: The periodization of the Fourier transform.

There is no overlap between copies because the spacing between δ 's of the Shah function III_Ω is Ω , and $\mathcal{F}f$ is 0 outside $[-\Omega/2, \Omega/2]$.

To recover the Fourier transform $(\mathcal{F}f)$ multiply the periodization (in the frequency domain) by a boxcar of width Ω (the bandwidth), Π_Ω

$$\boxed{\mathcal{F}f = \Pi_\Omega (\mathcal{F}f * \text{III}_\Omega)} \tag{9}$$

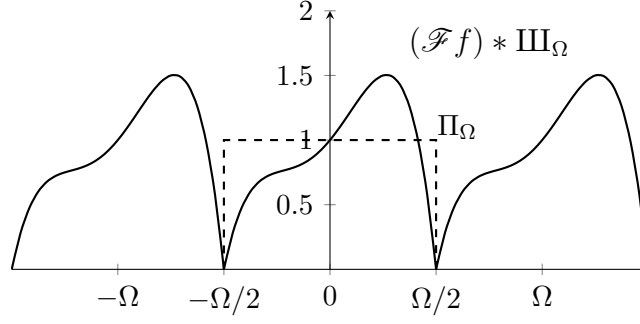


Figure 7: The periodization of the Fourier transform $(\mathcal{F}f) * \text{III}_\Omega$ (solid curves), and the boxcar of width Ω , III_Ω (dashed lines).

To recover the original signal take the inverse Fourier transform

$$\begin{aligned} f &= \mathcal{F}^{-1}(\mathcal{F}f) = \mathcal{F}^{-1}(\Pi_\Omega (\mathcal{F}f * \text{III}_\Omega)) \\ &= (\mathcal{F}^{-1}(\Pi_\Omega)) * (\mathcal{F}^{-1}(\mathcal{F}f * \text{III}_\Omega)) \end{aligned} \quad (10)$$

- The first part is

$$\begin{aligned} \mathcal{F}^{-1}(\Pi_\Omega) &= \mathcal{F}\left(\Pi_\Omega^{(-)}\right) \quad (\text{by duality}) \\ &= \mathcal{F}(\Pi_\Omega) \quad (\text{because } \Pi \text{ is even}) \\ &= \Omega \text{sinc}(\Omega t) \quad (\text{the Fourier transform of the boxcar function}) \end{aligned}$$

- The second part is

$$\mathcal{F}^{-1}((\mathcal{F}f) * \text{III}_\Omega) = (\mathcal{F}^{-1}\mathcal{F}f)(\mathcal{F}^{-1}\text{III}_\Omega)$$

By the scaling property of the Fourier transform, $\mathcal{F}(\kappa_{(a)}f) = (\kappa_{(1/a)}/|a|)(\mathcal{F}f)$ (where $\kappa_{(a)}$ scales the argument of the function by a factor of (a))

$$= f(t) \frac{1}{\Omega} \text{III}_{1/\Omega}(t)$$

Using the sampling property of the III -function (in the time domain)

$$\begin{aligned} &= \frac{1}{\Omega} f(t) \sum_{n=-\infty}^{\infty} \delta\left(t - \frac{n}{\Omega}\right) \\ &= \frac{1}{\Omega} \sum_{n=-\infty}^{\infty} f\left(\frac{n}{\Omega}\right) \delta\left(t - \frac{n}{\Omega}\right) \end{aligned}$$

The signal can now be written as

$$f(t) = \left(\Omega \text{sinc}(\Omega t)\right) * \left(\frac{1}{\Omega} \sum_{n=-\infty}^{\infty} f\left(\frac{n}{\Omega}\right) \delta\left(t - \frac{n}{\Omega}\right)\right)$$

The function f is unaffected by distributing the convolution inside the summation, only the $\text{sinc}()$ function and the δ -function contain (t)

$$= \sum_{n=-\infty}^{\infty} f\left(\frac{n}{\Omega}\right) \text{sinc}(\Omega t) * \delta\left(t - \frac{n}{\Omega}\right)$$

Use the shifting property of the δ -function to apply the convolution (see section on the Dirac δ -function)

$$= \sum_{n=-\infty}^{\infty} f\left(\frac{n}{\Omega}\right) \text{sinc}\left(\Omega\left(t - \frac{n}{\Omega}\right)\right)$$

This formula, which specifies the function at all values of t based on the sampled values, is known as the **Sampling (Nyquist-Shannon-Whittaker) Theorem/Formula**

$$\boxed{f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{\Omega}\right) \text{sinc}\left(\Omega\left(t - \frac{n}{\Omega}\right)\right)} \quad (11)$$

Writing out the sinc function

$$f(t) = \sum_{n=-\infty}^{\infty} f(t_n) \frac{\sin(\Omega \pi (t - t_n))}{\Omega \pi (t - t_n)}$$

There is a fair amount of jargon associated with the variables pertaining to this formula. The values, $f(n/\Omega)$, are the sample points. The bandwidth, (Ω) , is a frequency referred to in (Hertz) samples per second. The Ω used here is the same as that used in Briggs and Henson[5], and is the same as the variable p used by Osgood[4]. To recall the magnitude of the bandwidth, refer to the conceptual depiction of the Fourier transform. The bandwidth is the “distance” (in units of frequency) between $\omega_{max} = \Omega/2$ and $\omega_{min} = -\Omega/2$; between ν_{max} and ν_{min} or $2\nu_{max}$ in the Osgood notation[4], $2B$ on the wikipedia page[28]. The variable $\omega_{max} = \Omega/2$, ν_{max} (Osgood[4]), B (wikipedia[28]), is the highest frequency that can be resolved for a given Ω , and is called the Nyquist frequency.

Here there is a discrepancy between the definition of the Nyquist frequency stated by Osgood[4] and Briggs and Henson[5]. For Briggs and Henson the “Nyquist frequency is the highest frequency that can be resolved using a given sample spacing Δx ” (p.97, Ch.3, subsection on Aliasing, [5])

$$\Omega/2$$

For Osgood (p. 366, Ch 6. Sampling and Interpolation) the Nyquist frequency is

$$2\nu_{max}$$

Both statements cannot be true but Briggs and Henson definition is preferable.

The definition of the bandwidth is so important because $1/\Omega$ determines the grid spacing between sample points, Δt for time and Δx for space, denoted as T on the wikipedia page[28]. Counterintuitively, the largest grid spacing for which the sampling formula is true is called the Nyquist (or critical) sampling rate (p.96)[5] - although technically this quantity is not a rate. This leads to ambiguous statements like the Nyquist sampling rate is the “lowest” sampling rate that can reconstruct the signal. This statement makes sense from the stand point of a frequency, but the Nyquist rate is not a frequency, and the meaning of “low” in relation to grid spacing is not unambiguous.

The sampling theorem can be written without explicit reference to the frequency domain. The substitution, $n/\Omega = t_n$, gives the formula as written by Osgood (Ch.5 Sampling and Interpolation)[4] (in terms of Ω instead of p)

$$f(t) = \sum_{n=-\infty}^{\infty} f(t_n) \text{sinc}(\Omega(t - t_n))$$

The bandwidth, Ω , is related to the sample spacing, Δt , as $\Delta t = 1/\Omega$, which gives the wikipedia formula (Δt here is equivalent to T in the wikipedia article)

$$f(t) = \sum_{n=-\infty}^{\infty} f(t_n) \text{sinc}\left(\frac{t - t_n}{\Delta t}\right)$$

Writing out the sinc function gives the Briggs and Henson formula[5] (in time, t , rather than, x , in space)

$$f(t) = \Delta t \sum_{n=-\infty}^{\infty} f(t_n) \frac{\sin(\pi(t - t_n)/\Delta t)}{\pi(t - t_n)}$$

The sampling formula is the source of significant insights into the model, as well as a source of intuition working with the discrete Fourier transform (DFT). Most importantly the signal cannot be both time-limited and band-limited (note the mathematical model and reality disagree, in practice signals are limited in time and in frequency). Recall, the sampling theorem is true for band-limited signals. If the Fourier transform is zero outside the bandwidth (at frequencies beyond the absolute value of the Nyquist frequency)

$$(\mathcal{F}f)(s) \equiv 0 \quad \text{for} \quad |s| \geq \Omega/2$$

then $f(t)$ cannot be limited in time

$$f(t) \not\equiv 0 \quad \text{beyond some } t \text{ (large)}$$

(does not have compact support) goes on forever. To see this, apply the rationale used to derive the sampling formula from the Fourier transform without periodization (Osgood, The Fourier Transform and its Applications, Lect. 18)

$$\begin{aligned} \mathcal{F}f &= \Pi_{\Omega}(\mathcal{F}f) \\ \mathcal{F}^{-1}\mathcal{F}f &= \mathcal{F}^{-1}(\Pi_{\Omega}(\mathcal{F}f)) \\ f(t) &= \Omega \text{sinc}(\Omega t) * f(t_{\text{sampled}}) \end{aligned}$$

The sinc function is never identically 0 beyond a certain point so the right-hand side goes on forever. The sinc will not do this - act as the identity element for convolution - for all signals but it will for band-limited signals.

Likewise, if the signal is limited in time,

$$f(t) \equiv 0 \quad \text{for} \quad |t| \geq q/2 \quad (\text{outside some range})$$

then the Fourier transform cannot be limited in frequency

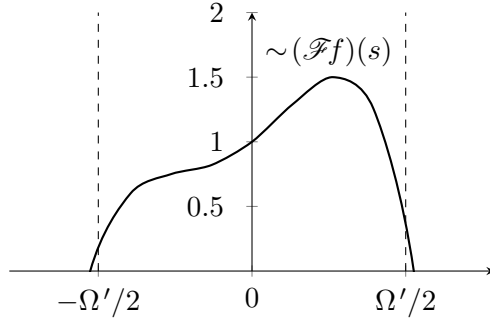
$$(\mathcal{F}f)(s) \not\equiv 0 \quad \text{beyond some } s \text{ (large)}$$

(does not have compact support). (Refer to Osgood text for better proofs).

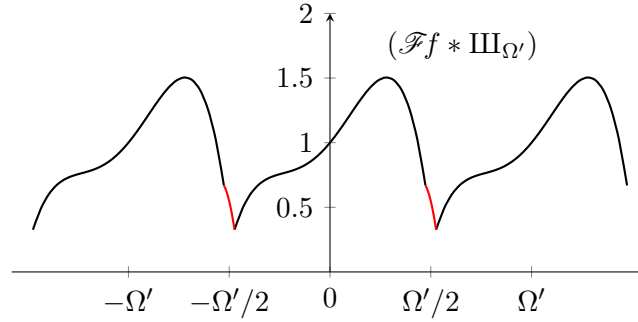
Some considerations to check off the list in what follows: Sampling theorem is still true for a greater bandwidth/ more sampling (sample points spaced closer together). But what happens if the sample rate is too low? Interpolation relies on an infinite number of sample points. In practice, however, only a finite number of sample points is possible, which introduces error such that the sampling theorem becomes an approximation rather than an equality.

8.2 Aliasing and Interpolation

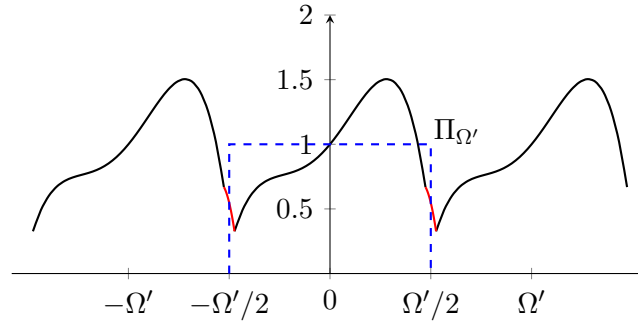
A common problem is the signal is band-limited, but the chosen bandwidth (Ω') is too small.



Performing the periodization $(\mathcal{F}f * \text{III}_{\Omega'})$ shifting by Ω' the shifted copies overlap, and overlapping contributions add up (shown in red).



Cutting off at Ω' using the rect-function (shown by dashed lines in blue) does not recover the original Fourier transform, because $\mathcal{F}f(s) \neq \Pi_{\Omega'}(\mathcal{F}f * \text{III}_{\Omega'})$



Taking the inverse Fourier transform $\mathcal{F}^{-1}(\Pi_{\Omega'}(\mathcal{F}f * \text{III}_{\Omega'}))$ gives

$$g(t) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{\Omega'}\right) \text{sinc}\left(\Omega' \left(t - \frac{k}{\Omega'}\right)\right) \neq f(t)$$

, which does not equal $f(t)$. Nevertheless, $g(t)$ and $f(t)$ agree *at the sample points* $\frac{m}{\Omega'}$ where $m \in \mathbb{Z}$.

$$g\left(\frac{m}{\Omega'}\right) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{\Omega'}\right) \text{sinc}\left(\Omega' \left(\frac{m}{\Omega'} - \frac{k}{\Omega'}\right)\right)$$

Note:

$$\text{sinc}(m - k) = \begin{cases} 1 & m = k \\ 0 & m \neq k \end{cases}, \text{ where } k, m \in \mathbb{Z}$$

In this sense $g(t)$ is an *alias* for $f(t)$.

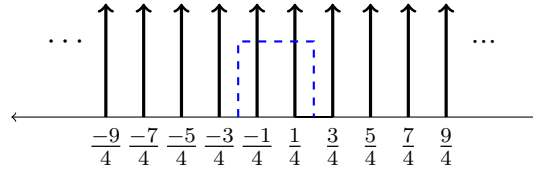
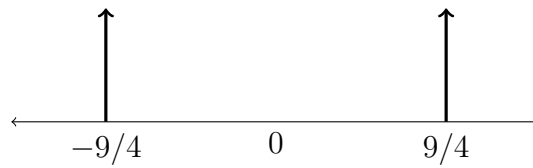


Figure 8: Periodization of the Fourier transform of the signal with the Shaw function III_1 and cut-off with Π_1 (blue).

Example: $f(t) = \cos\left(\frac{9\pi}{2}t\right)$

The Fourier transform is

$$\frac{1}{2}(\delta(s + 9/4) + \delta(s - 9/4))$$



Consider sampling at a rate of $\Omega = 1$. (To sample at a rate of $\Omega = 1$ means to convolve with the Shah function III_1 and to form the cut-off with the indicator function Π_1).

$$\frac{1}{2}(\delta(s + 9/4) + \delta(s - 9/4)) * \sum_{k=-\infty}^{\infty} \delta(x - k)$$

Recall, $\delta_a * \delta_b = \delta_{a+b}$. Only two δ 's in the periodization satisfy $-1/2 < s < 1/2$

$$\Pi_1(\mathcal{F}f * \text{III}_1) = \frac{1}{2}(\delta(s + 1/4) + \delta(s - 1/4))$$

Taking the inverse Fourier transform gives $g(t) = \cos(\frac{\pi}{2}t) \neq f(t) = \cos(\frac{9\pi}{2}t)$, but the two equations agree at the sample points for $t = \pm 1, \pm 2, \pm 3, \dots$ (...check)

9 The DFT

The DFT has two different definitions depending on where you decide to put the factor of $1/N$. In the spirit of No Pedagogy, this amounts to doing everything twice, which sucks but is maybe a price worth paying to satisfy everyone, and to be able to check everything.

9.1 Introductory statements (should be a summary of important practical facts/details)

AHEAD/SIDE: need to be able to display the “spectrogram” for signal analysis. What is multitaper? (Spectral Analysis for Physical Applications Percival and Walden)

Introductory statements: can only take a limited number of samples... real signals are limited in frequency and time...

“A common first step in data or signal analysis is to subtract the mean”(p.9[5].

WANT TO NAIL DOWN THE REASONING FOR THIS FACT! “if a signal is sampled once every Δt units, then a wave with a period less than $2\Delta t$ cannot be resolved accurately”(p.9[5]).

$$\omega_{max} = 1/(2\Delta t) \quad \text{units}^{-1}$$

There are multiple routes for deriving the DFT. Each reveals practical details of the calculation. But first, some preliminaries: conditions on f (the function in theory, the signal in practice), the bandwidth, cyclic complex exponentials, the spatio-temporal and frequency grids.

- only integer modes of a fundamental aka 'one' mode are accessible.
- there is a factor of $1/N$ that distinguishes the DFT from the Fourier transform, and which contributes to alternative definitions for the DFT.
- indices for N even on the symmetric grid range from $-N/2+1$ to $N/2$ (but in truth any consecutive range of N indices will work)
- end points - Briggs and Henson p.93 sec.3.4 “Averaging at Endpoints” [5]
- the DFT is an approximation for the Fourier coefficients
- The relationship between the continuous Fourier transform and the DFT is (p.23[5])

$$F(s_k) \approx LF[k]$$

- missing points: take average of adjacent points

9.2 prelim

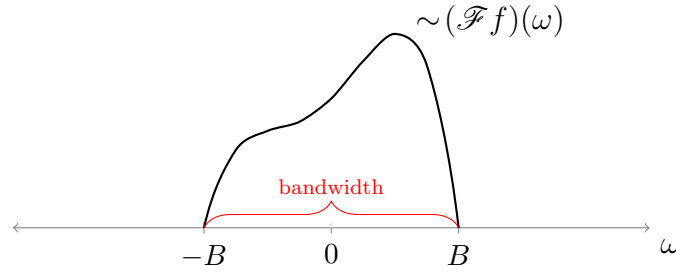
Need to specify conditions on f somewhere, ie:

$$\int f(x) dx < \infty$$

Need to state that the underlying assertion of the DFT involves a misapplication of the (classical) sampling theorem, that a signal cannot be both bandlimited and timelimited - oth terms are one-word, no hyphen, no space.

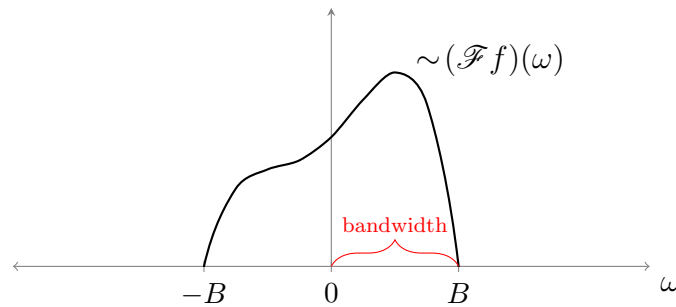
9.3 Review: The Bandwidth

Different communities use a different definition for the bandwidth (is it because they're using continuous or discrete Fourier transform?). The definition I use is based on the following illustration[5]



The benefit of this definition is it directly relates the bandwidth to the sample spacing, $\Delta t = 1/\Omega$.

Sometimes when people refer to the bandwidth they are referring to the depiction below, (although I think this is less standard). Maybe $2B$ is introduced to avoid issues for an even or odd number of samples (N) when indices are introduced for the frequencies.



Essentially, these two definitions for the bandwidth differ by a factor of two, $\Omega = 2B$. According to the definition for the bandwidth (Ω) adopted here, this alternative definition for the bandwidth (B) is equivalent to the Nyquist frequency given by $\Omega/2$.

The definition and concept of the “bandwidth” needs to be understood before diving into the DFT. Its definition gets to the point behind the primary application of the DFT, to reconstruct the signal (from its Fourier series representation). It, furthermore, identifies where limitations in the DFT arise. Lastly, understanding the bandwidth is intimately tied to the intrinsic differences between the DFT and the Fourier transform (... cyclic v. non-cyclic complex exponentials).

One way to understand the concept of the bandwidth is by what it means for a signal to be band-limited. Briggs and Henson are careful to define a periodic function as one which

- only need be piece-wise smooth on the interval $0, L$, $(-L/2, L/2$, however you want to set things up).

Then, the signal is band-limited if

- The Fourier coefficients are zero for frequencies

$$\omega \geq \Omega/2$$

- If f is sampled at least twice per period

$$\Delta x \leq \frac{1}{\Omega}$$

And, the DFT can reproduce the signal exactly.

*** AHEAD: criteria for signals that are “essentially bandlimited” (p.97 Ch.3 [5])

9.4 The Cyclic Complex Exponential

Many differences between the continuous and discrete Fourier transforms stem from the cyclic complex exponential

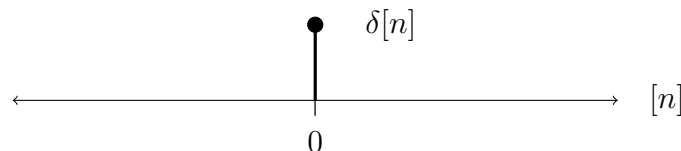
$$e^{-i2\pi sx} \quad \text{v.} \quad e^{-i2\pi kn/N}$$

Briggs and Henson adopt the notation[5] for the cyclic complex exponential

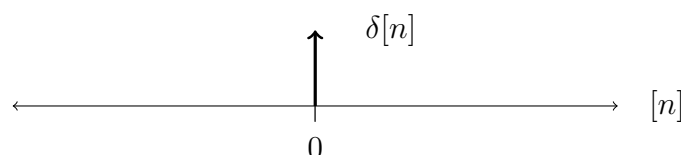
$$\omega_N^{kn} = e^{i2\pi kn/N}$$

COME BACK: also do $\delta[n - m]$ AND SHOW IMAGES TO SHOW THEY ARE THE SAME show symmetry to emphasize that you must make a choice when defining the Kronecker delta for a particular setting, ie: when defining the III function!!! The Kronecker delta

$$\delta[n] = \begin{cases} 1 & \text{for } n = 0 \\ 0 & \text{else} \end{cases}$$

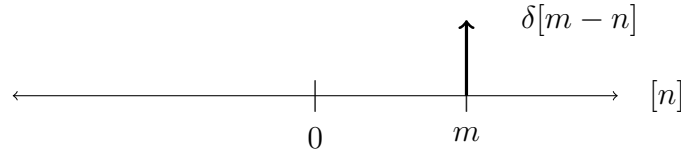


plays the role of the Dirac delta in the discrete context. Sums over the Kronecker delta behave like integrals over the Dirac delta. While the Dirac delta is infinitely high and infinitely narrow at a point along the axis, the Kronecker delta needs only to be equal to one to achieve the desired behavior. In that sense it is fair to represent the Kronecker delta as equal to 1 at a point along the axis. However, in practice the Kronecker delta serves more like a sampling function and sort of conceptually can be thought of as a distribution - that way you are not tempted to evaluate the sum when defining the Shah, III, function as the sum of shifted Kronecker deltas - from this perspective it is still appealing to represent the Kronecker delta as a spike with an arrow. (COME BACK: I must accept that this is a pedagogical stance...and I will have to come back and revise this/ make a decisive choice on how the Kronecker delta should be represented and conceptualized...but right now i cannot say for sure what works best...)

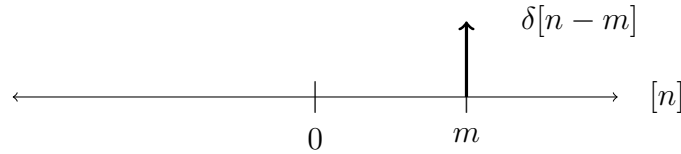


The shifted Kronecker delta

$$\delta[m - n] = \begin{cases} 1 & \text{for } n = m \\ 0 & \text{else} \end{cases}$$



The Kronecker delta is symmetric, $\delta[m - n] = \delta[n - m]$, in either case the function spikes where $n = m$



so that one must make a choice as to how to define the shifted Kronecker delta when implementing it, AHEAD this comes up when defining the discrete Shah function.

The Kronecker delta defined using the cyclic complex exponentials is where the DFT derives its cyclic property (along with the usual orthogonality property). Understanding this cyclic/modular[5] Kronecker delta, (δ_N) , is germane to manipulations involving the DFT.

For the modular Kronecker delta, Briggs and Henson use the notation (p.29[5])

$$\hat{\delta}_N(j - k)$$

I use the notation

$$\delta_N[m - n]$$

The hat over delta is ostensibly to suggest the discrete context. I use square brackets to indicate discreteness when enclosing the arguments of a function. *** (NOW YOU CAN UPDATE NOTATION - no more hat!) ***

Periodicity

The cyclic complex exponential is N -periodic. For $z \in \mathbb{Z}$.

$$\begin{aligned} \omega_N^{n+zN} &= e^{i2\pi(n+zN)/N} = e^{i2\pi n/N} e^{i2\pi zN/N} \\ &= e^{i2\pi n/N} (\cos(2\pi z) + i \sin(2\pi z)) = e^{i2\pi n/N} (1) \\ &= \omega_N^n \end{aligned}$$

where $N \in \mathbb{Z} > 1$ (ie: the number of samples) is the period.

Modular Geometric Series

The geometric series formed by summing the terms of the geometric progression/sequence ($|a| < 1$)...Recall:

$$S_0 = \sum_{n=0}^{N-1} a^n = 1 + a + a^2 + \dots + a^{N-1}$$

$$S_1 = aS_0 = \sum_{n=1}^N a^n = a + a^2 + \dots + a^N$$

$$S_0 - S_1 = S_0(1 - a) = 1 - a^N$$

$$\boxed{\sum_{n=0}^{N-1} a^n = \frac{1 - a^N}{1 - a}}$$

More generally, truncating the sum at the beginning

$$S = \sum_{n=M}^{N-1} a^n = a^M + a^{M+1} + \dots + a^{N-1}$$

$$S' = aS = \sum_{n=M}^{N-1} a^{n+1} = a^{M+1} + a^{M+2} + \dots + a^N$$

$$S - S' = S - aS = S(1 - a) = a^M - a^N$$

$$\boxed{\sum_{n=M}^{N-1} a^n = \frac{a^M - a^N}{1 - a}}$$

For $a = \omega_N$, (a) is N -periodic

$$a[p \pm N] = w_N^{p \pm N} = w_N^p w_N^{\pm N} = w_N^p(1) = a[p]$$

Any list of N consecutive integers, ie: p to $p + N - 1$, will result in the same geometric series

$$S_0 = \sum_{n=p}^{p+N-1} a^n = a^p + a^{p+1} + \dots + a^{p+N-1}$$

$$S_1 = aS_0 = \sum_{n=p}^{p+N-1} a^{n+1} = a^{p+1} + a^{p+2} + \dots + a^{p+N}$$

$$S_0 - S_1 = S_0(1 - a) = a^p - a^{p+N}$$

$$S_0 = \frac{a^p - a^{p+N}}{1 - a}$$

$$\boxed{\sum_{n=p}^{p+N-1} a^n = \frac{a^p - a^{p+N}}{1 - a} \quad (\text{not general, specific to } \textit{cyclic} (a))}$$

Orthogonality The cyclic/modular Kronecker delta $\hat{\delta}_N(m - n)$ (where the subscript N specifies the periodicity) ...the sum of the product of cyclic complex exponentials w_N^{kn} and $w_N^{-\ell n}$ Now, define the cyclic Kronecker delta $\delta_N(m - n)$ and derive the orthogonality of cyclic complex exponentials

$$\sum_{k=0}^{N-1} e^{i2\pi km/N} e^{-i2\pi kn/N} = \sum_{k=0}^{N-1} e^{i2\pi k(m-n)/N} = N\hat{\delta}_N(m - n)$$

$$\boxed{\sum_{k=p}^{p+N-1} \omega_N^{k(m-n)} = \frac{\omega_N^p - \omega_N^{p+N}}{1 - \omega_N} \equiv N\hat{\delta}_N(m - n)} = \begin{cases} N & \text{if } m = n \\ 0 & \text{if } m \neq n \text{ because } \omega_N^p = \omega_N^{p+N} \end{cases}$$

, which can be used to derive the DFT, what else?...

I will adopt a slightly different notation. Rather than use the hat over the delta function to denote discreteness, I will use the square brackets instead of parentheses. The subscript is useful for denoting the cyclic/modular behavior so I keep it.

$$\sum_{k=p}^{p+N-1} \omega_N^{k(m-n)} = \frac{\omega_N^p - \omega_N^{p+N}}{1 - \omega_N} \equiv N \delta_N[m - n] = \begin{cases} N & \text{if } m = n \\ 0 & \text{if } m \neq n \text{ because } \omega_N^p = \omega_N^{p+N} \end{cases}$$

The Dirichlet Kernel

... establishing the Dirichlet kernel (Percival + Walden prob.1.2.c)

$$\begin{aligned} \sum_{t=0}^{N-1} e^{i2\pi st} &= \sum_{t=0}^{N-1} (e^{i2\pi s})^t = \frac{1 - e^{i2\pi sN}}{1 - e^{i2\pi s}} \\ &= \frac{e^{i\pi sN} (e^{-i\pi sN} - e^{i\pi sN})}{e^{i\pi s} (e^{-i\pi s} - e^{i\pi s})} \\ &= \frac{e^{i\pi sN} e^{-i\pi s} (e^{-i\pi sN} - e^{i\pi sN}) (-1/i2)}{(e^{-i\pi s} - e^{i\pi s}) (-1/i2)} \\ &= e^{i\pi s(N-1)} \frac{\sin(N\pi s)}{\sin(\pi s)} \end{aligned}$$

The ratio of $\sin()$'s is continuous at the limit $s \rightarrow z \in \mathbb{Z}$, where s takes on integer values. Applying L'Hospital's rule on the form $\frac{0}{0}$

$$\begin{aligned} \lim_{s \rightarrow z \in \mathbb{Z}} \frac{\sin(N\pi s)}{\sin(\pi s)} &= \frac{\frac{d}{ds} \sin(N\pi s)}{\frac{d}{ds} \sin(\pi s)} \\ &= \frac{\cos(N\pi(z \in \mathbb{Z}))(N\pi)}{\cos(\pi(z \in \mathbb{Z}))(\pi)} \\ &= N \frac{(-1)^{Nz}}{(-1)^z} = N(-1)^{(N-1)z} \end{aligned}$$

gives a modified version of the **Dirichlet kernel** (partly due to the range, from 0 to N , ie: an asymmetric grid),

$$D_N(s) = \begin{cases} \frac{\sin(N\pi s)}{\sin(\pi s)} & \text{for } s \notin \mathbb{Z} \\ N(-1)^{(N-1)s} & \text{for } s \in \mathbb{Z} \end{cases}$$

Percival and Walden (prob.1.2.c, p.17 [15]) play a weird "song and dance" with a factor of N that allows them to write the Dirichlet kernel as

$$D_N(f) = \begin{cases} \frac{\sin(N\pi f)}{N \sin(\pi f)} & \text{for } f \notin \mathbb{Z} \\ (-1)^{(N-1)f} & \text{for } f \in \mathbb{Z} \end{cases}$$

, but which requires them to multiply by a factor of N in the calculation of the geometric progression - defines the DFT - as follows

$$\sum_{t=0}^{N-1} e^{i2\pi ft} = N e^{i(N-1)\pi f} D_N(f)$$

I'll stick with my notation / derivation, it's simpler.

Brad Osgood derivation using a geometric progression on the symmetric grid from $-N$ to N looks more like the standard definition of the Dirichlet kernel (ie: p.189, Rudin, "Principles of Mathematical Analysis, Ed.3") but not exactly

$$\begin{aligned}
\sum_{k=-N}^N e^{i2\pi kt} &= \frac{e^{-i2\pi Nt} - e^{i2\pi(N+1)t}}{1 - e^{i2\pi t}} \\
&= \frac{e^{i\pi t} \left(\exp\left(-i2\pi Nt - i\left(\frac{2}{2}\right)\pi t\right) - \exp\left(i2\pi(N+1)t - i\left(\frac{2}{2}\right)\pi t\right) \right)}{e^{i\pi t}(e^{-i\pi t} - e^{i\pi t})} \\
&= \frac{\exp\left(-i2\pi\left(N + \frac{1}{2}\right)t\right) - \exp\left(i2\pi\left(N + 1 - \frac{1}{2}\right)t\right)}{(e^{-i\pi t} - e^{i\pi t})} \\
&= \frac{\exp\left(-i2\pi\left(N + \frac{1}{2}\right)t\right) - \exp\left(i2\pi\left(N + \frac{1}{2}\right)t\right)}{(e^{-i\pi t} - e^{i\pi t})} \\
&= \frac{(-1) \left(\exp\left(i2\pi\left(N + \frac{1}{2}\right)t\right) - \exp\left(-i2\pi\left(N + \frac{1}{2}\right)t\right) \right)}{(-1)(e^{i\pi t} - e^{-i\pi t})/(i2)} \\
&= \frac{\sin\left(2\pi\left(N + \frac{1}{2}\right)t\right)}{\sin(\pi t)}
\end{aligned}$$

9.5 The Reciprocity Relations

The two variables that you can control when using the DFT are N and L ...

Taking the view that the "reciprocity relations are the keystone of the DFT that holds its entire structure in place." (p.20, [5], they are stated without hesitation and an effort is made to not overlook fine details. These fundamental relationships are derived from the temporal (or spatial) and frequency grids. Briggs and Henson take spatial (temporal) domain $[-A/2, A/2]$ with grid spacing $\Delta x = L/N$ and grid points at $x_n = n\Delta x$; and the frequency domain also consists of N grid points with equal spacing $\Delta\omega$, denoted $\omega_k = k\Delta\omega$, where the indices range from $k = -N/2 + 1 : N/2$. They shirk derivation by "assuming that both the spatial and frequency grids consist of N points" (p.21, [5]). They simply take for granted the congruence between indices, and let $n = -N/2 + 1 : N/2$. Here, the opportunity is taken to clarify any assumptions that may introduce ambiguity in details or in practical considerations of the respective grids.

In practice, it is only possible to obtain a finite number of samples (N). From the perspective of sampling in time it is natural to think of the first sample as occurring at time zero, $t = 0$, and continuing for some length of time L . For temporal data it is also natural to index N samples from 0 to $N - 1$. This can introduce some ambiguity when constructing the frequency grid with the same set of indices. The way to understand both, the motivation for indexing samples from $n = [0 : N - 1]$, and how to adjust the frequency grid to use the same set of indices for $k = [0 : N - 1]$, is by the N -periodicity of the DFT. Over the finite duration (L), which can be subdivided into N equally spaced intervals, the grid spacing (aka "sampling rate" (p.7 [5])) is given by $\Delta t = L/N$.

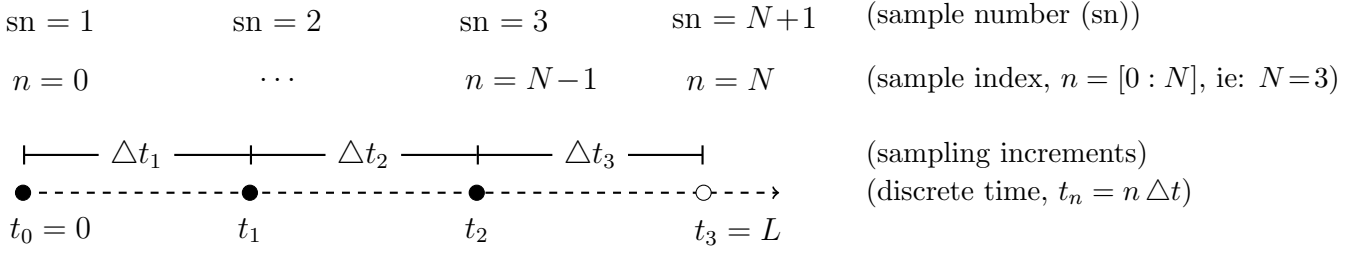


Figure 9: Depiction of a full sampling cycle, from 0 to L , for $N=3$ samples, and the onset of the next cycle. Notice the full extent of the sampling period is not accounted for until the cycle repeats. The $N+1^{\text{th}}$ -sample corresponding to the N^{th} sampling index ($n=N$), the first sample of the next sampling sequence (hollow circle), bridges the gap between the sampling extent at the last sample in a cycle and the complete sampling length. The subtle point I am trying to point out is that, while the total number of sample points and the number of sampling intervals is the same, (N), it is not enough to look at the elapsed time at the last sample in the cycle to determine the length of the sampling duration, rather, $L = N\Delta t$.

It is natural to think of frequencies as being essentially zero or higher. The DFT only permits modes with a complete number of cycles over the sampling length to be used to reconstruct the signal. (This is in stark contrast to the continuous case, where the possible modes are not limited by any boundary and frequencies can take on any value). All available modes are multiples of the fundamental mode, which consists of one period over the full sampling length, depicted below.

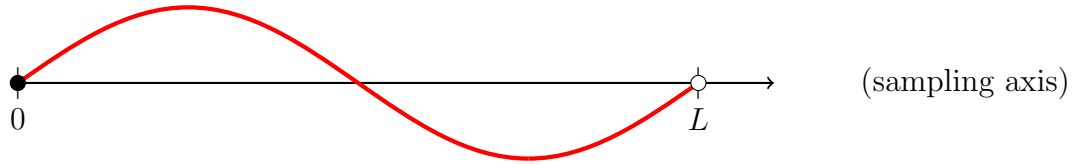


Figure 10: The fundamental mode, aka “one-mode” (red), which determines the grid spacing in the frequency domain $\Delta s = 1/L$. The dots indicate where this frequency is sampled during the present (black) and ensuing (hollow) sampling cycle.

Therefore, the grid spacing in the frequency domain is given by the fundamental frequency as, $\Delta s = 1/L$. BUT where does N come from in determining the number of frequency-samples, and the number of frequency intervals that comprise the full extent of the frequency domain? ??? must somehow get to $\Omega = N\Delta s$??? How to motivate indices starting from 0??? There is clearly some desire to have matching indices for points along the respective grids, but does this occur naturally??? Is there a correspondence between indices used in the respective grids??? (...) Given the bandwidth is equal to $\Omega = N\Delta s = N/L$, the grid spacing in the time domain is related to the bandwidth by $\Delta t = 1/\Omega$.

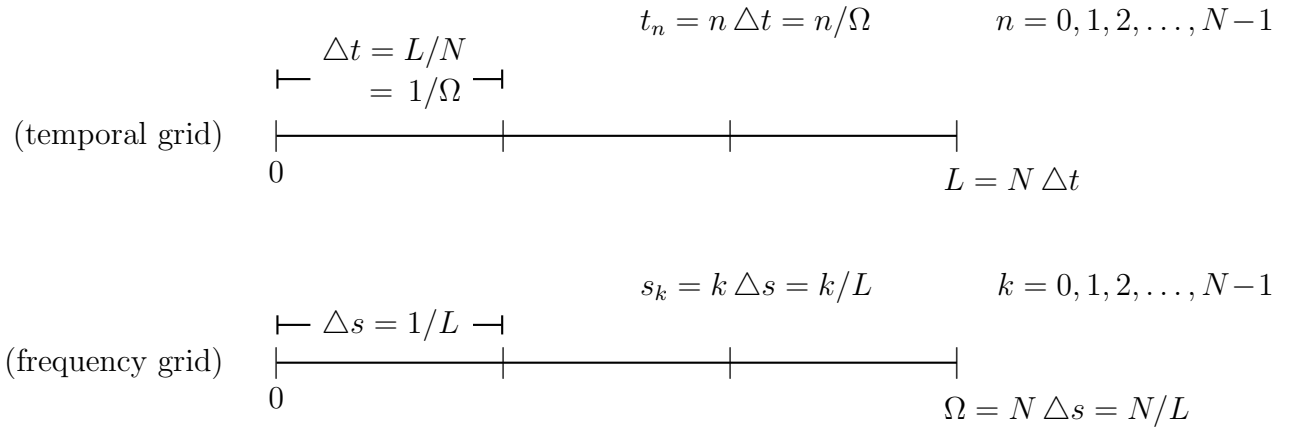


Figure 11:

Does the frequency $s_0 = 0$ correspond to the average value (the DC component, check out Denis Freeman MIT OCW Signals and Systems Lect.14)? How is s_0 handled by subtracting the average value at the outset? Think a little bit more about what is going on with a variable/quantity like this $s_{N-1} = \frac{N-1}{L}$. Based on the temporal (spatial) and frequency grids the reciprocity relations are

$$\Omega L = N \quad (12a)$$

$$\Delta s \Delta t = \frac{1}{N} \quad (12b)$$

Both time and frequency grids intuitively start from zero, as negative values of these quantities are merely mathematical constructs without actual physical meaning. It is trivial to shift the temporal axis to include negative values, ie: for the sake of symmetry. Importantly, and slightly less trivial is how the intuitive conceptual representation for the frequency domain must adjust to accommodate the complex exponential representation of the normal modes. Recall, corresponding positive and negative indices are needed to reconstruct the $|k|^{\text{th}}$ mode. To preserve the bandwidth associated with the old grid when transitioning to the new grid involves a shift along the frequency axis, followed by accompanying indices, depicted below.

“Shift to symmetric grid”

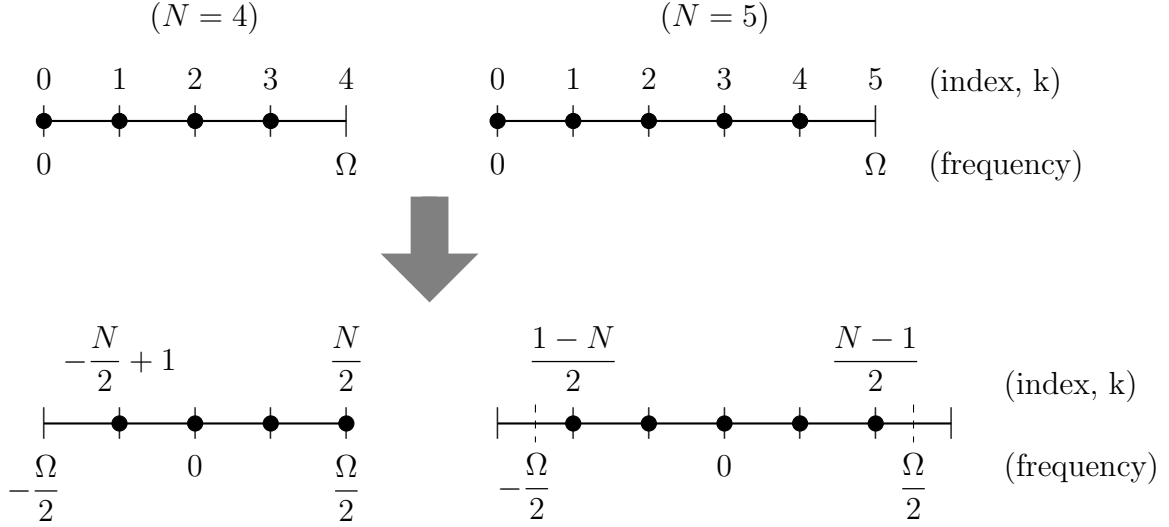


Figure 12: Depiction (for $N = 4$ (left), and $N = 5$ (right) sample points) of the shift to the symmetric frequency grid imposed by the complex exponential representation of the normal modes. (These indices match those presented by Briggs and Henson (p.28 [5]))

In practice, it is common to use indices with the same set of values for both the temporal and frequency grids. This is primarily for convenience, there is no inherent correspondence between locations on the time/position grid and locations on the frequency grid with the same index value. Rather, the relationship is linear; just as all data points along the discrete axis $\{t_n\}$ are needed to determine a single component (s_k) in the frequency spectrum $\{s_k\}$, the entire spectrum is needed to reconstruct each data point, $f(t_n)$, of the signal. To restore the frequency domain so it matches the temporal domain uses the N -periodicity. The aliasing of negative indices to higher frequency positive indices is depicted below. Note, this shows that indices beyond $N/2$ up until N are actually the counter part to indices from 0 to $N/2$ (some vagueness due to N even v. odd).

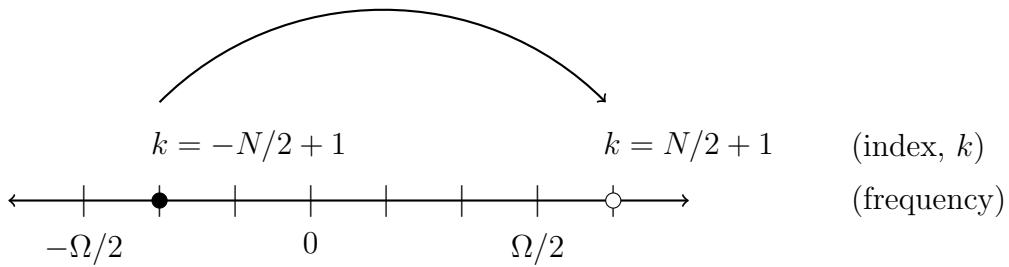


Figure 13: Reconstructing the asymmetric frequency domain from the symmetric frequency grid via aliasing based on the N -periodicity of the DFT.

Observe, for N samples over a spatial/temporal extent L the DFT can at most resolve a maximum frequency of

$$s_{max} = \frac{\Omega}{2}$$

, which corresponds to the index $k = N/2$ whether the frequency grid is constructed as symmetric or asymmetric. This shows that to resolve a mode with k number of wavelengths/periods over L at least $2k$ samples over L are required. To kick a dead horse, consider the case of $N = 1$. For the DFT defined

by $L\Omega = N$ the maximum frequency attainable is $s_{max} = \Omega/2$. The corresponding index is $k_{max} = N/2$; given $N = 1$, then $k_{max} = 1/2$. But k is only allowed to take on integer values.

MOVE THIS!!! You can choose the sampling rate Δt and the number of samples N , then the resolution in frequency Δs is fixed (and *vice versa*). (zero-padding: technique for dealing with this issue)

Note: L and Ω are determined by available memory and the sampling apparatus, respectively...

9.6 Derivation of the DFT via Misapplication of the Sampling Theorem

Assume the signal is both bandlimited and time-limited (although based on the sampling theorem both cannot be true simultaneously) It is perhaps worth noting at this point that the Fourier transform is derived by taking the limits of integration from a single period to infinity but in practice there are limits on the boundary of integration for a specific application. To go from the continuous and analog to the discrete and digital requires

1. A discrete approximation to the signal $f(t)$.
2. A discrete approximation to the Fourier transform $(\mathcal{F}f)(s)$.
3. A way to go recover the discrete signal from the DFT that mimics the continuous case.

To define the sampled/discrete form use the Shaw function. The sampled form of $f(t)$ for N samples is

$$\begin{aligned} f(t)_{sampled} &= f(t) \sum_{n=0}^{N-1} \delta(t - t_n) \\ &= \sum_n f(t) \delta(t - t_n) = \sum_n f(t_n) \delta(t - t_n) \\ &= \sum_n f_n \delta(t - t_n) \end{aligned}$$

This is a reasonable approximation to $f(t)$ in that (based on the sampling theorem) interpolation at the recorded values should reconstruct the signal exactly. Note the choice to keep the Shaw function in the formula, essentially because in principle the Shaw function could have been applied to the product of multiple functions not just $f(t)$ alone. To get rid of the Shaw function it needs to be integrated out.

Taking the Fourier transform of the sampled form of $f(t)$

$$(\mathcal{F}f_{sampled})(s) = \int e^{-i2\pi st} \sum_n f_n \delta(t - t_n) dt$$

Given the Fourier transform is a linear operator

$$\begin{aligned} &= \sum_n f_n \int e^{-i2\pi st} \delta(t - t_n) dt \\ &= \sum_n f(t_n) e^{-i2\pi st_n} \end{aligned}$$

, which is still a function of a continuous variable, s .

To get an approximate discrete version of $\mathcal{F}f_{\text{sampled}}$ take N samples spaced $1/L$ apart over an interval of length Ω in the frequency domain, $s_0 = 0, s_1 = \frac{1}{L}, s_2 = \frac{2}{L}, \dots, s_{N-1} = \frac{N-1}{L}$. The sampled form of the Fourier transform is the sampled form of the Fourier transform (in frequency) of the sampled form of $f(t)$ (in time)

$$\begin{aligned}
(\mathcal{F}f_{\text{sampled}})(s)_{\text{sampled}} &= (\mathcal{F}f_{\text{sampled}})\left(\sum_{k=0}^{N-1} \delta(s - s_k)\right) \\
&= \left(\sum_{n=0}^{N-1} f_n e^{-2\pi i s t_n}\right) \left(\sum_{k=0}^{N-1} \delta(s - s_k)\right) \\
&= \sum_k \sum_n f_n e^{-2\pi i s t_n} \delta(s - s_k) \\
&= \sum_k \sum_n f_n e^{-2\pi i s_k t_n} \delta(s - s_k)
\end{aligned}$$

The explicit reference to time and frequency maintains the connection to the continuous picture. The last step to define the DFT is to eliminate any reference to the continuous variables, t and s . The sample points in the time domain are

$$t_n = \frac{n}{\Omega}$$

The sample points in the frequency domain are

$$s_k = \frac{k}{L}$$

The product of s_k and t_n is

$$s_k t_n = \frac{kn}{\Omega L} = \frac{kn}{N}$$

because $\Omega L = N$, leaving only discrete indices, k and n .

Take the inverse Fourier transform to recover the discrete signal

$$\begin{aligned}
\mathcal{F}^{-1}((\mathcal{F}f_{\text{sampled}})(s)_{\text{sampled}}) &= \int e^{i2\pi s t} \sum_{k=0}^{N-1} \sum_{n=0}^{N-1} f_n e^{-i2\pi s t_n} \delta(s - s_k) ds \\
&= \sum_k \sum_n f_n e^{-i2\pi s_k t_n} \int e^{i2\pi s t} \delta(s - s_k) ds \\
&= \sum_k \sum_n e^{-i2\pi s_k t_n} e^{i2\pi s_k t} = \sum_k \sum_n f_n e^{i2\pi s_k (t - t_n)} \\
&= \sum_k \sum_n f_n \exp\left(i2\pi \frac{k}{L} \left(\frac{mL}{N} - \frac{nL}{N}\right)\right) \\
&= \sum_k \sum_n f_n e^{i2\pi k(m-n)/N} = \sum_n f_n \underbrace{\sum_k e^{i2\pi k(m-n)/N}}_{N\hat{\delta}_N(m-n)} \\
&= N \sum_n f_n \hat{\delta}_N(m-n) = f_m \quad (m \text{ equals the index } n \text{ at the data points})
\end{aligned}$$

The question remains, how should the factor of N be distributed in defining the DFT and IDFT? Both options are in use. The form used by Briggs & Henson, Bracewell, Weaver, etc.[5, 9, 24] is

$$\mathcal{F}f = (\mathcal{F}(f(\{x\}_n))(s_k) = F[k] = \frac{1}{N} \sum_{n=0}^{N-1} f_n e^{-i2\pi kn/N} \quad k = 0, 1, 2, \dots, N-1 \quad (13a)$$

$$\mathcal{F}^{-1}F = (\mathcal{F}^{-1}(F(\{s\}_k))(x_n) = f[n] = \sum_{k=0}^{N-1} F_k e^{i2\pi kn/N} \quad n = 0, 1, 2, \dots, N-1 \quad (13b)$$

, where square brackets are used to connote indices for discrete transforms, curly braces are used for sequence notation (as distinguished from the continuous variable), and the symbol \mathcal{F} denotes the DFT.

The alternative form, used by Osgood[4], Brigham[25], Oppenheim & Schafer[12], Percival & Warden[15], etc. is

$$F[k] = \sum_{n=0}^{N-1} f_n e^{-i2\pi kn/N} \quad k = 0, 1, 2, \dots, N-1 \quad (14a)$$

$$f[n] = \frac{1}{N} \sum_{k=0}^{N-1} F_k e^{i2\pi kn/N} \quad n = 0, 1, 2, \dots, N-1 \quad (14b)$$

The $1/N$ factor is another feature of the DFT that distinguishes it from the continuous Fourier transform, and is an important practical detail. I do not have a mathematical explanation for why the $1/N$ factor *should* be associated with the IDFT. But there is a significant computational motivation that serves as justification. In practice, the DFT will probably be calculated repeatedly in search of a good estimate, or continuously as data is streaming. The additional computational cost of including this $1/N$ factor up front in the calculation is enormous, while providing little additional information. In principle, the IDFT is only calculated once for the purpose of signal reconstruction. The additional computational cost imposed at this final stage is insignificant by comparison. **So this is the one you should use!**

A pragmatic rationale for associating the factor of $1/N$ with the IDFT is that if *you forgot* to include it in your calculation of the DFT, it's totally fine you can reach for it when you need it.

9.7 Derivation of the DFT via Numerical Approximation (end points)

...Important practical details for implementing the DFT can be seen by looking at other derivations of the DFT... Here we will explore:

- how to handle data at the end-points?
- what does the DFT really calculate?

The DFT can be derived as a numerical approximation to the Fourier transform.

$$\hat{f}(s) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi s x} dx$$

Assume the function/signal is zero outside the finite sampling domain.

$$\hat{f}(s_k) = \int_L f(x) e^{-i2\pi s_k x} dx$$

The area of a trapazoid based on the average of the two parallel sides is

$$\left(\frac{g(x) + g(x+1)}{2} \right) \Delta x$$

such that the trapazoid rule is

$$\int_L g(x) dx \approx \frac{\Delta x}{2} \left(g(x_0) + (2) \sum_{n=1}^{N-1} g(x_n) + g(x_N) \right)$$

Assume (this is an important practical detail)

$$\boxed{g(x_0) = g(x_N)}$$

(this is reasonable when the periodic extension of the signal is often invoked but that is not the case here as the procedure clearly employs zero-padding); for an asymmetric interval $g(0) = g(L)$; for an symmetric interval $g(-L/2) = g(L/2)$. Then

$$= \Delta x \sum_n g(x_n)$$

Then

$$\hat{f}(s_k) \approx \frac{L}{N} \sum_n f(x_n) e^{-i2\pi s_k x_n}$$

Using the reciprocity relation, $\Delta s^{-1} \Delta x^{-1} = 1/N$, it follows that the set of points $s_k x_n = kn/N$, which allows the substitution

$$(\mathcal{F}f)(s)_k = F(s_k) = \hat{f}(s_k) \approx L \underbrace{\frac{1}{N} \sum_n f(x_n) e^{-i2\pi kn/N}}_{F[k]}$$

This expression mirrors the continuous case (see sections on Fourier coefficients and the derivation of the Fourier transform)

$$F(s_k) = Lc_k$$

suggesting that the DFT is an approximation for the Fourier coefficients (an important practical detail).
The relationship between the continuous Fourier transform and the DFT is then (p.23[5])

$$\boxed{F(s_k) \approx LF[k]}$$

9.8 Derivation of the DFT via Approximation to the Fourier Coefficients (series convergence and discontinuities)

Recall, the Fourier coefficients are obtained from

$$c_k = \frac{1}{L} \int_L f(x) e^{-i2\pi kx} dx$$

... The assumption

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kx}$$

requires that the series converges. Convergence of Fourier series is deep ... Briggs + Henson are content with the definition

def. **piecewise continuous** on $[a, b]$ There are finite number of points where x_i is either not defined or discontinuous but the limits

$$f(x_i^+) = \lim_{h \rightarrow 0^+} f(x_i + h)$$

$$f(x_i^-) = \lim_{h \rightarrow 0^-} f(x_i + h)$$

exist and are finite.

def. **piecewise smooth** on $[a, b]$ f is piecewise continuous, and

$$f'(x_i^+) = \lim_{h \rightarrow 0^+} \frac{f(x_i + h) - f(x_i)}{h}$$

$$f'(x_i^-) = \lim_{h \rightarrow 0^-} \frac{f(x_i + h) - f(x_i)}{h}$$

Theorem: Convergence of Fourier Series (ref.158[5])

If f is L -periodic and piecewise smooth then

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kx}$$

converges (pointwise) to

$$\frac{f(x^+) + f(x^-)}{2}$$

for every x , which reveals an important practical detail, points of discontinuity take the average value of the points to the left and to the right.

If f is undefined outside the sampling interval or is aperiodic over L , the Fourier series converges to the periodic extension. A subtle but important consequence of this periodic extension is the practical detail that the end points must be equal, ie: if $f(x_0 = 0) \neq f(x_N = L)$ then one value, eg: the latter, end point can be set such that (p.38 [5])

$$f(x_N) = \frac{f(x_0^+) + f(x_N^-)}{2}$$

For N even and ranging from 0 to $N - 1$ there is no data point for $f(x_N = L)$ to terminate the sequence. The first point in the periodic extension is used as the $(N+1)$ -data point needed to compute the N -point DFT. To initial data point, at index $n = 0$, and the last data point, at index $n = N$, are both supplied manually as

$$f[N] = \frac{f[0] + f[N - 1]}{2}$$

to ensure continuity of the periodic extension. For N even and N ranges from $-N/2 + 1$ to $N/2$ there is no data point for $f(x_{-N/2} = -L/2)$, The first data point, at index $n = -N/2$, is supplied manually as

$$f[-N/2] = \frac{f[-N/2 + 1] + f[N/2]}{2}$$

What to do in practice for N odd?...

The DFT can be derived by writing the integral for the Fourier coefficients as a series approximation

$$c_k = \frac{1}{L} \int_L f(x) e^{-i2\pi sx} dx \approx \frac{1}{L} \sum_{n=0}^{N-1} f_n e^{-i2\pi kn/N} \Delta x$$

$$= \frac{1}{L} \sum_n f_n \omega_N^{-kn} \left(\frac{L}{N} \right)$$

$$= \frac{1}{N} \sum_n f_n \omega_N^{-kn} = F[k] = (\mathfrak{F}(f(\{x\}_n)))(s_k)$$

, which shows that the DFT is an approximation to the Fourier coefficients

$$c_k \approx F[k]$$

9.9 “The DFT from Trigonometric Approximation” [5]

Idea: use trigonometric polynomial, a linear combination of sines and cosines, to fit data. Model data using trigonometric polynomial

$$\psi(x) = \sum c_k e^{i2\pi kx/L}$$

(This instance it is more convenient to work with an odd no. samples using the symmetric representation...) The (discrete) least square error

$$E = \sum |f_n - \psi(x_n)|^2$$

is non-negative real-valued function.

(...) Minimize the error by writing the “normal equations” (prob.26) (...) Write the DFT based on the coefficients...(? I thought the formula for the FT and the FT coefficients are different by a factor of $1/L$!?!)...(another way prob.28)

(...) Parseval’s relation for the DFT

$$\sum_n |f_n|^2 = N \sum_n |c_n|^2$$

*** In practice this serves as a useful way to check your work (without having to reconstruct the signal).

9.10 The DFT Matrix

The DFT for a specific k is the inner product of the discrete signal with the corresponding row of the DFT matrix. The elements of the DFT matrix are $\omega_N^{-kn} = e^{-i2\pi kn/N}$, where the k -index is associated with the row index, and the n -index is associated with the column index. For the N -point DFT indexed from 0 to $N - 1$

$$F[k] = \frac{1}{N} \sum_{n=0}^{N-1} f_n \omega_N^{-kn}$$

has a DFT matrix W_{kn}

$$W_{kn} = \frac{1}{N} \begin{bmatrix} \omega_N^{-(0)(0)} & \omega_N^{-(0)(1)} & \dots & \omega_N^{-(0)(n)} & \dots & \omega_N^{-(0)(N-1)} \\ \omega_N^{-(1)(0)} & \omega_N^{-(1)(1)} & \dots & \omega_N^{-(1)(n)} & \dots & \omega_N^{-(1)(N-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \omega_N^{-(k)(0)} & \omega_N^{-(k)(1)} & \dots & \omega_N^{-(k)(n)} & \dots & \omega_N^{-(k)(N-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \omega_N^{-(N-1)(0)} & \omega_N^{-(N-1)(1)} & \dots & \omega_N^{-(N-1)(n)} & \dots & \omega_N^{-(N-1)(N-1)} \end{bmatrix}$$

where k is the row and n specifies the column.

10 Properties of the DFT

10.1 DFT Duality

The definitions of the DFT and the IDFT operators

$$\mathcal{F} = \frac{1}{N} \sum_n w_N^{-kn} \quad \text{and} \quad \mathcal{F}^{-1} = \sum_k w_N^{kn}$$

lead to the duality statements

$$\boxed{\mathcal{F}(f^{(-)}) = \frac{1}{N} (\mathcal{F}^{-1} f)} \quad (15a)$$

$$\boxed{\mathcal{F}^{-1}(f^{(-)}) = (N) (\mathcal{F} f)} \quad (15b)$$

The definitions of the DFT and the IDFT operators

$$\mathcal{F} = \sum_n w_N^{-kn} \quad \text{and} \quad \mathcal{F}^{-1} = \frac{1}{N} \sum_k w_N^{kn}$$

$$\boxed{\mathcal{F} f^{(-)} = (N) (\mathcal{F}^{-1} f)} \quad (16a)$$

$$\boxed{\mathcal{F}^{-1}(f^{(-)}) = \frac{1}{N} (\mathcal{F} f)} \quad (16b)$$

Duality begins with the reverse signal, where the signal starts with the last index and continues counting down to the starting index, which correspond to the last and first sampling points, respectively. The N -element index $[n]$ proceeds from 0 up to $N - 1$ as

$$[n] : 0, 1, 2, \dots, N - 1$$

The $[N - n]$ -index counts down from N to 1 as

$$[N - n] : N, N - 1, N - 2, \dots, 1$$

Then the reverse signal can be written as

$$f[n]^{(-)} = f[N - n]$$

The DFT of the reverse signal is

$$\mathcal{F}(f[n]^{(-)}) = \frac{1}{N} \sum_n e^{-i2\pi kn/N} f_n^{(-)} = \frac{1}{N} \sum_n e^{-i2\pi kn/N} f[N - n] = \frac{1}{N} \sum_n e^{-i2\pi kn/N} f[-n]$$

by the N -periodicity of the DFT. (Recall:

$$\begin{aligned} \mathcal{F}(f[0]) &= \mathcal{F}(f[\pm N]) &= \mathcal{F}(f[\pm 2N]) &= \mathcal{F}(f[\pm 3N]) &\dots \\ \mathcal{F}(f[1]) &= \mathcal{F}(f[\pm N + 1]) &= \mathcal{F}(f[\pm 2N + 1]) &= \mathcal{F}(f[\pm 3N + 1]) &\dots \\ \mathcal{F}(f[-1]) &= \mathcal{F}(f[\pm N - 1]) &= \mathcal{F}(f[\pm 2N - 1]) &= \mathcal{F}(f[\pm 3N - 1]) &\dots \\ &\vdots \\ \mathcal{F}(f[\pm n]) &= \mathcal{F}(f[\pm N \pm n]) &= \mathcal{F}(f[\pm 2N \pm n]) &= \mathcal{F}(f[\pm 3N \pm n]) &\dots \end{aligned}$$

from the cyclic behavior of the complex exponentials, $w_N^{k(n \pm N)} = w_N^{kn}$).

Substituting (n) for $(-n)$ is straight forward enough but let's be pedantic here, if there is any place to be long winded if explicit it is with the earlier simpler material. Let $m = -n$. The $-n$ index goes from

$$\begin{aligned} [m] &= [-n] : 0, -1, -2, -3, \dots, -(N-1) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} e^{-i2\pi k(n)/N} f[-n] = \frac{1}{N} \sum_{m=0}^{-(N-1)} e^{i2\pi k(m)/N} f[m] \end{aligned}$$

... the cyclic property of w_N means any consecutive N integers will do...

$$\frac{1}{N} \sum_{m=0}^{-(N-1)} e^{i2\pi k(m)/N} f[m] = \frac{1}{N} \sum_{n=0}^{N-1} e^{i2\pi k(n)/N} f[n]$$

in the argument to f and in the complex exponentials gives

$$= \frac{1}{N} \sum_n e^{-i2\pi k(n)/N} f[(-n)] = \frac{1}{N} \sum_n e^{-i2\pi k(-n)/N} f[(n)] = \frac{1}{N} \sum_n e^{i2\pi kn/N} f[n]$$

Here it helps to be flexible with our interpretation of the indices/variables, and say that the DFT is essentially the sum of cyclic complex exponentials with a negative exponent while the IDFT is the sum of cyclic complex exponentials with a possitive exponent. Then the DFT and the IDFT applied to the same function ignoring the meaning or physical significance of the summation indices are

$$\mathcal{F}(f[n]) = \frac{1}{N} \sum_n e^{-i2\pi kn/N} f[n] \quad \text{and} \quad \mathcal{F}^{-1}(f[k]) = \sum_k e^{i2\pi kn/N} f[k]$$

From this perspective, the above equation shows the DFT applied to the reverse signal is equal to $1/N$ times the IDFT

$$\boxed{\mathcal{F}(f^{(-)}) = \frac{1}{N} \mathcal{F}^{-1}(f)}$$

, albeit not with the expected index.

Similarly, the IDFT of the reverse signal

$$\mathcal{F}^{-1}(f^{(-)}) = \sum_k f[k]^{(-)} w_N^{kn}$$

By the N -periodicity of the IDFT

$$= \sum_k f[N-k] w_N^{kn} = \sum_k f[-k] w_N^{kn}$$

Substituting k for $-k$

$$= \sum_k f[k] w_N^{-kn} = \frac{1}{N} \mathcal{F}(f[k])$$

where the k -index actually indexes over the function values via the spatial/temporal grid.

$$\boxed{\mathcal{F}^{-1}(f^{(-)}) = \frac{1}{N} \mathcal{F}f}$$

Now do this for the alternative versions of the DFT and IDFT...

10.2 The DFT & Parseval's Theorem

Do discrete form of Parseval's theorem for both forms of the DFT equations...where to put the $\frac{1}{N}$ factor?...

COME BACK: (this is a practical and conceptual check) Taken from Numerical Methods p.608 s.12.1.2: If the DFT is defined as

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{-i2\pi kn/N}$$

then the discrete form the Parseval's theorem is

$$\sum_{k=0}^{N-1} |h_k|^2 = \sum_{n=0}^{N-1} |H_n|^2$$

10.3 The DFT & Discrete Convolution

Using the definitions for DFT and IDFT

$$\mathcal{F}f = \sum_n f[n] w_N^{-kn} \quad \text{and} \quad \mathcal{F}^{-1}F = \frac{1}{N} \sum_k F[k] w_N^{kn}$$

convolution in the spatial/time domain is defined as

$$(f * g)[m] \equiv \sum_n f[n] g[m - n] \quad (17)$$

where $[m]$ is a dummy index that references the $[n]$ grid. And the convolution theorem consists of the expressions

$$\begin{aligned} (\mathcal{F}f) (\mathcal{F}g) &= \mathcal{F}(f * g) & (18a) \\ (\mathcal{F}^{-1}(f)) (\mathcal{F}^{-1}(g)) &= \frac{1}{N} \mathcal{F}^{-1}(f * g) & (18b) \\ \mathcal{F}(f) * \mathcal{F}(g) &= N \mathcal{F}(f g) & (18c) \\ (\mathcal{F}^{-1}f) * (\mathcal{F}^{-1}g) &= \mathcal{F}^{-1}(f g) & (18d) \end{aligned}$$

For this topic it may be useful to explore the different versions of the DFT (where to put the factor of $1/N$) to try and understand why various authors adopt one convention over the other.

In the continuous case, the definition of the convolution was prompted by the question, "Given (f) and (g) , is there an (h) such that"

$$FG = \mathcal{F}(f) \mathcal{F}(g) = \mathcal{F}(h)$$

For the DFT this becomes

$$F G = \mathcal{F}(f) \mathcal{F}(g) \stackrel{?}{=} \mathcal{F}(h)$$

To find (h) take the IDFT

$$\mathcal{F}^{-1}(\mathcal{F}(f) \mathcal{F}(g)) = \mathcal{F}^{-1} \mathcal{F}(h) = h$$

The functions F and G are both functions of the k -index, which refers to the frequency grid (in general, F and G are independent, and need not refer to the same frequency/ have the same index, ie: $F[k]G[\ell]$).

$$\mathcal{F}f \mathcal{F}g = F[k]G[k]$$

Separate indices are needed to destinguish each DFT's series summation over the spatial/temporal grid

$$= \sum_n f_n e^{-i2\pi kn/N} \sum_{n'} g_{n'} e^{-i2\pi kn'/N} = \sum_n f_n w_N^{-kn} \sum_{n'} g_{n'} w_N^{-kn'}$$

Writing out the multiplication of the series

$$\begin{aligned} &= (f_0 w^{-k(0)} + f_1 w^{-k(1)} + \dots + f_{N-1} w^{-k(N-1)}) (g_0 w^{-k(0)} + g_1 w^{-k(1)} + \dots + g_{N-1} w^{-k(N-1)}) \\ &= f_0 w^{-k(0)} (g_0 w^{-k(0)} + g_1 w^{-k(1)} + \dots + g_{N-1} w^{-k(N-1)}) \\ &\quad + f_1 w^{-k(1)} (g_0 w^{-k(0)} + g_1 w^{-k(1)} + \dots + g_{N-1} w^{-k(N-1)}) \\ &\quad \vdots \\ &\quad + f_{N-1} w^{-k(N-1)} (g_0 w^{-k(0)} + g_1 w^{-k(1)} + \dots + g_{N-1} w^{-k(N-1)}) \end{aligned}$$

shows it can be thought of as

$$= \sum_n \left(e^{-i2\pi kn/N} f_n \sum_{n'} e^{-i2\pi kn'/N} g_{n'} \right) \stackrel{\text{and}}{=} \sum_{n'} \left(e^{-i2\pi kn'/N} g_{n'} \sum_n e^{-i2\pi kn/N} f_n \right)$$

Grouping the terms inside the parentheses

$$= \sum_n \left(\left(e^{-i2\pi kn/N} f_n \right) \left(\sum_{n'} e^{-i2\pi kn'/N} g_{n'} \right) \right)$$

suggests, by associativity, the rearrangement

$$= \sum_n \sum_{n'} f_n g_{n'} e^{-i2\pi kn/N} e^{-i2\pi kn'/N}$$

The double sum produces all $N \times N$ combinations of the indices in constructing the series

$$\begin{aligned} &= f_0 g_0 w^{-k(0)} w^{-k(0)} + f_0 g_1 w^{-k(0)} w^{-k(1)} + \dots + f_0 g_{N-1} w^{-k(0)} w^{-k(N-1)} \\ &\quad + f_1 g_0 w^{-k(1)} w^{-k(0)} + f_1 g_1 w^{-k(1)} w^{-k(1)} + \dots + f_1 g_{N-1} w^{-k(1)} w^{-k(N-1)} \\ &\quad \vdots \\ &\quad + f_{N-1} g_0 w^{-k(N-1)} w^{-k(0)} + f_{N-1} g_1 w^{-k(N-1)} w^{-k(1)} + \dots + f_{N-1} g_{N-1} w^{-k(N-1)} w^{-k(N-1)} \end{aligned}$$

To uncover (h) take the IDFT

$$\begin{aligned} \mathcal{F}^{-1}(\mathcal{F}f \mathcal{F}g)[m] &= \mathcal{F}^{-1}(F[k] G[k])[m] \\ &= \mathcal{F}^{-1} \left(\left(\sum_n f_n w_N^{-kn} \right) \left(\sum_{n'} g_{n'} w_N^{-kn'} \right) \right) [m] \\ &= \frac{1}{N} \sum_k w_N^{+km} \left(\left(\sum_n f_n w_N^{-kn} \right) \left(\sum_{n'} g_{n'} w_N^{-kn'} \right) \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{N} \left(w^{+(0)m} + w^{+(1)m} + \dots + w^{+(N-1)m} \right) \\
&\quad \left(f_0 g_0 w^{-k(0)} w^{-k(0)} + f_0 g_1 w^{-k(0)} w^{-k(1)} + \dots + f_0 g_{N-1} w^{-k(0)} w^{-k(N-1)} \right. \\
&\quad \left. + f_1 g_0 w^{-k(1)} w^{-k(0)} + f_1 g_1 w^{-k(1)} w^{-k(1)} + \dots + f_1 g_{N-1} w^{-k(1)} w^{-k(N-1)} \right. \\
&\quad \vdots \\
&\quad \left. + f_{N-1} g_0 w^{-k(N-1)} w^{-k(0)} + f_{N-1} g_1 w^{-k(N-1)} w^{-k(1)} + \dots + f_{N-1} g_{N-1} w^{-k(N-1)} w^{-k(N-1)} \right) \\
&= \frac{w^{+(0)m}}{N} \left(f_0 g_0 w^{-k(0)} w^{-k(0)} + \dots + f_{N-1} g_{N-1} w^{-k(N-1)} w^{-k(N-1)} \right) \\
&\quad + \frac{w^{+(1)m}}{N} \left(f_0 g_0 w^{-k(0)} w^{-k(0)} + \dots + f_{N-1} g_{N-1} w^{-k(N-1)} w^{-k(N-1)} \right) \\
&\quad \vdots \\
&\quad + \frac{w^{+(N-1)m}}{N} \left(f_0 g_0 w^{-k(0)} w^{-k(0)} + \dots + f_{N-1} g_{N-1} w^{-k(N-1)} w^{-k(N-1)} \right) \\
&= \frac{1}{N} \sum_k w_N^{+km} \left(\sum_n \sum_{n'} f_n g_{n'} w_N^{-kn} w_N^{-kn'} \right) \\
&= \frac{1}{N} \left(\sum_n \sum_{n'} f_n g_{n'} w_N^{-k(n+n')} \right) \sum_k w_N^{+km} \\
&= \frac{1}{N} \sum_n \sum_{n'} f_n g_{n'} \sum_k w_N^{+km} w_N^{-k(n+n')} \\
&= \frac{1}{N} \sum_n \sum_{n'} f_n g_{n'} \sum_k w_N^{k(m-n-n')} \\
&= \frac{1}{N} \sum_n \sum_{n'} f_n g_{n'} (N \hat{\delta}_N(m-n-n')) \\
&= \sum_n \sum_{n'} f_n g_{n'} \hat{\delta}_N(m-n-n')
\end{aligned}$$

Performing the sum over n'

$$\begin{aligned}
&= \sum_n f[n] \sum_{n'} g[n'] \hat{\delta}_N(m-n-n') \\
&= \sum_n f[n] g[m-n]
\end{aligned}$$

This derivation leads to the definition of the discrete convolution

$$\boxed{(f * g)[m] \equiv \sum_n f[n] g[m-n]} \tag{19}$$

and shows that the product of DFTs is the DFT of the convolution

$$\boxed{(\mathcal{F}f)(\mathcal{F}g) = \mathcal{F}(f * g)} \tag{20}$$

The IDFT version of this procedure is

$$\mathcal{F}(\mathcal{F}^{-1}(f)[n] \mathcal{F}^{-1}(g)[n])[\ell] = \mathcal{F}^{-1}(h)$$

$$\begin{aligned}
&= \sum_n e^{-i2\pi\ell n/N} \left(\left(\frac{1}{N} \sum_k f_k e^{+i2\pi kn/N} \right) \left(\frac{1}{N} \sum_{k'} g_{k'} e^{+i2\pi k' n/N} \right) \right) \\
&= \sum_n w_N^{-\ell n} \left(\frac{1}{N^2} \sum_k \sum_{k'} f_k g_{k'} w_N^{(k+k')n} \right) \\
&= \frac{1}{N^2} \sum_k \sum_{k'} f_k g_{k'} w_N^{(k+k')n} \sum_n w_N^{-\ell n} \\
&= \frac{1}{N^2} \sum_k \sum_{k'} f_k g_{k'} \sum_n w_N^{(k+k'-\ell)n} \\
&= \frac{1}{N^2} \sum_k \sum_{k'} f_k g_{k'} N \hat{\delta}_N(k+k'-\ell)
\end{aligned}$$

Performing the summation over k'

$$= \frac{N}{N^2} \sum_k f[k] \sum_{k'} g[k'] \hat{\delta}_N(k+k'-\ell)$$

selects the value, of g , when $k' = \ell - k$

$$= \frac{1}{N} \sum_k f[k] g[\ell - k] = \frac{1}{N} (f * g)$$

$$\boxed{\mathcal{F}^{-1}(f) \mathcal{F}^{-1}(g) = \frac{1}{N} \mathcal{F}^{-1}(f * g)} \quad (21)$$

Using the definition of the convolution, the convolution of DFTs is

$$\begin{aligned}
(\mathcal{F}f)[k] * (\mathcal{F}g)[k] &= F[k] * G[\ell - k] \\
&= \sum_k F[k] G[\ell - k] \\
&= \sum_k \left(\left(\sum_n f_n w_N^{-kn} \right) \left(\sum_{n'} g_{n'} w_N^{-(\ell-k)n'} \right) \right) \\
&= \sum_n f_n \sum_{n'} g_{n'} \sum_k w_N^{-kn} w_N^{-(\ell-k)n'} \\
&= \sum_n f_n \sum_{n'} g_{n'} \sum_k w_N^{-kn} w_N^{-\ell n'} w_N^{+kn'} \\
&= \sum_n f_n \sum_{n'} g_{n'} w_N^{-\ell n'} \sum_k w_N^{-kn} w_N^{+kn'} \\
&= \sum_n f_n \sum_{n'} g_{n'} w_N^{-\ell n'} \sum_k w_N^{k(n'-n)} \\
&= \sum_n f_n \sum_{n'} g_{n'} w_N^{-\ell n'} \left(N \hat{\delta}_N(n' - n) \right) \\
&= (N) \sum_n f_n \sum_{n'} g_{n'} w_N^{-\ell n'} \hat{\delta}_N(n' - n)
\end{aligned}$$

Performing the summation over n'

$$= (N) \sum_n f[n] g[n] w_N^{-\ell n}$$

, which is the DFT of the product of f and g evaluated at ℓ

$$= (N)(\mathcal{F}(fg))[\ell]$$

The (discrete) convolution theorem also refers to the convolution of DFTs, which under the given definitions for the DFT pair, is

$$\boxed{\mathcal{F}(f) * \mathcal{F}(g) = N \mathcal{F}(fg)} \quad (22)$$

Using the definition of the convolution, the convolution of IDFTs is

$$\begin{aligned} (\mathcal{F}^{-1}f) * (\mathcal{F}^{-1}g) &= \left((\mathcal{F}^{-1}f)[n] \right) * \left((\mathcal{F}^{-1}g)[m] \right) \\ &= \sum_n (\mathcal{F}^{-1}f)[n] (\mathcal{F}^{-1}g)[m-n] \\ &= \sum_n \left(\left(\frac{1}{N} \sum_k f_k w_N^{+kn} \right) \left(\frac{1}{N} \sum_{k'} g_{k'} w_N^{+k'(m-n)} \right) \right) \\ &= \frac{1}{N^2} \sum_n \sum_k f_k \sum_{k'} g_{k'} w_N^{kn} w_N^{k'(m-n)} \\ &= \frac{1}{N^2} \sum_k f_k \sum_{k'} g_{k'} \sum_n w_N^{kn} w_N^{k'm} w_N^{-k'n} \\ &= \frac{1}{N^2} \sum_k f_k \sum_{k'} g_{k'} w_N^{k'm} \sum_n w_N^{(k-k')n} \\ &= \frac{1}{N^2} \sum_k f_k \sum_{k'} g_{k'} w_N^{k'm} N \hat{\delta}_N(k-k') \end{aligned}$$

Performing the summation over k' the discrete delta, $\hat{\delta}_N$, selects the k^{th} element

$$= \frac{N}{N^2} \sum_k f[k] g[k] w_N^{km}$$

, which is the IDFT evaluated at IDFT[m]

$$= \frac{1}{N} \sum_k f[k] g[k] w_N^{km} = (\mathcal{F}^{-1}(fg))[m]$$

meaning the convolution of IDFTs is

$$\boxed{(\mathcal{F}^{-1}f) * (\mathcal{F}^{-1}g) = \mathcal{F}^{-1}(fg)} \quad (23)$$

Using the alternative definitions for the DFT and the IDFT

$$(\mathcal{F}f)[k] = \frac{1}{N} \sum_n w_N^{-kn} \quad \text{and} \quad (\mathcal{F}^{-1}f)[n] = \sum_k w_N^{kn}$$

the procedure for deriving the discrete convolution becomes

$$\begin{aligned}
\mathcal{F}^{-1}(\mathcal{F}f \mathcal{F}g)[m] &= \mathcal{F}^{-1}(F[k] G[k])[m] \\
&= \sum_k w_N^{km} \left(\left(\frac{1}{N} \sum_n f_n w_N^{-kn} \right) \left(\frac{1}{N} \sum_{n'} g_{n'} w_N^{-kn'} \right) \right) \\
&= \frac{1}{N^2} \left(\sum_n \sum_{n'} f_n g_{n'} w_N^{-k(n+n')} \right) \sum_k w_N^{km} \\
&= \frac{1}{N^2} \sum_n \sum_{n'} f_n g_{n'} \sum_k w_N^{k(m-n-n')} \\
&= \frac{1}{N^2} \sum_n \sum_{n'} f_n g_{n'} N \hat{\delta}_N(m-n-n') \\
&= \frac{1}{N} \sum_n f[n] g[m-n]
\end{aligned}$$

and the definition of the convolution under this definition of the DFT ($\mathcal{F}f = \frac{1}{N} \sum_n f_n w_N^{-kn}$) is

$$(f * g)[m] \equiv \frac{1}{N} \sum_n f[n] g[m-n]$$

MOVE THESE NOTES TO END OF SECTION AND CLEAN UP... Briggs and Henson define the discrete convolution as (p.81 [5])

$$f_n * g_n = \sum_j f_j g_{n-j}$$

I find this notation incredibly confusing, f and g ostensibly share an index, n , yet it is not the index referenced in the summation? But their description of the sequences is interesting... To form the sequence g_{n-j} first form the sequence g_{-j} from the reverse sequence then shift this to the left by n ... need to think more about this... since the convolution index feels to me like it should be a dummy index that eventually supplants the old index (even if in the end it refers to the same grid).

*** IMPORTANT FORMALISM DETAILS BY YOU!!! *** At some point you need to lay down your formalism about discrete delta functions and the need to perform a sum in order to evaluate them, to “collapse the delta function” ... ***

MIT OCW 6.003 Signals and Systems, Fall 2011 Lect. 8 Convolution def. convolution

$$y[n] = \sum_k x[k] h[n-k] = (x * h)[n]$$

convolution is applied to (entire) signals not samples. Admittedly confusing conventional notation (min 14)

$$y[n] = \sum_k x[k] h[n-k] = x[n] * h[n]$$

...really, neither x or h is a function of n but rather $(x * h)[n]$ is a function of n . A better notation is

$$(f[n] * g[n])[m] = \sum_n f[n] g[m-n]$$

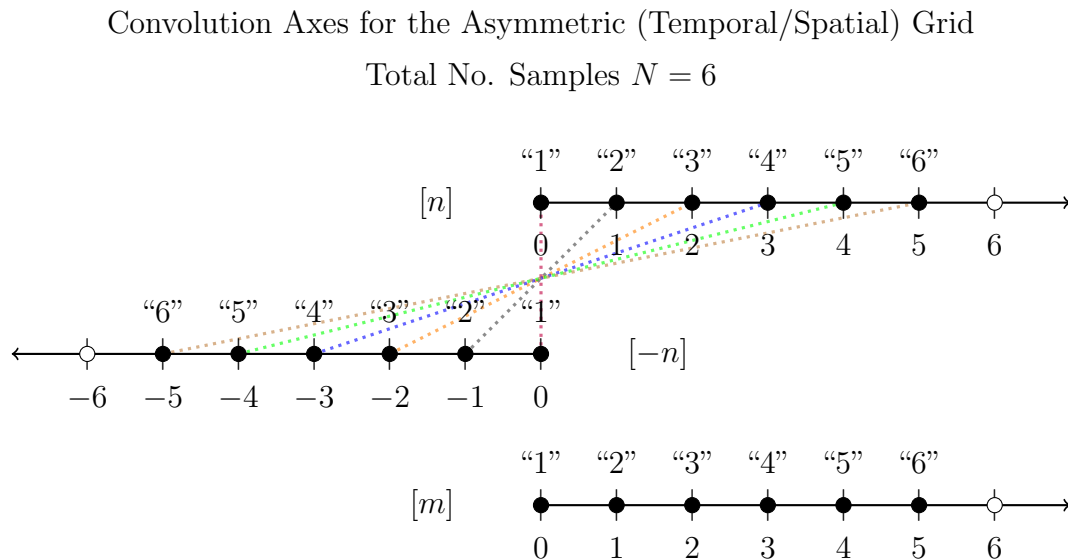
It is important to note the index $[m]$ has not yet been defined but it is essentially an alternative (dummy) index for the same grid.

Freeman does a visualization that should be good reference for convolution but WORKING ON IT A semi-illustration of the convolution algorithm

NOTE: Ans. it is $h[k]$ reflected about the origin of the $[k]$ axis $[k = 0]$ min 25:41 (but we do not know if N is even or odd, although it is on a symmetric axis, in the example shown).

The discrete convolution for N even, ie: $N = 8$, on a symmetric grid (axis centered at zero) of f convolved with itself NOTE: change this from $N = 8$ to $N = 6$ so you can show the steps of convolution on a row...

Just focus on the axes...put this with the convolution-algorithm it is associated with...do the other version for the alternative convolution-algorithm and place it accordingly...



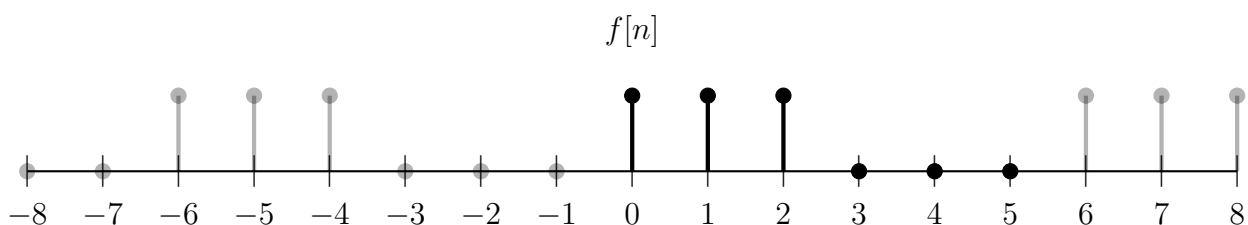
The $[m - n]$ itself is an index with elements

$$\begin{aligned}
 [0 - n] &: [0, -1, -2, \dots, -(N - 1)] \\
 [1 - n] &: [1, 0, -1, \dots, 1 - (N - 1)] \\
 &\vdots \\
 [m - n] &: [m, m - 1, m - 2, \dots, m - n, \dots, m - (N - 1)] \\
 &\vdots \\
 [(N - 1) - n] &: [(N - 1), (N - 1) - 1, (N - 1) - 2, \dots, 0]
 \end{aligned}$$

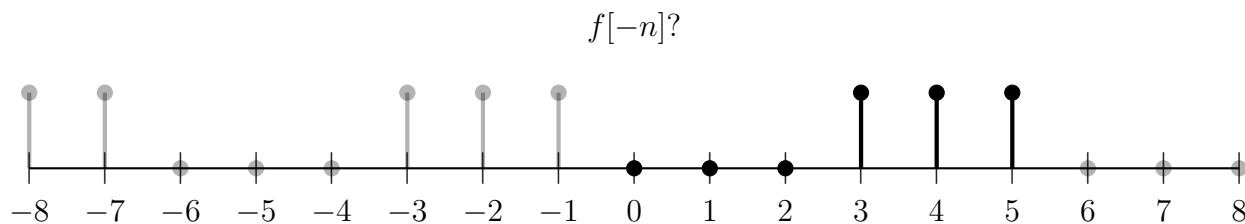
The question remains, what does $f[-n]$, and by extension $f[m - n]$, look like? To investigate this question, consider the convolution of $f[n]$ with itself

$$(f * f)[m] = \sum_{n=0}^{N-1} f[n]f[m - n]$$

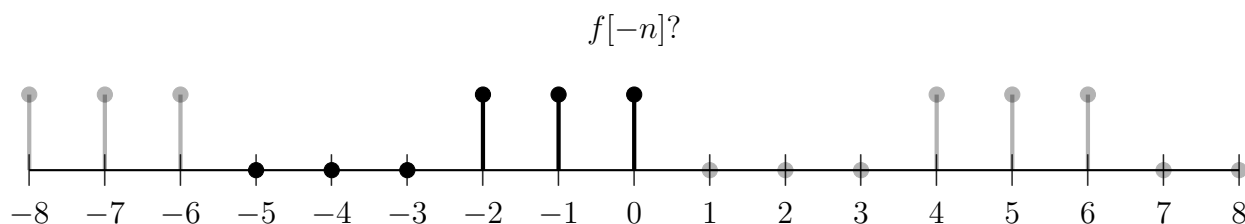
where $f[n]$ is the 6 sample boxcar function defined over the asymmetric grid, including its periodic extension, as



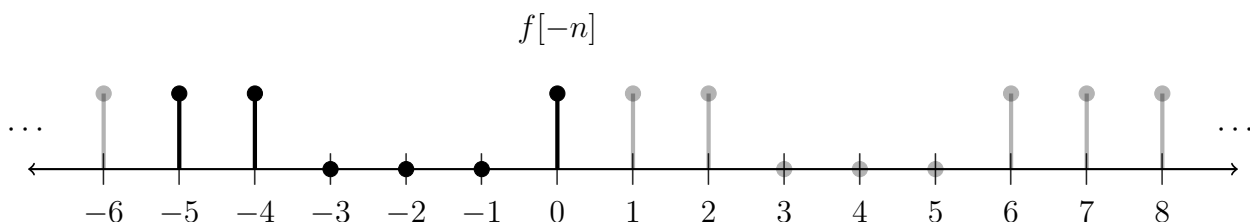
Then one version of what $f[-n]$ might be is a literal reversal of the data sampled, ie: the samples are reported in exactly the opposite order of how they were originally recorded



Another version of what $f[-n]$ might be the reflection (or 180 degree rotation) of the data/signal about the vertical axis where the n -index equals zero (But this is effectively the same as the signal itself)



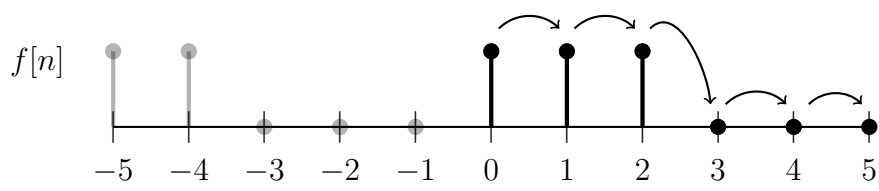
Another way to construct $f[-n]$ might be to preserve $f[n]$ in its original form but count the samples in the opposite direction (from right to left) starting from zero utilizing the periodic extension of the signal. This is the same as reflecting the signal about the vertical axis at $n = 0$ and recording the samples from left to right (do figure: MOVE THIS DOWN)



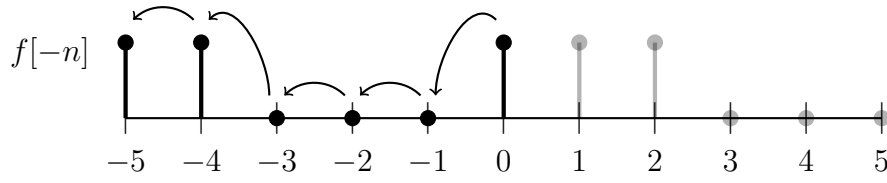
(The goal is “no pedagogy” (ie: make everything seem a coherent progression, such that there is little need for providing motivation for definition)...What is the right check/comparison?) The possibilities are not endless. Which one is correct? Let’s proceed by looking at the simplest example possible $f * f$... It would be nice to be clear about what m is here... By default, m refers to the same axis as n , such that $f[n]g[m - n]$ is intrinsically meaningful - and, m is therefore an alias for what is commonly referred to as n , the n -index is associated with distance in time or in space.

note: signals do not even have to be the same length to perform convolution[16](amazing ref on discrete convolution p.49)

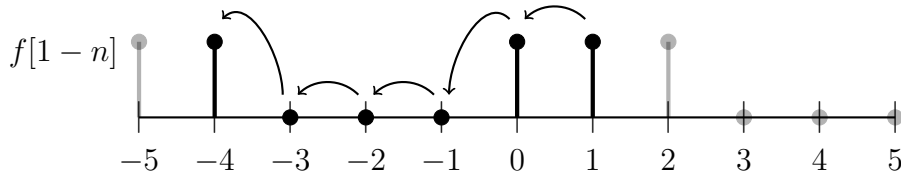
(No idea why this image is not located correctly...) Computing the convolution requires a definition for $f[m - n]$. The approach shown here is an index-centric approach in that all of the manipulations to compute the convolution involve the index exclusively, the signal is untouched and remains intact. The periodic extension is invoked to include negative indices. Signal samples associated with negative indices can be counted in the negative direction from right to left. The reverse signal, $f[-n]$, is constructed from $f[n = 0]$ and using the periodic extension to proceed sampling along negative indices. One complete cycle of the reverse signal, $f[-n]$, extends from $f[n = 0]$ to $f[n = -(N - 1)]$ and proceeds in the negative direction along indices $n = 0, -1, -2, \dots, -(N - 1)$.



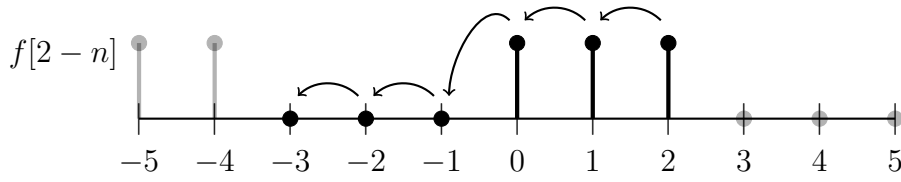
$$(f * f)[m] = \sum_n f[n] f[m - n]$$



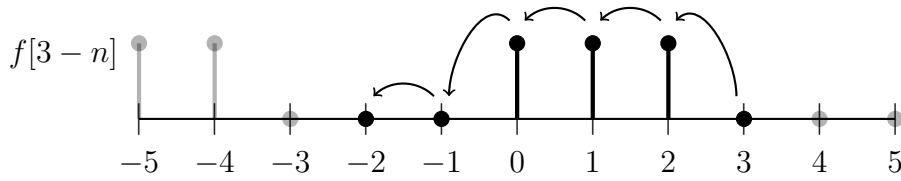
$$\Rightarrow (f * f)[0] = 1$$



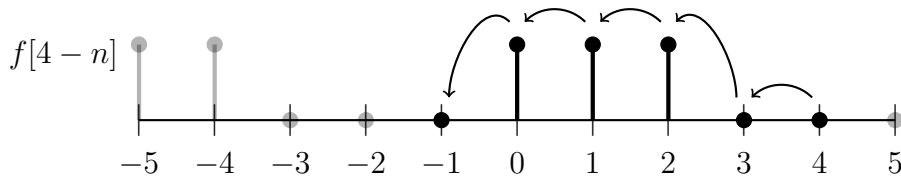
$$\Rightarrow (f * f)[1] = 2$$



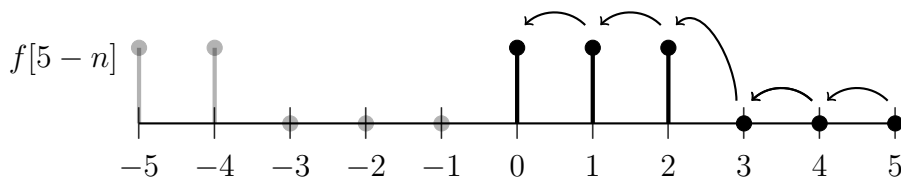
$$\Rightarrow (f * f)[2] = 3$$



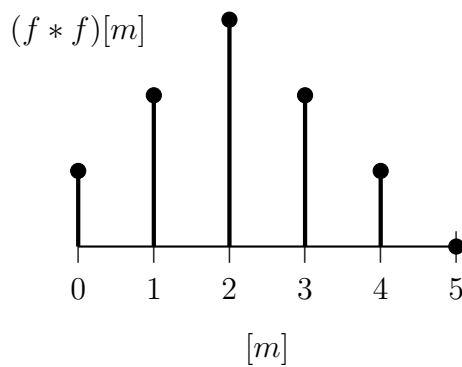
$$\Rightarrow (f * f)[3] = 2$$



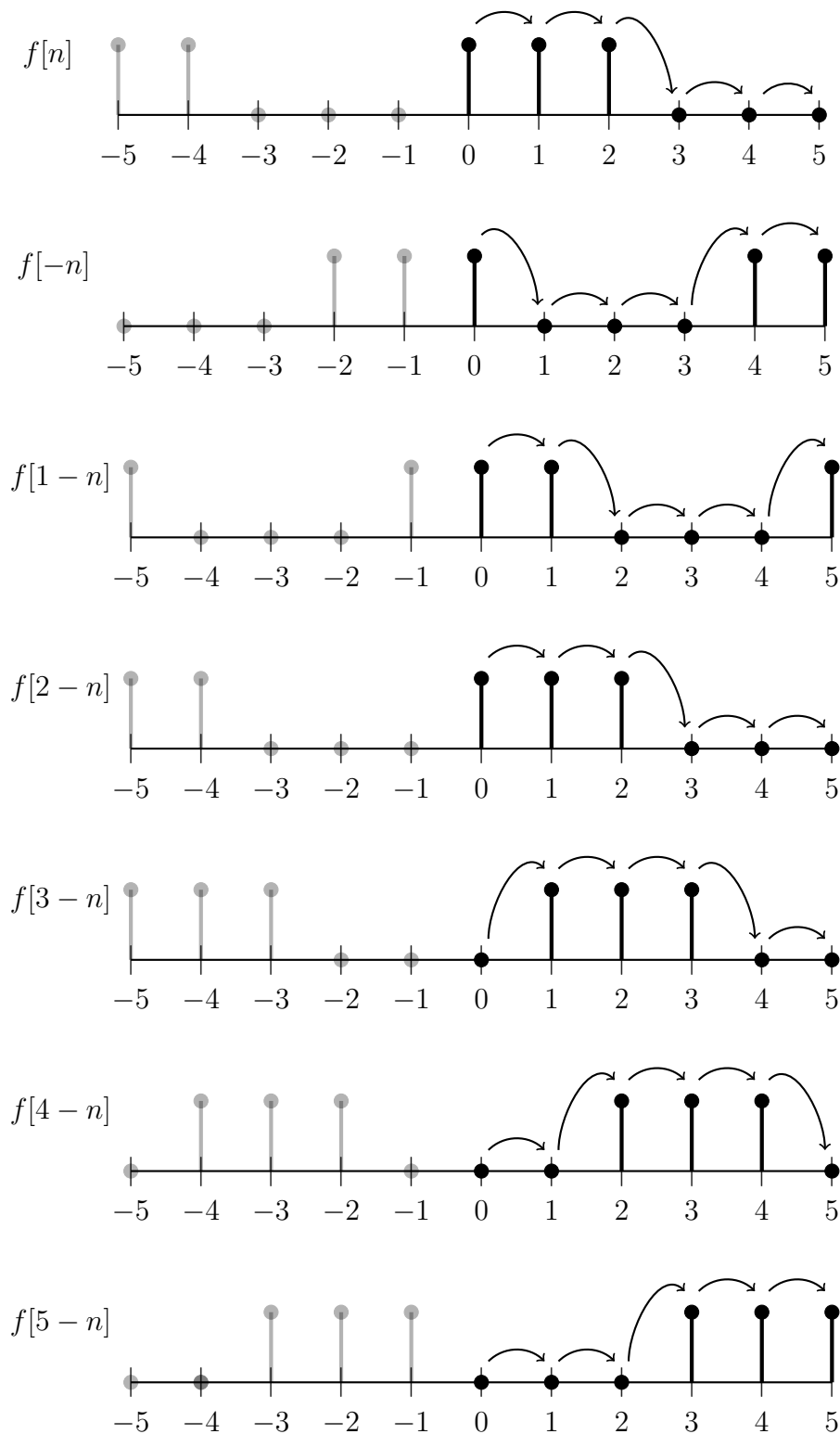
$$\Rightarrow (f * f)[4] = 1$$



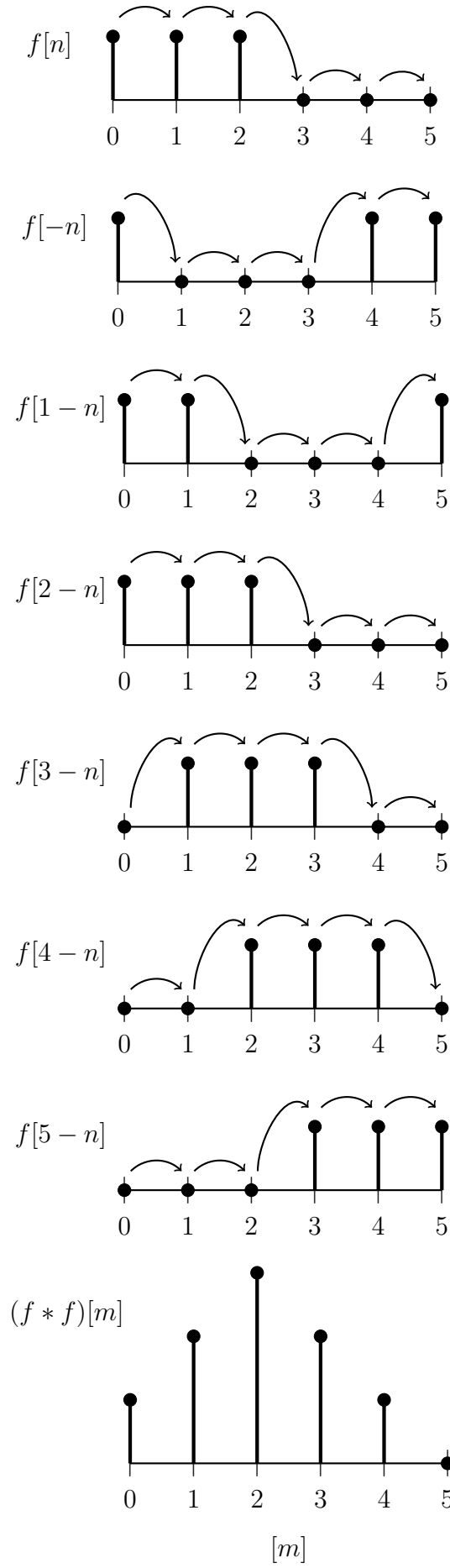
$$\Rightarrow (f * f)[5] = 0$$



The other option...is a signal-centric approach. Reflect signal about $n = 0$ and shift signal to the right by m to generate $f[m - n]$. Take the signal from $n = 0$ to $n = N - 1$ and its periodic extension in the negative indices from $-(N - 1)$ to -1 . The index n ranges from $[0, \dots, N - 1]$



This approach of using the periodic extension when shifting the reverse signal, $f[-n]$, to the right in the positive direction to generate the $f[m - n]$ is equivalent to cycling the signal values: “popping the tail and appending it to the head of the list, reindexing samples to match the new order”.



This approach works for even or odd N , obviously. It also works for signals on the symmetric grid. To implement this for N odd on the symmetric grid is simple. To implement this for N even on the

symmetric grid is maybe worth showing... Also the convolution does not have to involve signals that are the same length (same no. samples)...Definitely show how to do this!!!

MIT OCW Signals and Systems Spring 2011, Alan V. Oppenheim, Lec 4 Convolution Sometimes convolution of x and h is denoted

$$x[n] * h[n] = \sum_k x[k]h[n - k]$$

this is confusing because it does not appear that x is a function of n but rather a function of k . But $x[n]$ can be written as (cite?)

$$x[n] = \sum_k x[k]\delta[n - k]$$

, which makes the notation a little more justified.

11 The DFT of Common Signals

NOTE: the definition for the DFT used in these examples lacks the $1/N$ factor most if not all of the time in the subsequent examples and treatment...(UPDATE this or modify my own convention?)

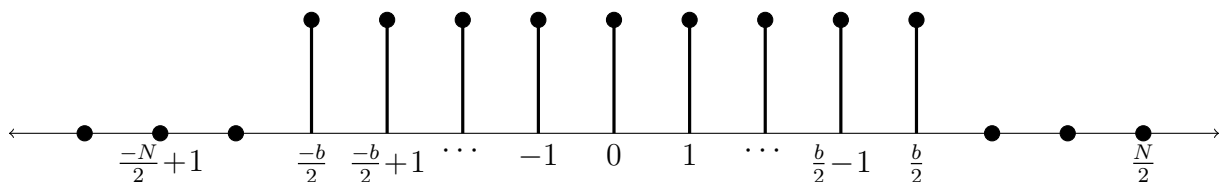
11.1 The DFT of the Boxcar, Π

The DFT of the boxcar function is so important because it is used for signal reconstruction in the discrete form of the sampling theorem.

(boxcar with $2N$ data points - p.60 , Digital Filters (Hanning))

To center the boxcar function, Π , on the symmetric axis grid it must have an odd number of data points in the region where it is non-zero for it to remain symmetrical. An even number of points in the non-zero region would produce an asymmetric construction were it assembled on a grid axis centered at zero (DO THIS NEXT...This is essentially what sampling with N even is). The discrete boxcar function of width (b), Π_b , is

$$\Pi_b = \begin{cases} 1 & \text{if } |k| \leq b/2 \\ 0 & \text{if } |k| > b/2 \end{cases}$$



For $k = 0$ the DFT of the boxcar, $\mathcal{F}\Pi_b$, is

$$F[0] = \sum_n \Pi_b e^{-i2\pi(0)n/N} = \sum_{n=-b/2}^{b/2} 1 = b + 1$$

The DFT of the boxcar for $k \neq 0$ is

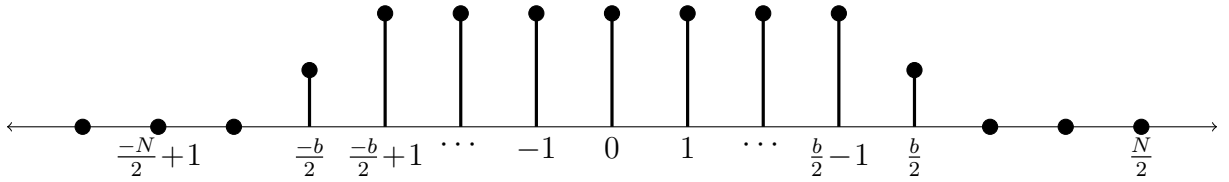
$$\begin{aligned}
F[k] &= \sum_n \Pi_b e^{-i2\pi kn/N} = \sum_{n=-b/2}^{b/2} e^{-i2\pi kn/N} = \frac{e^{i2\pi kb/(2N)} - e^{-i2\pi k(b/2+1)/N}}{1 - e^{-i2\pi k/N}} \\
&= \frac{e^{-i\pi k/N} (e^{\frac{i2\pi k}{N} \frac{b}{2} + \frac{i\pi k}{N}} - e^{\frac{-i2\pi k}{N} (\frac{b}{2}+1) + \frac{i\pi k}{N}})}{e^{-i\pi k/N} (e^{i\pi k/N} - e^{-i\pi k/N})} \\
&= \frac{e^{\frac{i2\pi k}{N} \frac{b}{2} + \frac{i2\pi k}{2N}} - e^{(\frac{-i2\pi k}{2N} (b+2) + \frac{i2\pi k}{2N})}}{e^{i\pi k/N} - e^{-i\pi k/N}} \\
&= \frac{e^{\frac{i2\pi k}{2N} (b+1)} - e^{\frac{-i2\pi k}{2N} (b+2-1)}}{e^{i\pi k/N} - e^{-i\pi k/N}} \\
&= \frac{(e^{\frac{i\pi k}{N} (b+1)} - e^{\frac{-i\pi k}{N} (b+1)}) / (i2)}{(e^{i\pi k/N} - e^{-i\pi k/N}) / (i2)} \\
&= \frac{\sin(\pi k(b+1)/N)}{\sin(\pi k/N)}
\end{aligned}$$

Then the DFT of the boxcar with (b) odd on the symmetric grid is

$$\mathcal{F} \Pi_b = \begin{cases} b+1 & k \bmod N = 0 \\ \frac{\sin(\pi k(b+1)/N)}{\sin(\pi k/N)} & k \bmod N \neq 0 \end{cases}$$

(...motivation...) Taking the average at the discontinuities, the boxcar function of width (b) odd becomes

$$\Pi_b = \begin{cases} 1 & \text{if } |k| < b/2 \\ 1/2 & \text{if } |k| = b/2 \\ 0 & \text{if } |k| > b/2 \end{cases}$$



Taking the average at the discontinuities, the DFT of the boxcar of width (b) odd becomes

$$\begin{aligned}
F[0] &= \sum_n \Pi_b e^{-i2\pi(0)n/N} = \frac{1}{2}e^{-i2\pi(0)(-b/2)/N} + \frac{1}{2}e^{-i2\pi(0)(b/2)/N} + \sum_{n=-b/2+1}^{b/2-1} 1 = 1 + (b-1) = b \\
F[k \neq 0] &= \sum_n \Pi_b e^{-i2\pi kn/N} = \frac{1}{2}e^{-i2\pi k(-b/2)/N} + \frac{1}{2}e^{-i2\pi k(b/2)/N} + \sum_{n=-b/2+1}^{b/2-1} e^{-i2\pi kn/N} \\
&= \frac{e^{i2\pi k(b/2)/N} + e^{-i2\pi k(b/2)/N}}{2} + \frac{e^{-i2\pi k(-b/2+1)/N} - e^{-i2\pi k(b/2)/N}}{1 - e^{-i2\pi k/N}} \\
&= \cos(\pi kb/N) + \frac{e^{-i\pi k/N}(e^{i2\pi k(b/2-1)/N+i\pi k/N} - e^{-i2\pi k(b/2)/N+i\pi k/N})}{e^{-i\pi k/N}(e^{i\pi k/N} - e^{-i\pi k/N})} \\
&= \cos(\pi kb/N) + \frac{e^{\frac{i2\pi k}{2N}(b-2)+\frac{i2\pi k}{2N}} - e^{-\frac{i2\pi kb}{2N}+\frac{i2\pi k}{2N}}}{e^{i\pi k/N} - e^{-i\pi k/N}} \\
&= \cos(\pi kb/N) + \frac{e^{\frac{i2\pi k}{2N}(b-2+1)} - e^{-\frac{i2\pi k}{2N}(b-1)}}{e^{i\pi k/N} - e^{-i\pi k/N}} \\
&= \cos(\pi kb/N) + \frac{e^{\frac{i\pi k}{N}(b-1)} - e^{-\frac{i\pi k}{N}(b-1)}}{e^{i\pi k/N} - e^{-i\pi k/N}} \\
&= \cos(\pi kb/N) + \frac{\sin(\pi k(b-1)/N)}{\sin(\pi k/N)}
\end{aligned}$$

Distributing the denominator, collecting terms, and focusing on the numerator

$$\cos(\pi kb/N) \sin(\pi k/N) + \sin(\pi k(b-1)/N) = \cos\left(\frac{\pi kb}{N}\right) \sin\left(\frac{\pi k}{N}\right) + \sin\left(\frac{\pi kb}{N} - \frac{\pi k}{N}\right)$$

Using the identity $\sin(\theta - \phi) = \sin \theta \cos \phi - \cos \theta \sin \phi$

$$\begin{aligned}
&= \cos(\pi kb/N) \sin(\pi k/N) + \sin(\pi kb/N) \cos(\pi k/N) - \cos(\pi kb/N) \sin(\pi k/N) \\
&= \sin(\pi kb/N) \cos(\pi k/N)
\end{aligned}$$

Taking the average at the discontinuities, the DFT of the boxcar of width (b) odd simplifies to

$$\mathcal{F} \Pi_b = \begin{cases} b & k \bmod N = 0 \\ \frac{\sin\left(\frac{\pi k}{N}b\right) \cos\left(\frac{\pi k}{N}\right)}{\sin\left(\frac{\pi k}{N}\right)} & k \bmod N \neq 0 \end{cases}$$

Osgood refers to this as the discrete sinc function with the notation `dinc`. I HATE the function name “`dinc`”!!![4]. The name `sinc` is short for sine cardinal. Maybe a nomenclature like discrete sine cardinal, `disc`, or `dsc`, would be better. Let’s just call it what it is, the DFT of the discrete boxcar function, which is such a special thing that it deserves to stand out and gets the elaborate notation \mathcal{B} . It would have been nice to go with a lowercase notation because it’s a function but it’s not available for math fonts `mathscr` or `mathcal`, and lowercase `mathfrak`, `b`, looks terrible. The `mathcal` version, \mathcal{B} , looks ok and might allude visually to the DFT from which it came, but to jar your memory and to emphasize its importance let’s make the DFT of the boxcar stand out and look as distinct from the standard notation regarding the DFT and as baroque as possible.

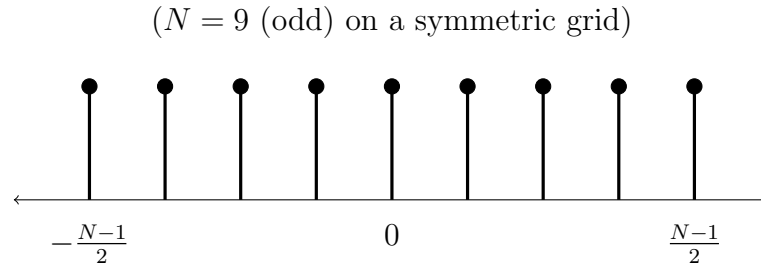
$$\boxed{\mathcal{F} \Pi \equiv \mathcal{B}}$$

NEXT: - extend boxcar to edge of box to make it a window function (that is what is actually used in the discrete form of the sampling theorem, as the boxcar encapsulates the entire frequency spectrum...): -

N even, and odd on symmetric grid - asymmetric grid (N even and odd same) - windows with average at end points (but maybe check back on how the input signal gets modified prior to the DFT calculation to make sure you are not technically doing something redundant by averaging the endpoints via the window function!...) - boxcar for $N \in [0 : N - 1]$ (asymmetric window) - N even symmetric - N even asymmetric boxcar boxcar with even number of points in its width ...

The DFT of the N -odd symmetric boxcar window ...

OMG THIS SHIT BY Dr. Osgood IS ALL WRONG!!! THE EDGE OF THE WINDOW (or boxcar) CANNOT HAVE INDEX OF $N/2$ IF N IS ODD, because if N is odd then $N/2$ is not an integer and all grid indices must have integer values !!!!!!!!



If $k = 0$

$$F[0] = \sum_n \Pi e^{-i2\pi(0)n/N} = \sum_{n=-(N-1)/2}^{(N-1)/2} 1 = N$$

If $k \neq 0$

$$F[k] = \sum_n \Pi e^{-i2\pi kn/N} = \sum_{n=-(N-1)/2}^{(N-1)/2} \left(e^{-i\frac{2\pi}{N}k} \right)^n$$

Using

$$\sum_{k=p}^{p+N-1} \omega_N^{k(m-n)} = \frac{\omega_N^p - \omega_N^{p+N}}{1 - \omega_N} \equiv N \delta_N[m - n] = \begin{cases} N & \text{if } m = n \\ 0 & \text{if } m \neq n \text{ because } \omega_N^p = \omega_N^{p+N} \end{cases}$$

$$\begin{aligned} &= \frac{\exp\left(-i\frac{2\pi}{N}k\left(-\frac{(N-1)}{2}\right)\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{(N-1)}{2} + 1\right)\right)}{1 - \exp\left(-i\frac{2\pi}{N}k\right)} \\ &= \frac{\exp\left(i\frac{2\pi}{N}k\left(\frac{N-1}{2}\right)\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{N+1}{2}\right)\right)}{1 - \exp\left(-i\frac{2\pi}{N}k\right)} \\ &= \frac{\exp\left(i\frac{\pi}{N}k(N-1)\right) - \exp\left(-i\frac{\pi}{N}k(N+1)\right)}{1 - \exp\left(-i\frac{2\pi}{N}k\right)} \\ &= \frac{\exp\left(-i\frac{\pi}{N}k\right)\left(\exp\left(i\frac{\pi}{N}k(N)\right) - \exp\left(-i\frac{\pi}{N}k(N)\right)\right)}{\exp\left(-i\frac{\pi}{N}k\right)\left(\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)\right)} \\ &= \frac{\exp(i\pi k) - \exp(-i\pi k)}{\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)} \\ &= \frac{\exp(i\pi k) - \exp(-i\pi k)}{\exp(i\pi k/N) - \exp(-i\pi k/N)} \end{aligned}$$

Using

$$\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{i2}$$

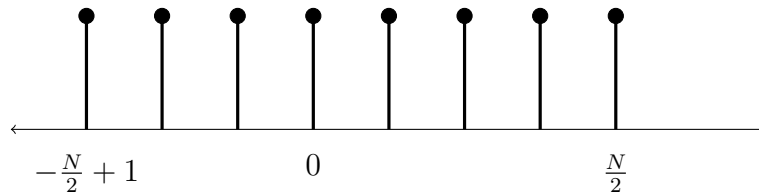
$$\begin{aligned} &= \frac{\left(\exp(i\pi k) - \exp(-i\pi k)\right)/i2}{\left(\exp(i\pi k/N) - \exp(-i\pi k/N)\right)/i2} \\ &= \frac{\sin(\pi k)}{\sin(\pi k/N)} \end{aligned}$$

If N is odd, the DFT of the symmetric window is

$$\mathcal{F}\Pi = \mathcal{W} = \begin{cases} N & \text{if } k \bmod N = 0 \\ \frac{\sin(\pi k)}{\sin(\pi k/N)} & \text{if } k \bmod N \neq 0 \end{cases} \quad (24)$$

The DFT of the N -even symmetric boxcar window ...

($N = 8$ even on a symmetric grid)



If $k = 0$

$$F[0] = \sum_n \Pi e^{-i2\pi(0)n/N} = \sum_{n=-N/2+1}^{N/2} 1 = N$$

If $k \neq 0$

$$F[k] = \sum_n \Pi e^{-i2\pi kn/N} = \sum_{n=-N/2+1}^{N/2} \left(e^{-i\frac{2\pi}{N}k} \right)^n$$

Using
$$\sum_{k=p}^{p+N-1} \omega_N^{k(m-n)} = \frac{\omega_N^p - \omega_N^{p+N}}{1 - \omega_N} \equiv N \delta_N[m - n] = \begin{cases} N & \text{if } m = n \\ 0 & \text{if } m \neq n \text{ because } \omega_N^p = \omega_N^{p+N} \end{cases}$$

****FINISH****

$$\begin{aligned} &= \frac{\exp\left(-i\frac{2\pi}{N}k\left(-\frac{N}{2} + 1\right)\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{N}{2} + 1\right)\right)}{1 - \exp\left(-i\frac{2\pi}{N}k\right)} \\ &= \frac{\exp\left(i\frac{2\pi}{N}k\left(\frac{N}{2} - 1\right)\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{N}{2} + 1\right)\right)}{\exp\left(-i\frac{\pi}{N}k\right)\left(\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)\right)} \end{aligned}$$

One approach is to absorb the factor in the denominator into the terms in the numerator

$$\begin{aligned} &= \frac{\exp\left(+i\frac{\pi}{N}k\right)\left(\exp\left(i\frac{2\pi}{N}k\left(\frac{N}{2} - 1\right)\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{N}{2} + 1\right)\right)\right)}{\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)} \\ &= \frac{\exp\left(i\frac{2\pi}{N}k\left(\frac{N}{2} - 1\right) + i\frac{2\pi}{2N}k\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{N}{2} + 1\right) + i\frac{2\pi}{2N}k\right)}{\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)} \\ &= \frac{\exp\left(i\frac{2\pi}{2N}k(N - 2 + 1)\right) - \exp\left(-i\frac{2\pi}{2N}k(N + 2 - 1)\right)}{\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)} \\ &= \frac{\exp\left(i\frac{\pi}{N}k(N - 1)\right) - \exp\left(-i\frac{\pi}{N}k(N + 1)\right)}{\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)} \end{aligned}$$

However, the numerator does not simplify to the sin function as is - there is a phase factor that needs to be factored out

$$\begin{aligned} &= \frac{\exp\left(-i\frac{\pi}{N}k\right)\left(\exp\left(i\frac{\pi}{N}k(N)\right) - \exp\left(-i\frac{\pi}{N}k(N)\right)\right)}{\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)} \\ &= e^{-i\pi k/N} \frac{\sin(\pi k)}{\sin(\pi k/N)} \end{aligned}$$

****The OLD****

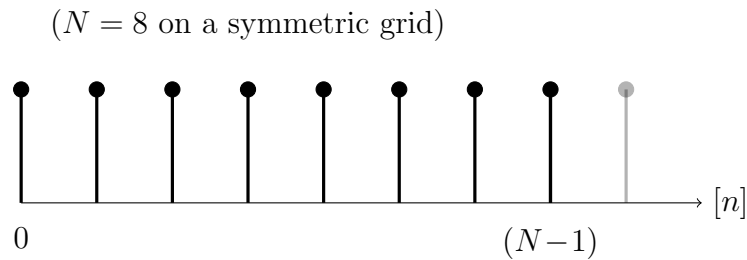
$$\begin{aligned}
&= \frac{\exp\left(-i\frac{2\pi}{N}k\left(-\frac{(N-1)}{2}\right)\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{(N-1)}{2} + 1\right)\right)}{1 - \exp\left(-i\frac{2\pi}{N}k\right)} \\
&= \frac{\exp\left(i\frac{2\pi}{N}k\left(\frac{N-1}{2}\right)\right) - \exp\left(-i\frac{2\pi}{N}k\left(\frac{N+1}{2}\right)\right)}{1 - \exp\left(-i\frac{2\pi}{N}k\right)} \\
&= \frac{\exp\left(i\frac{\pi}{N}k(N-1)\right) - \exp\left(-i\frac{\pi}{N}k(N+1)\right)}{1 - \exp\left(-i\frac{2\pi}{N}k\right)} \\
&= \frac{\exp\left(-i\frac{\pi}{N}k\right)\left(\exp\left(i\frac{\pi}{N}k(N)\right) - \exp\left(-i\frac{\pi}{N}k(N)\right)\right)}{\exp\left(-i\frac{\pi}{N}k\right)\left(\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)\right)} \\
&= \frac{\exp(i\pi k) - \exp(-i\pi k)}{\exp\left(i\frac{\pi}{N}k\right) - \exp\left(-i\frac{\pi}{N}k\right)} \\
&= \frac{\exp(i\pi k) - \exp(-i\pi k)}{\exp(i\pi k/N) - \exp(-i\pi k/N)}
\end{aligned}$$

Using

$$\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{i2}$$

$$\begin{aligned}
&= \frac{(\exp(i\pi k) - \exp(-i\pi k))/i2}{(\exp(i\pi k/N) - \exp(-i\pi k/N))/i2} \\
&= \frac{\sin(\pi k)}{\sin(\pi k/N)}
\end{aligned}$$

The DFT of the asymmetric boxcar window ...



(TO BE CONTINUED...)

12 The DFT of III - the Discrete Sampling Theorem

Ahead: why it's important to use $1/2$ at the discontinuities in the discrete boxcar function ($\alpha = 1/2$ in Osgood text).

class of signals: periodic, bandlimited ($F[k] = 0$ for $|k| \geq (b = \Omega)/2$), length N (indexed from $-N/2+1 : N/2$)

In discrete-land functions are more like vectors consisting of N samples of the function. It is important to remember that the elements of this vector themselves are indexed from 0, which indicates the starting point or initial value/condition as no progress has been made at the beginning of the process being observed.

(...) Ingredients for the Discrete Sampling and Interpolation Theorem

1. discrete bandlimited functions of the DFT are

- periodic
- N samples (even, indices $[-N/2 + 1 : N/2]$)
- $\mathcal{F}f[k] = 0$ for $|k| = w = \Omega/2$

2. discrete Shah III - spacing distance, d , between sampling spikes, where N is divisible by d , therefore d is even also.

12.1 The Discrete Shah Function (does this need to be moved up?)

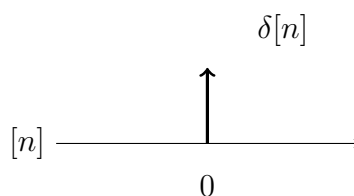
The definition for the DFT selected in this section is

$$F[k] = \mathcal{F}f = \sum_n f[n] w_N^{-kn} \quad \text{and} \quad f[n] = \mathcal{F}^{-1}F = \frac{1}{N} \sum_k F[k] w_N^{kn}$$

(Show others' definitions and ask How to denote the discrete Shah function? Ans: denote discreteness with square-bracket notation)

Notice, the argument in square brackets of the DFT and IDFT is not the summation index. In the definition of the convolution the argument is also not the summation index. Fair enough, these are transforms, which take functions of one variable and produce functions of another variable. The III function might be thought of as a transform, but it is also natural to think of it as simply a collection of delta functions. Regardless of how we chose to think about III , the notation used to denote III should facilitate working with this object. Specifically, the DFT of the Shah function, $\mathcal{F}\text{III}$, should be natural to write and compute. The DFT operates on signals that are a function of $[n]$ so it would be nice to be able to write, $\mathcal{F}(\text{III}[n])$, in the same way that it is natural to write, $\mathcal{F}(f[n])$. The III can certainly be a function of $[n]$ as it can exist on the $[n]$ grid but the question arises, what does $[n]$ refer to in the argument of the III function if not to the summation index? I would like to stick with the convention of not specifying the summation index as the argument in the case of defining a transform. In the discrete context, transforms involve summation over an index; the III is weird in that it is essentially a function that involves a summation in its definition. I think the solution to this is to think of III as a transform. The III is essentially a tool for sampling a function by targeting the index on which it resides.

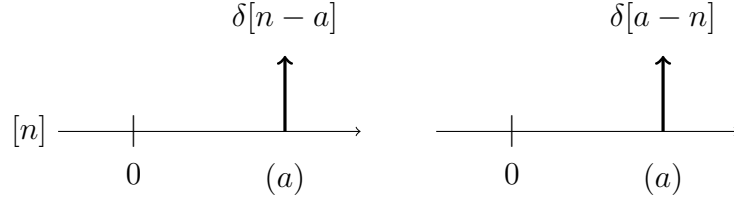
In $\delta[n]$ the argument $[n]$ is largely superfluous, other than it specifies the grid axis



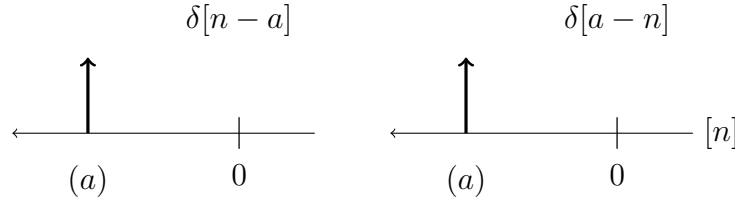
It is tricky to define the discrete Shah function due to the odd symmetry of the Kronecker delta (multiplying its argument by (-1) has no effect on the function/outcome)

$$\delta[n - a] = \delta[a - n]$$

Notice, (a) is deterministic, for $(a) > 0$



For, $(a) < 0$



Essentially it would be easier and just as clear to simply define the Kronecker delta without any reference to the grid axis, ie: $\delta[0]$, $\delta[a]$, and $\delta[-a]$. However, this solution for simplifying the notation for the Kronecker delta cannot be extended to \mathbb{III} , which must reference an index to generate the spike train of the Shah function.

The first question is, should the index reference the grid axis directly or indirectly via an alias? Secondly, how does the symmetry of the Kronecker delta influence this choice? I want to make a case for defining the Shah function as $\mathbb{III}[n]$, where $[n]$ is not the summation index but rather specifies the index of the target function, which will be swapped out for the index used to define \mathbb{III} . In this construction, the summation index used to define \mathbb{III} is an alias for the index/grid axis it ultimately targets.

The Shah function can be written as

$$\mathbb{III}_d[m] = \sum_n \delta[m - nd]$$

From this perspective the index is moving along its grid, and $[m]$ is a dummy index that only references the underlying grid associated with $[n]$ via its juxtaposition to $[n]$. (Note: maybe using the index $[n]$ will clash with other operators, ie: the DFT, use of the same index in their natural definition...)

It is entirely equivalent to represent the Shah function can be written as

$$\mathbb{III}_d[n] = \sum_m \delta[n - md]$$

In this notation, $[m]$ is a dummy index used to explicitly sample the $[n]$ index and by direct proxy its underlying grid.

Both of these conventions do not reference the summation index used in their definition, keeping with the convention for convolution. However, $\mathbb{III}[n]$, presents an opportunity to present a notational lure for the index/grid we want to target by hooking into another adjoining summation. According to this

view, the Shah function, $\text{III}[n]$, is primarily an object used to supplant the target index, $[n]$, of another function, ie: $f[n]$, with its own, the dummy index specified in the summation definition for the Shah function.

In comparison to the continuous case, III , where d is the distance between spikes/samples,

$$\text{III}_d(x) = \sum_{n=-\infty}^{\infty} \delta(x - nd)$$

the index n is used to iterate over all integers ($n \in \mathbb{Z}$). The variable x refers to the actual axis/dimension itself, whereas the index n is a dummy variable used to reference points along the axis. In contrast, for the DFT, the variable n

$$\mathcal{F}[k] = \sum_n^{(N)} e^{-i2\pi kn/N}$$

does refer directly to the (spatial/temporal) axis/dimension itself.

To maintain the use of a dummy variable to select points along the axis means keeping the reference to the physical dimensions in tact, n refers to x . So, the discrete Shah function needs a new dummy index to replace the n used in the continuous case to select points from the temporal/spatial grid where, ie: $n \in [0 : N - 1]$. This leads to the discrete Shah function using the dummy index, ie: $[m]$, in the summation definition

By comparison to the continuous case, III , where d is the distance between spikes/samples,

$$\text{III}_d(x) = \sum_{n=-\infty}^{\infty} \delta(x - nd)$$

the index n is used to iterate over all integers ($n \in \mathbb{Z}$). The variable x refers to the actual axis/dimension itself, whereas the index n is a dummy variable used to reference points along the axis. In contrast, for the DFT, the variable n

$$\mathcal{F}[k] = \sum_n^{(N)} e^{-i2\pi kn/N}$$

does refer directly to the (spatial/temporal) axis/dimension itself.

To maintain the use of a dummy variable to select points along the axis means keeping the reference to the physical dimensions in tact, n refers to x . So, the discrete Shah function needs a new dummy index to replace n in the continuous case to select points from the temporal/spatial grid where, ie: $n \in [0 : N - 1]$. This leads to the discrete Shah function using the dummy index, ie: $[m]$, in the summation definition

$$\text{III}_d[n] = \sum_{m=0}^{N/d-1} \delta[n - md]$$

to index over the grid set, $\{n\}$, of N consecutive integers. The implications of this construction are subtle but profound from the perspective of how this conceptualization translates into practical implementation. Using this formulation results in the Sha function effectively supplanting the index of a (discrete) function with its own. For example, the grid set $[n]$ of the DFT is targeted by $\text{III}_d[n]$ in

$$\mathcal{F}\text{III}_d[n] = \sum_n w_N^{-kn} \sum_m \delta[n - md]$$

Gathering terms then switching the order of summation

$$= \sum_m \sum_n w_N^{-kn} \delta[n - md]$$

And performing the sum over n

$$= \sum_m w_N^{-k(md)}$$

gives the DFT in terms of a new $[m]$ -index with an interval (d) . While the name of the index has changed, it is arbitrary and can be changed to suit the reference, including back to the previous grid reference.

The aforementioned paradigm of supplanting the index is in contrast to the Sha function as essentially sampling along an index

$$\text{III}_d[m] = \sum_n \delta[m - nd]$$

where the summation retains the n -index and $[m]$ is a dummy index that only references the grid via juxtaposition to $[n]$. In practice this leads to statements like the DFT of the Shah function is

$$\mathcal{F}\text{III}_d = \sum_n w_N^{-kn} \sum_n \delta[m - nd]$$

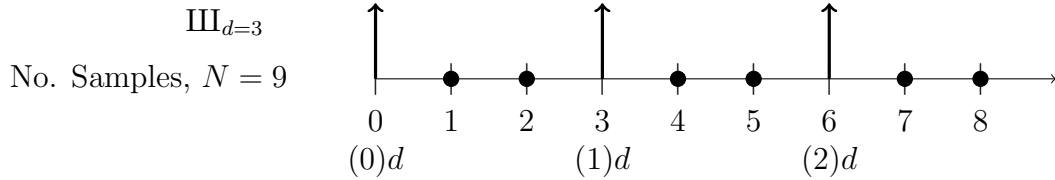
And then assert that the right most sum produces

$$= \sum_n w_N^{-k(m=nd)}$$

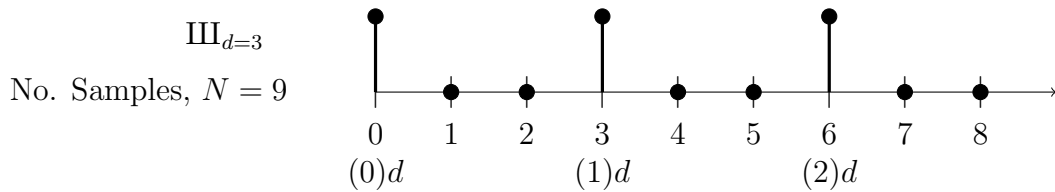
Which convention, $\text{III}[n]$ v. $\text{III}[m]$, I will adopt is yet to be determined. Ultimately, provided N is divisible by d

$$\text{III}_d = \begin{cases} 1 & m = 0, d, 2d, \dots, (N/d - 1)d \\ 0 & \text{else} \end{cases}$$

, which is the N -periodic (and N/d -periodic) sequence, ie:



To be pedantic without pedagogy, it is nice to represent the delta functions as arrows, and to think of them as such so that you are not tempted to actually perform the sum when thinking about the Shah III function, and rather to use the III function for what it is meant to do, which is sample another function. But in the discrete context it is also accurate and sometimes useful to simply regard the delta functions/spikes as 1's



, for example (ahead) when interpreting the DFT of the III function.

ULTIMATELY: The III function exists on a grid/axis associated with a specific index. The III function employs a dummy index to access this grid. ...perhaps worth noting that the dummy index can ultimately be replaced/renamed, ie: to reference the grid.

Sampling Property of the Discrete Shah Function

Recall, in the continuous case, $\text{III}(x)$ can be thought of as convolution with the delta function

$$(f(x) * \delta(x)) = \sum_n f(x) \delta(n - x) = f[n] = f \cdot \text{III}$$

to produce a discretized version of the original function.

In the discrete context, III switches out one index for another, ie: converts a function of $[n]$ to a function of $[\ell]$ (the summation index used to define III)

$$\begin{aligned} f[n] \cdot \text{III}_d[n] &= f[n] \cdot \sum_{\ell} \delta[n - d\ell] \\ &= \sum_{\ell=0}^{N/d-1} f[n] \delta[n - d\ell] = f[d\ell] \end{aligned}$$

Written like this, $\text{III}[n]$, suggests the Shah function is targeting the n -index. The discrete Shah function, $\text{III}[n]$, samples the function, $f[n]$, at scaled intervals along the grid axis, $n = d\ell$, supplanting the n -index with a dummy index, $[d\ell]$, where the $[\ell]$ -index is an alias for the $[n]$ -index.

Periodization Property of the Discrete Shah Function

The convolution takes place between two functions of the same variable and returns a new function of a different variable. Recall the definition of the convolution of $f[n]$ and $g[n]$

$$(f * g)[m] \equiv \sum_n f[n] g[m - n]$$

where $[m]$ is a dummy index that references the same grid axis as the $[n]$ -index via its juxtaposition to the $[n]$ -index (but maybe this is purely the default option and $[m]$ could in principle refer to something else?). The $[n]$ -index typically corresponds to the spatial/temporal grid but the convolution can be applied on any grid. Notice, the argument to the convolution is not the summation index while referring to the same axis.

The Shah function, $\text{III}_d[n]$, where (d) is some distance or duration

$$\text{III}[n] = \sum_{\ell} \delta[n - \ell d]$$

targets the $[n]$ -index.

The convolution of $f[n]$ and $\text{III}[n]$ is

$$\begin{aligned} (f * \text{III})[m] &= \sum_n f[n] \text{III}[m - n] \\ &= \sum_n f[n] \left(\sum_{\ell} \delta[(m - n) - \ell d] \right) \\ &= \sum_n \sum_{\ell} f[n] \delta[(m - n) - \ell d] \end{aligned}$$

Switching the order of summation, and performing the sum over the $[n]$ -index

$$\begin{aligned} &= \sum_{\ell} \sum_n f[n] \delta[(m - \ell d) - n] \\ &= \sum_{\ell} f[m - \ell d] \end{aligned}$$

Note the restriction on the summation index upon swapping out $[n]$ for $[\ell]$

$$\begin{aligned}
&= \sum_{\ell} \sum_{n=0}^{N-1} f[n] \delta[(m - \ell d) - n] \\
&= \sum_{\ell=0}^{N/d-1} f[m - \ell d]
\end{aligned}$$

Recall the $[\ell]$ -index is a dummy index and references the grid axis via its juxtaposition to the target index. At the end of the derivation it is appropriate to abandon this semantic exercise and reference the grid axis directly using its

Let's see what happens when the index used to iterate over the grid axis is used as the summation index

$$\text{III}[\ell] = \sum_n \delta[\ell - nd]$$

in the definition of the Shah function. This presents some ambiguity in defining the convolution as now the convolution, $(f * \text{III})[m]$, must take place between two functions that are not of the same variable, $f[n]$ and $\text{III}[\ell]$. Given the definition of the convolution

$$(f[n] * \text{III}[\ell])[m] = \sum_n f[n] \text{III}[m - n]$$

it is not clear how to substitute the variable, (ℓ) , of $\text{III}[\ell]$, into this expression. This predicament motivates a break with convention and refer to the summation index in the argument to the Shah function; in obvious contrast to the definition of the convolution which is notated as a function of $[m]$ while the summation takes place over $[n]$; but at least now the convolution can be written in the natural way

$$(f[n] * \text{III}[n])[m] = \sum_n f[n] \text{III}[m - n]$$

The next question that inevitably arises, is how to substitute $[m - n]$ into the Shah function? A possible substitution might be

$$(f * \text{III})[m] = \sum_n f[n] \left(\sum_n \delta[\ell - (m - n)d] \right)$$

Another possibility might be

$$(f * \text{III})[m] = \sum_n f[n] \left(\sum_n \delta[m - \ell + nd] \right)$$

but either option would make the Shah function so overly complicated that it is no longer facile to use.

All this ambiguity is before even getting to the issue of there being redundant summations involving the $[n]$ -index. After the first summation is performed the function that remains is no longer a function of $[n]$ but of $[m]$ and $[\ell]$. There is no way to resolve what remains: a summation over $[n]$ involving a function of $[m]$ and $[\ell]$.

This is the core context in which it becomes very beneficial to adopt my notation and construction of the Shah function!!!

$$\text{III}[n] = \sum_{\ell} \delta[n - \ell d]$$

This definition keeps with the convention that is already in place for defining other objects that exist on a specific grid, ie: the DFT, and the convolution, where the argument does not not reference

the summation index. Furthermore, the construction adopted here resembles the classical/continuous definition of the Shah function

$$\text{III}(x) = \sum_n \delta(x - nd)$$

where the summation index, (n) , is a dummy index that only references the (x) -axis via juxtaposition. In the discrete context, $[n]$, refers directly to the grid axis, and $[\ell]$ is the dummy index that can be scaled to select points at regular intervals. The variable, (ℓ) , is chosen to define the dummy index in the summation of the discrete Shah function because it resembles a spike; (ℓ) is better than (l) because it reminds you that the Shah function is involved; but use whatever you want. The Shah function written this way differs from objects like the DFT and the convolution in that, in contrast these objects, the Shah function does not involve a summation over the grid axis.

According to my construction the Shah function should be thought of as targeting a specific grid, which is specified by its argument. An important feature of this construction is that an accompanying summation is needed for $\text{III}[n]$ to reach its target. Think of using the target grid axis as a lure in the definition of III that ultimately hooks into the summation coming from another object, ie: the DFT, or the convolution. By performing the summation coming from an auxiliary function, the Shah function effectively supplants the index with the dummy index. Ultimately, while the survival of this dummy index presents a source of complexity, it carries the identity of the axis it targets, making it trivial to rename in the end.

The DFT of the Discrete Shah Function (...Osgood notation sux...adapt yours!...)

THINK: it's time to get some clarity about the choice of notation... Although, the DFT is defined as a series over $[n]$ it is notated as a function of $[k]$, as in $\mathcal{F}[k] = \sum_n e^{-i2\pi kn/N}$

*****COME BACK: FIX to include proper definition of III and be explicit with def of DFT...

The DFT of the discrete Shah function is

$$(\mathcal{F} \text{III}_d[m])[k] = \sum_{n=0}^{N/d-1} e^{-i2\pi kn/N} \delta[m - nd]$$

The exponential only pops up on nd intervals

$$\begin{aligned} &= \sum_{n=0}^{N/d-1} e^{-i2\pi k(nd)/N} \\ &= \sum_{n=0}^{N/d-1} (e^{-i2\pi kd/N})^n \end{aligned}$$

For $kd = 0 \bmod N$, and for $k = 0 \bmod N/d$, for ie: $k = 0, N/d, 2N/d, 3N/d, \dots \in [0 : N - 1]$

$$\begin{aligned} (\mathcal{F} \text{III}_d)[k = 0 \bmod N/d] &= \sum_{n=0}^{N/d-1} w_N^{-(k=zN/d)dn} \quad , \text{ where } z \in \mathbb{Z} \\ &= \sum_{n=0}^{N/d-1} (1)^n = \frac{N}{d} \end{aligned}$$

For $kd \neq 0 \bmod N$, for and $k \neq 0 \bmod N/d$

$$\begin{aligned} &= \sum_{n=0}^{N/d-1} w_N^{kdn} = \frac{w^0 - w^{-kd(N/d)}}{1 - w^{-kd}} \\ &= \frac{1 - e^{-i2\pi kd/N(N/d)}}{1 - e^{-i2\pi kd/N}} = \frac{(1 - e^{-i2\pi k} = 0)}{1 - e^{-i2\pi kd/N}} \end{aligned}$$

Collecting these cases the DFT of the Shah function of spacing (b) is

$$(\mathcal{F} \text{III}_d)[k] = \begin{cases} N/d & \text{for } k = 0 \bmod N/d \\ 0 & \text{for } k \neq 0 \bmod N/d \end{cases}$$

, where k is any set of N consecutive integers. The result is a uniform set of (N/d) -magnitude values spaced at regular (N/d) -sized intervals along the axis, $[k]$. While not formally defined as a spike train of Kronecker deltas, this function is nonetheless in essence the discrete Shah function that has been scaled both vertically and horizontally (analagous to the continuous case ref. formula)

$$\boxed{(\mathcal{F} \text{III}_d)[k] = \left(\frac{N}{d}\right) \text{III}_{(N/d)}} \quad (25)$$

The IDFT of the Discrete Shah Function

COME BACK: use duality (and the symmetry of the Shah function)

$$\mathcal{F}f^{(-)} = (N) (\mathcal{F}^{-1}f)$$

$$\mathcal{F}^{-1}f = \frac{1}{N} \mathcal{F}f^{(-)}$$

...

$$(\mathcal{F}^{-1} \text{III}_d[\ell])[n] = \frac{1}{N} \sum_{k=0}^{N/b-1} e^{+i2\pi kn/N} \delta[\ell - kb]$$

- use the subscript b for the discrete Shah function in the frequency domain to suggest some “(b)andwidth” between samples.

$$\text{III}[\ell]_b = \sum_{k=0}^{N/b-1} \delta[\ell - kb]$$

$$\mathcal{F}^{-1} \text{III}_d = \frac{1}{N} \mathcal{F} \text{III}_d^{(-)}$$

...By the symmetry and cyclic periodidcity of III ...

$$\begin{aligned} &= \frac{1}{N} \mathcal{F} \text{III}_d \\ &= \frac{1}{N} \left(\frac{N}{d}\right) \text{III}_{(N/d)} \\ &= \left(\frac{1}{d}\right) \text{III}_{(N/d)} \end{aligned}$$

12.2 Derivation of the Discrete Sampling Theorem

Using the definitions for the DFT and the IDFT

$$F[k] = \mathcal{F}f = \sum_n f[n] w_N^{-kn} \quad \text{and} \quad f[n] = \mathcal{F}^{-1}F = \frac{1}{N} \sum_k F[k] w_N^{kn}$$

Following the derivation of the classical sampling theorem, the DFT is equivalent to

$$\mathcal{F}f = \Pi_{\Omega} (\mathcal{F}f * \mathbb{I}_{\Omega})$$

the periodization of the DFT, truncated by a boxcar function at the bandwidth. To recover the signal, take the IDFT

$$f = \mathcal{F}^{-1}(\Pi_{\Omega} (\mathcal{F}f * \mathbb{I}_{\Omega}))$$

By the convolution theorem for the aforementioned definitions of the DFT pair

$$\mathcal{F}^{-1}(fg) = (\mathcal{F}^{-1}f) * (\mathcal{F}^{-1}g)$$

the IDFT of the product is equal to the convolution of the IDFTs

$$= \mathcal{F}^{-1}(\Pi_{\Omega}) * \mathcal{F}^{-1}(\mathcal{F}f * \mathbb{I}_{\Omega})$$

By (discrete) duality

$$\mathcal{F}^{-1}(f^{(-)}) = \frac{1}{N} (\mathcal{F}f)$$

and because Π_{Ω} is symmetric, the IDFT of the boxcar function is

$$(\mathcal{F}^{-1}\Pi_{\Omega})[n] = \frac{1}{N} \mathcal{F}\Pi_{\Omega}^{(-)} = \frac{1}{N} \mathcal{F}\Pi_{\Omega} = \frac{1}{N} \mathcal{B}_{\Omega}[n]$$

NOTE: - (p.449[4]) Osgood claims this is where the reasoning comes in for using the convention of modeling the discontinuities / setting the edges of the boxcar to 1/2; so that a factor of p becomes available (here, equivalent to Ω) to cancel with a factor of p that comes from the expression right of the convolution sign involving the DFT of \mathbb{I}_{Ω} . - In my derivation this factor is absent...so it seems appropriate to use which ever version of the boxcar function. - maybe instead of trying to write a discrete version of the sinc function, a better convention/notation would be to refer to a generic DFT of the boxcar function using a notation like, ie: \mathcal{B} , \mathcal{B}_{Ω} , \mathfrak{B} . By the (discrete) convolution theorem

$$\mathcal{F}^{-1}(f) \mathcal{F}^{-1}(g) = \frac{1}{N} \mathcal{F}^{-1}(f * g)$$

the IDFT of the DFT convolved with \mathbb{I}_{Ω} - the right side of the outer convolution sign - is

$$\mathcal{F}^{-1}(\mathcal{F}f * \mathbb{I}_{\Omega}) = (N) \left(\mathcal{F}^{-1}(\mathcal{F}f) \right) \left(\mathcal{F}^{-1}(\mathbb{I}_{\Omega}) \right) = (N) f[n] \cdot (\mathcal{F}^{-1}\mathbb{I}_{\Omega})[n]$$

where the IDFT of the \mathbb{I} function is notated explicitly as a function of $[n]$.

Rewriting the signal construction as a convolution based on the two aforementioned formulas

$$\begin{aligned} f &= \left(\frac{1}{N} \mathcal{B}_{\Omega} \right) * \left((N) f \cdot (\mathcal{F}^{-1}\mathbb{I}_{\Omega}) \right) \\ &= \mathcal{B}_{\Omega} * \left(f \cdot (\mathcal{F}^{-1}\mathbb{I}_{\Omega}) \right) \end{aligned}$$

Writting the IDFT of the \mathbb{I}

$$= \mathcal{B}_{\Omega} * \left(f \cdot \frac{1}{\Omega} \mathbb{I}_{(N/\Omega)} \right)$$

By the commutativity of the convolution

$$= \frac{1}{\Omega} (f \cdot \mathbb{I}_{(N/\Omega)})[n] * \mathcal{B}_{\Omega}[n]$$

Under the given definitions of the DFT pair, the formula for the discrete convolution is

$$(f * g)[m] = f[n] * g[n] = \sum_n f[n] g[m - n]$$

where the dummy index $[m]$ references the same grid as $[n]$ through its juxtaposition to $[n]$. Notice, written out, the signal is a function of the dummy variable, $[m]$, used to define the convolution

$$\begin{aligned} f[m] &= \frac{1}{\Omega} \sum_n (f \cdot \text{III}_{(N/\Omega)})[n] \mathcal{B}_\Omega[m - n] \\ &= \frac{1}{\Omega} \sum_n \left(f \cdot \sum_\ell \delta \left[n - \frac{N}{\Omega} \ell \right] \right) \mathcal{B}_\Omega[m - n] \\ &= \frac{1}{\Omega} \sum_\ell \sum_n f[n] \mathcal{B}_\Omega[m - n] \delta \left[n - \frac{N}{\Omega} \ell \right] \end{aligned}$$

Performing the summation over $[n]$ swaps out the index $[n]$ for $[(N/\Omega)\ell]$

$$= \frac{1}{\Omega} \sum_\ell f \left[\frac{N}{\Omega} \ell \right] \mathcal{B}_\Omega \left[m - \frac{N}{\Omega} \ell \right]$$

and (m) is no longer a simple index because of its juxtaposition now with $[(N/\Omega)\ell]$, formerly $[n]$.

The $[\ell]$ -index was used by III to target the $[n]$ -index, so at this stage the index $[\ell]$ can be swapped out for its alias, $[n]$

$$f[m] = \frac{1}{\Omega} \sum_n f \left[\frac{N}{\Omega} n \right] \mathcal{B}_\Omega \left[m - \frac{N}{\Omega} n \right]$$

The variable (m) can be recast as an integer index $[m]$ by recognizing that the signal reconstruction is a function of the quantity $[(N/\Omega)m]$

$$\boxed{f \left[\frac{N}{\Omega} m \right] = \frac{1}{\Omega} \sum_n f \left[\frac{N}{\Omega} n \right] \mathcal{B}_\Omega \left[\frac{N}{\Omega} (m - n) \right]} \quad (26)$$

where $[m]$ indexes over the same range as $[n]$ (and given the reciprocal relation $N = L\Omega$, $L = N/\Omega$). Recall, L and Ω are determined by available memory and the sampling apparatus, respectively. This formula provides the essential criteria for checking the fidelity of the signal reconstruction based on sampling at a given bandwidth. Notice this construction does not reference the sinc function or even to a discrete version of the sinc function, as in ie: Osgood[4], and others ... The subtle advantage of this construction is that \mathcal{B} need not refer to the DFT of the box function, but can refer to the DFT of any window function that is applied to the signal.

NOW YOU HAVE TO CHECK THIS!!! This formula is very important for understanding aliasing, upsampling, downsampling in DSP.

Compare this result with formulas by others:

Brad Osgood[4] A discrete sampling theorem (p.449)

$$f[m] = \sum_{k=0}^{p-1} f \left[\frac{Nk}{p} \right] \text{dinc}_p \left[m - \frac{Nk}{p} \right]$$

Briggs and Henson[5] Theorem 3.2 Shannon Sampling Theorem. “Let f be a band-limited function whose Fourier transform is zero outside of the interval $[-\Omega/2, \Omega/2]$. If Δx is chose so that

$$\Delta x \leq \frac{1}{\Omega}$$

then f may be reconstructed exactly from its samples $f_n = f(n\Delta x) = f(x_n)$ by

$$f(x) = \sum_{n=-\infty}^{\infty} f_n \operatorname{sinc}\left(\frac{\pi(x - x_n)}{\Delta x}\right)$$

$$f(x) = \Delta x \sum_{n=-\infty}^{\infty} f_n \frac{\sin(\pi(x - x_n)/\Delta x)}{\pi(x - x_n)}$$

...what i was missing before was the IDFT of the Sha function, AND a actual definition/implementation of the convolution...COME BACK AND FINISH UP! NOTE: this derivation differs from Osgood's - Osgood defines the dinc function with a factor of $1/p$ - for me, $1/\Omega$ - which allows him to write the discrete analog to the continuous case as $p\operatorname{dinc}_p$, but this is really meaningless as it amounts to multiplyig by p/p . - ... this factor of p on the dinc function gets canceled with the $1/p$ associated with the IDFT of the Sha function $\operatorname{III}_p = (1/p)\operatorname{III}_{N/p}$. + But, in my derivation, I simply keep the p -factor in the IDFT of III_p + there are nonetheless differences between our respective derivations... + Osgood does propose a check in the end ... p.449[4] is very important!

13 Errors in the DFT and IDFT

A question that comes up in Briggs and Henson wrt errors is, when are we approximating the Fourier coefficients v approximating the Fourier transform? Recall the two expressions differ by a factor of $1/L$ (p.53-59[5]).

13.1 Roundoff Error

...section(1.4.1) in Van Loan's book on computational Frameworks for the FFT[23]

13.2 Errors in the DFT

Briggs and Henson present this topic through the three different scenarios: 1. (discretization error) the signal is L -periodic, and use the DFT to approximate the Fourier series coefficients. 2. (discretization error) the signal is spatially limited, ie: $f(x) = 0$ when $|x| > L/2$, and use the DFT to approximate the Fourier transform (one frequency at a time for the frequencies permissible for given L and N). 3. (truncation error) the signal is not time-limited, ie: $f(x) \neq 0$ for some $|x| > L/2$, and the DFT is used to

Important intuition about errors in the DFT comes from the Discrete Poisson Summation Formula[5], which is derived by taking the DFT of the Fourier series of f (move this derivation here...?)

13.2.1 (beta) Leakage

Leakage refers to the effect when "frequency components that are not present in the original waveform to leak into the DFT"(p.98[5]).

"leakage occurs when a periodic signal is truncated and sampled on an interval that is not an integer multiple of the period"(prob. 59, 99[5])

13.3 (beta) DFT Discretization Error

Discretization error is a result of finite sample size, ie: it arises because it is only possible to take a finite number of sample points. To investigate discretization error consider taking $N \rightarrow \infty$

$$F_k = \frac{1}{N} \sum_n f_n e^{-i2\pi kn/N}$$

Consider calculating the Fourier coefficients THINK: don't want to send all exponential terms to 0, so make substitution

$$n/N = x_n/L \quad \text{given} \quad x_n = n\Delta x = nL/N$$

to remove N from the exponential terms

$$F_k = \frac{1}{N} \sum_n f_n e^{-i2\pi kx_n/L}$$

To eliminate reference to the number of samples N except in the summation, multiply by L such that

$$LF_k = \frac{L}{N} \sum_n f_n e^{-i2\pi kx_n/L} = \Delta x \sum_n f_n e^{-i2\pi kx_n/L}$$

Now take the limit as $N \rightarrow \infty$, and $\Delta x \rightarrow 0$

$$\begin{aligned} \lim_{\substack{N \rightarrow \infty \\ \Delta x \rightarrow 0}} LF_k &= \lim_{\substack{N \rightarrow \infty \\ \Delta x \rightarrow 0}} \Delta x \sum_n f_n e^{-i2\pi kx_n/L} \\ &= \int_L f_n e^{-i2\pi kx_n/L} = L \underbrace{\left(\frac{1}{L} \int_L f_n e^{-i2\pi kx_n/L} \right)}_{a_k} = La_k \end{aligned}$$

Which implies

$$\lim_{\substack{N \rightarrow \infty \\ \Delta x \rightarrow 0}} F[k] = a_k$$

The DFT approximates the Fourier coefficients with an increasing number of samples. There is a discretization error introduced by the non-zero albeit small Δx , which can be reduced with increasing number of samples over L .

The DFT does not become the Fourier transform as $N \rightarrow \infty$ (p.54 [5]).

13.4 (beta) DFT Truncation Error

Truncation error arises working with non-time-limited signals due to the fact that the DFT is inherently defined over a finite boundary. Therefore, want to consider the limit as $L \rightarrow \infty$ (p.54[5]).

$$\begin{aligned} \lim_{\substack{N \rightarrow \infty \\ \Delta x \rightarrow 0}} L(\mathcal{F}f)[k] &= \lim_{\substack{N \rightarrow \infty \\ \Delta x \rightarrow 0}} \Delta x \sum_n f_n e^{-i2\pi s_k x_n} \\ &= \int_L f(x) e^{-i2\pi s_k x} dx \end{aligned}$$

, which is an approximation to the Fourier transform

$$\approx \int_{-\infty}^{\infty} f(x) e^{-i2\pi s_k x} dx = \hat{f}(s_k)$$

Taking the limits $L \rightarrow \infty$, and concomitantly $\Delta s \rightarrow 0$

$$\lim_{\substack{L \rightarrow \infty \\ \Delta s \rightarrow 0}} \int_L f(x) e^{-i2\pi s_k x} dx = \hat{f}(s_k)$$

13.5 Pester 1: $M = N/2$ (Periodic Band-limited Signals)

Recall, if f is piecewise smooth over L then its Fourier series

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kx/L}$$

converges to its periodic extension, and to the average value of its neighbors at points of discontinuity. And, the Fourier coefficients are given by

$$c_k = \frac{1}{L} \int_L f(x) e^{i2\pi kx/L} dx$$

where the k^{th} -mode has k periods/wavelengths over L .

Pester: $M = N/2$ (would this be considered an example of aliasing?) - recall: $N/2$ is the highest frequency the DFT can resolve.

Note: Briggs and Henson always work with the symmetric interval, $-L/2, L/2$.

Difference between N even v. odd (prob.126).

Case N even: $k = [-N/2 + 1, N/2] \in \mathbb{Z}$ DO: consider the continuous case, where there is no problem between the k and $-k$ indices. - Recall, to get the Fourier coefficients you need the positive and the negative k components. - *** the $-N/2$ and $N/2$ modes are indistinguishable due to the cyclic exponential in the DFT. - note: here, $-N/2$ is outside the set of indices. - this is the same as for indices 0 to N . - the same is true of any pair of indices separated by N number of indices. - Take away, $\Delta x \leq 1/\Omega$. - To satisfy the criteria of having two sample point per period

Take away: for an even number of samples N - $N = 2$ times the index of the Nyquist frequency ($\Omega/2$), is the minimum required to resolve the Nyquist frequency. - BUT: this is often not enough, 2 times this value, that is 4 times the index of the Nyquist frequency, is used in practice. - consider case of N = the index of the Nyquist frequency ("undersampling").. (prob.122 Briggs+Henson)

13.6 DFT Zero Padding - for mitigating error (which?)

p.56,...[5]

Zero padding is a technique for refining the frequency grid by artificially increasing the number of samples albeit all with value zero.

Zero padding for N' even, with n -index domain $[0, \dots, N-1]$

$$g[n] = \begin{cases} f[n] & \text{if } n \in [0, \dots, N-1] \\ 0 & \text{if } n \in [N, \dots, N'-1] \end{cases}$$

the N' -point DFT of $g[n]$ is

$$\begin{aligned} G_k &= (\mathcal{F}g)[k] = \frac{1}{N'} \sum_{n=0}^{N'-1} g[n] w_{N'}^{-kn} \\ &= \frac{1}{N'} \sum_{n=0}^{N-1} f[n] w_{N'}^{-kn} \\ &= \left(\frac{N}{N'}\right) \frac{1}{N'} \sum_{n=0}^{N-1} f[n] \exp\left(-i2\pi \frac{kn}{N'} \left(\frac{N}{N'}\right)\right) \\ &= \left(\frac{N}{N'}\right) \frac{1}{N} \sum_{n=0}^{N-1} f[n] \exp\left(-i2\pi \left(\frac{kN}{N'}\right) \left(\frac{n}{N}\right)\right) \\ &= \left(\frac{N}{N'}\right) (\mathcal{F}f)[kN/N'] = \left(\frac{N}{N'}\right) F_{kN/N'} \end{aligned}$$

13.7 Errors in the IDFT

Again, the IDFT is presented via 3 cases 1. (truncation error) the coefficients $F[k]$ are known, and f is over L , then the IDFT approximates the f at the sequence of points $\{x_n\}$ suffers from truncation error due to a finite number of frequency samples. 2. (discretization error) regard $F[k]$ as a set of samples of a band-limited ($\hat{f}(s) = 0$ for $|s| > \Omega/2$) signal, use the IDFT to approximate the inverse Fourier transform for $f(x_n) = f[n]$. 3. (discretization error and truncation error) non-band-limited signals are common and want to reconstruct the signal using the IDFT, then

$$\lim_{\substack{\Omega \rightarrow \infty \\ \Delta x \rightarrow 0}} \lim_{\substack{N \rightarrow \infty \\ \Delta s \rightarrow 0}} \Delta s F[k] = f[n]$$

- truncation error due to fundamental inability to take the limit $\Omega \rightarrow \infty$ in practice. - discretization error due to finite grid size used for the integral approximation.

13.8 (beta) IDFT Truncation Error

Note: def. band-limited (p.57[5]).

Truncation error occurs because the IDFT is a truncated version of the (infinite) Fourier series representation (p.57[5]), ie: it is only possible to sample a finite number of frequencies.

Consider f on a closed interval, and being given the Fourier coefficients, then the IDFT is

$$f[n] = \sum_k F_k e^{i2\pi kx/L}$$

approximates the Fourier series synthesis at individual points x_n .

13.9 (beta) IDFT Discretization Error

$$\Delta s f[n] = \Delta s \sum_k F_k e^{i2\pi kn/L} = \Delta s \sum_k F_k e^{i2\pi s_k x_n}$$

Taking the limit as $N \rightarrow \infty$

$$= \lim_{\substack{N \rightarrow \infty \\ \Delta s \rightarrow 0}} \Delta s \sum_k F_k e^{i2\pi s_k x_n}$$

Assuming f is band-limited

$$\begin{aligned} &= \int_{-\Omega/2}^{\Omega/2} \hat{f}(s) e^{i2\pi s x_n} ds \\ &= f(x_n) = f[n] \end{aligned}$$

14 Errors in the DFT

Investigate sequence of cases that cover all common DFT use cases. What are the sources of error in each case? "The form of the input sequence dictates how the DFT is used and how its output should be interpreted.

2 Cases: Use the DFT to ... 1. approx the Fourier coefficients of the input. 2. approx (samples of) the FT of the input (?).

def. "pester", ie: treatment of endpoints, even/od number of data points, ... - cannot be ignored. - contain a hidden lesson.

14.1 Periodic Band-limited Input and the Nyquist Frequency

Note: it is important to specify that we are referring to "functions" here, as these are idealizations - in practice sampling produces a discrete data set, the continuous model comes later. For periodic band-limited input the question is, how well does the DFT approximate the Fourier coefficients of f ? Assume f is piecewise smooth and periodic on $[-L/2, L/2]$, then Fourier series representation

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kx/L}$$

converges to f on the interval $[-L/2, L/2]$, and to its periodic extension beyond, and to its average value at points of discontinuity, (and how to handle endpoints?). And the coefficients are given by

$$c_k = \frac{1}{L} \int_{-L/2}^{L/2} dx f(x) e^{-i2\pi kx/L}, \quad k = 0, \pm 1, \pm 2, \dots$$

where the k -th "mode" specifies the number of periods/cycles/wavelengths on the interval $[-L/2, L/2]$.

For periodic f , given N (even) uniform samples, $x_n = n\Delta x = \frac{nL}{N}$, where $n = -N/2 + 1 : N/2$ (the sampled value assigned to f at the right endpoint ($f_{N/2}$) is the average of its endpoint values (itself and the first sample value?))

$$f_n = f(x_n) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kx_n/L} = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi kn/N}$$

Taking the DFT (for $k = -N/2 + 1 : N/2$)

$$\begin{aligned}
\mathfrak{F}(\{f_n\}) &= \frac{1}{N} \sum_{n=-N/2+1}^{N/2} f_n e^{-i2\pi kn/N} \\
&= \frac{1}{N} \sum_{n=-N/2+1}^{N/2} \left(\sum_{j=-\infty}^{\infty} c_j e^{i2\pi jn/N} \right) e^{-i2\pi kn/N} \\
&= \frac{1}{N} \sum_{j=-\infty}^{\infty} c_j \left(\sum_{n=-N/2+1}^{N/2} e^{i2\pi(j-k)n/N} \right) \\
&= \frac{1}{N} \sum_{j=-\infty}^{\infty} c_j N \hat{\delta}_N(j-k) \\
&= \sum_{j=-\infty}^{\infty} c_{k+jN} \\
&\boxed{F_k = c_k + \sum_{j=1}^{\infty} c_{k+jN} + c_{k-jN}}
\end{aligned}$$

Briggs and Henson refer to the last expression as the “Discrete Poisson Summation Formula” [5]. For bandlimited signals, the Fourier coefficients should be zero beyond the Nyquist frequency. Hence, the DFT reproduces the signal exactly, there is no error in the Fourier coefficients.

14.2 Periodic Non-band-limited Input and Aliasing

Now, when considering

$$F_k = c_k + \sum_{j=1}^{\infty} c_{k+jN} + c_{k-jN}$$

higher order components cannot be ignored. def. aliasing effect - higher order modes appear to contribute to lower frequency modes essentially because they agree at the data points. (prob.116, 117)

$$\begin{aligned}
e^{i2\pi s_k x_n} &= \exp\left(i2\pi \frac{k}{L} \frac{nL}{N}\right) = e^{i2\pi kn/N} \\
e^{i2\pi s_{k+jN} x_n} &= \exp\left(i2\pi \frac{k+jN}{L} \frac{nL}{N}\right) = e^{i2\pi kn/N} (1) = e^{i2\pi s_k x_n}
\end{aligned}$$

def. “smoothness” as number of continuous derivatives (refs. 30, 70, 115, 158). Theorem: rate of decay of Fourier coefficients (v1) - assm. for $p \geq 1$ the derivative of $f^{(p-1)}$ is periodic and continuous over L . - assm. $f^{(p)}$ bounded with a finite number of discontinuities over L . - then,

$$|c_k| \leq \frac{C}{|k|^p} \quad \forall k$$

- pf. (repeated integration by parts)...

def. piecewise monotone - L can be split (evenly?) into a finite number of subintervals over which f is nonincreasing or nondecreasing. Theorem: rate of decay of Fourier coefficients (v2) - assm. for

$p \geq 1$ the derivative of $f^{(p-1)}$ is periodic and continuous over L . - assm. $f^{(p)}$ bounded and piecewise monotone over L (note: $f^{(0)}$ is the function itself). - then,

$$|c_k| \leq \frac{C}{|k|^{p+1}} \quad \forall k$$

- pf. p integrations by-parts... - *finite number of discontinuities with infinite jumps. - *Dirichlet's conditions ref.30 - cover all conditions under which the theorem is true.

NOW: DFT Error for periodic non-bandlimited signals - $f^{(p-1)}$ for $p \geq 1$ are periodic and continuous over L . - $f^{(p)}$ bounded and piecewise monotone over L .

$$|F_k - c_k| \leq \frac{C}{N^{p+1}} \quad \forall k$$

pf. (really want ref.76, Henson's PhD thesis on image reconstruction.) Using the discrete Poisson summation formula

$$F_k - c_k = \sum_{j=1}^{\infty} (c_{k+jN} + c_{k-jN}) \quad \forall k$$

Using the previous theorem

$$\begin{aligned} F_k - c_k &\leq \sum_{j=1}^{\infty} \frac{C'}{|k+jN|^{p+1}} + \frac{C'}{|k-jN|^{p+1}} \\ &= \frac{C'}{N^{p+1}} \sum_{j=1}^{\infty} \frac{1}{|j + \frac{k}{N}|^{p+1}} + \frac{1}{|j - \frac{k}{N}|^{p+1}} \end{aligned}$$

The series converges for $p \geq 1$ (cite: Henson PhD thesis ref.76, or just do it!), hence

$$|F_k - c_k| \leq \frac{C}{N^{p+1}} \quad \forall k$$

Consider this formula under N large...(compare this formula with Table of DFT Errors)

How to evaluate "smoothness" based on discrete data? Maybe, approximate with polynomial to determine how many derivatives it has? SNAGS! p.189 I'm not sure I understand this section well! It's hard to see how important it really is But it might entail some important features of signals that are important to analyze!... we want to take derivatives!!! 1. ($p = 1$), (details/conditions, ie: $\cos(\pi x)$ sampled on $[-1/2, 1/2]$, error $\sim N^{-2}$, ie: real even harmonics and the triangle wave... 2. ($p > 1$), difficult to compute analytically, require numerical methods. 3. ($p = 0$), only the function f is bounded and piecewise monotone over L ... The Theorem above does not cover this case!!! - Cases: discontinuities either within the sequence or at the endpoints. - complex harmonic - real odd harmonic - linear - rectangular wave - square pulse - exponential - these cases follow asymptotic behavior:

$$F_k - c_k \sim CkN^{-2}$$

for all k as $N \rightarrow \infty$. * two regimes (p.189 [5]): case1: for $|k| \ll N/2$ low frequenc, then $E \sim N^{-2}$ case2: for $|k| \approx N/2$ low frequenc, then $E \sim N^{-1}$

14.3 Periodic Non-band-limited Input II

aliasing - occurs when the period/wavelength of a periodic function is greater than the sampling duration/extent. ie: $f(x) = \cos(\pi x)$ sampled on $[-1/2, 1/2]$ results in discontinuity in 1st derivative at the end points. This is the case of $p = 1$, mentioned above. Increasing the number of samples (N) cannot overcome the limit in sampling duration/extent. Higher frequency modes are said to be "aliased" onto (corresponding) lower frequency modes (might have to do the exercise to actually see this (cannot understand the terminology!)). (Should this still be called aliasing? maybe this is like high frequencies leaking IN TO the DFT calculation (v. leaking OUT)) obs: error for low-frequency coefficients is lower than for high-frequency coefficients.

14.4 Leakage in General

LIFE SAVING RESOURCE: <https://libgen.rs/>

NOTE: we should be able to answer “what is the difference between aliasing and leakage”? (would be a good YouTube video) Note: make YouTube videos that are problem-solution pairs.

Leakage is a source of error in the DFT that is in addition to aliasing error. Brigham in states “leakage ... is inherent in the DFT because of the required time-domain truncation” (p.103 [25]), but this is a bit of an over simplification. Leakage occurs when a periodic function is sampled over a duration that is not a multiple of the period of the signal or of its components. It should be appreciated that without prior knowledge this is generally the case. This is such an important point that deserves to be stated openly from the outset, that it is almost certain that the DFT calculation will suffer from leakage error.

E.g. assm. f is periodic over L with Fourier coefficients $\{c_k\}$ Consider taking N samples over the region $[-L/2, L/2]$, and pN sample over $[-pL/2, pL/2]$ where $p > 1 \in \mathbb{Z}$ (p.190[5])

$$c'_k = \begin{cases} c_{k/p} & \text{if } k/p \in \mathbb{Z} \\ \text{otherwise,} & \text{the frequency is not shared by the two sets of indices} \end{cases}$$

This obviously holds for the F'_k as well (prob.124). This isn't something deep it is simply a relationship between indices for the DFT over L and the DFT over pL . OBS: the variable p is being used in two different contexts in close proximity, previously to describe the smoothness of a function and here to describe the elongation of the sampling interval (editor pho-pa)!

If you must...demonstrate this in practice (ie: $f(x) = \cos(2x)$) ... $\cos(2x)$ is a difficult example to work with! ... How about a simpler function, ie: $f(x) = \cos(\pi x)$? ACTUALLY: stick with $\cos(2x)$ and supply x values. (this relationship is maintained for the DFT approximation of the Fourier transform) Note: this sounds like a way to refine the DFT calculation to make sure the frequencies you're sampling are valid.

THIS WHOLE SHIT IS SUPER IMPORTANT (IT IS THE DIFFERENCE BETWEEN CONTINUOUS AND DISCRETE) !!! THIS IS SUCH AN IMPORTANT EXERCISE because it is our first encounter with what it is like to work with a real example... Now consider the less ideal and more realistic situation. Assume f is L -periodic over the symmetric interval $[-L/2, L/2]$ (is this setup essential?). It therefore can be represented as a Fourier series with a set of coefficients, $\{c_k\}$. And sample N data points on region $[-L/2, L/2]$, and pN data points on $[-pL/2, pL/2]$ where $p > 1 \notin \mathbb{Z}$ (p.190[5]). Restricting the focus to a single mode

$$f(x) = e^{i2\pi k_1 x/L}$$

(MUST DERIVE:) The Fourier coefficients of f on $[-pL/2, pL/2]$ are (eqn. 6.5, p.191 [5])

$$c'_k = \frac{\sin \pi(k - pk_1)}{\pi(k - pk_1)}$$

None of the coefficients c'_k vanish because $(k - pk_1) \notin \mathbb{Z}$ (even if f is band-limited). Coefficients with k nearest pk_1 are often the largest, with characteristic oscillatory “sidelobes” decaying as $(k - pk_1)^{-1}$ symptomatic of leakage. (MUST DERIVE: problem 125) The DFT coefficients (what's the difference between these and the Fourier coefficients?) suffers similarly in that the pN -point DFT is

$$F'_k = \frac{\sin(\pi(k - pk_1)) \sin(2\pi(k - pk_1)/pN)}{2pN \sin^2(\pi(k - pk_1)/pN)}$$

for $k = -pN/2 + 1 : pN/2$. *** I don't understand why arbitrary k modes are described with reference to some arbitrary mode k_1 associated with f over $[-L/2, L/2]$??? ***

Note: p could change N from even to pN odd, ie: $10 \times 1.5 = 15$. Want to stick to something simpler like $10 * 1.2 = 12$.

NOTICE: error is reported as $|c_k - F_k|$

START SIMPLE: We should be able to first do the continuous case. (Consider) One mode from an L -periodic function

$$f(x) = e^{i2\pi k_L x/L}$$

over the expanded region pL has a Fourier series representation with the Fourier coefficients

$$c'_k = \frac{\sin(\pi(k - k_L))}{\pi(k - k_L)}$$

Proof: Fourier's trick for recovering the coefficients in the complex exponential series representation

$$c'_k = \frac{1}{pL} \int_{pL} dx f(x) e^{-i2\pi kx/pL}$$

Substituting for f

$$\begin{aligned} c'_k &= \frac{1}{pL} \int_{pL} dx e^{i2\pi k_L x/L} e^{-i2\pi kx/pL} \\ &= \frac{1}{pL} \int_{pL} dx e^{-i2\pi(k - pk_L)x/pL} \\ &= \frac{1}{pL} \int_{pL} dx e^{-i2\pi(k - pk_L)x/pL} \\ &= \frac{1}{pL} \left[\frac{-pL}{i2\pi(k - pk_L)} e^{-i2\pi(k - pk_L)x/pL} \right]_{x=-pL/2}^{x=pL/2} \\ &= \frac{-1}{i2\pi(k - pk_L)} \left[\exp\left(-i2\pi \frac{(k - pk_L)}{pL} \left(\frac{pL}{2}\right)\right) - \exp\left(-i2\pi \frac{(k - pk_L)}{pL} \left(-\frac{pL}{2}\right)\right) \right] \\ &= \frac{1}{i2\pi(k - pk_L)} \left(e^{i\pi(k - pk_L)} - e^{-i\pi(k - pk_L)} \right) \\ &= \frac{1}{\pi(k - pk_L)} \left(\frac{e^{i\pi(k - pk_L)} - e^{-i\pi(k - pk_L)}}{i2} \right) \\ &= \frac{\sin(\pi(k - pk_L))}{\pi(k - pk_L)} \end{aligned}$$

NOTE: What about the asymmetric region $x \in [0, L]$.

And then the discrete case

$$\begin{aligned} f(x) &= e^{i2\pi k_L x/L} \\ f_n &= e^{i2\pi k_L n \Delta x/L} = e^{i2\pi k_L n (L/N)/L} = e^{i2\pi k_L n/N} \\ F'_k &= \frac{\sin(\pi(k - pk_L)) \sin(2\pi(k - pk_L)/pN)}{2pN \sin^2(\pi(k - pk_L)/pN)} \end{aligned}$$

Proof: (This thing is called a “geometric progression” and has the trig-representation above) (This result is also related to the “Dirichlet kernel”, aka the aliased sinc function) (Formulas with this structure occur repeatedly in applicaitons: p.267 + prob.156 (digital filters))

$$F'_k = \frac{1}{pN} \sum_{n=-pN/2+1}^{pN/2} \left(e^{i2\pi k_L n/N} \right) e^{-i2\pi kn/(pN)}$$

$$\begin{aligned}
&= \frac{1}{pN} \sum_n e^{i2\pi p k_L n / (pN)} e^{-i2\pi k n / (pN)} \\
&= \frac{1}{pN} \sum_n e^{i2\pi (p k_L - k) n / (pN)}
\end{aligned}$$

Use the formula for geometric sum with a -cyclic

$$\begin{aligned}
&\sum_{n=p}^{p+N-1} a^n = \frac{a^p - a^{p+N}}{1 - a} \\
&= \frac{1}{pN} \frac{e^{i2\pi (p k_L - k) (-pN/2+1) / (pN)} - e^{i2\pi (p k_L - k) (pN/2+1) / (pN)}}{1 - e^{i2\pi (p k_L - k) n / (pN)}}
\end{aligned}$$

convert $(p k_L - k)$ to $(k - p k_L)$

$$= \frac{1}{pN} \frac{e^{-i2\pi (k - p k_L) (-pN/2+1) / (pN)} - e^{-i2\pi (k - p k_L) (pN/2+1) / (pN)}}{1 - e^{-i2\pi (k - p k_L) n / (pN)}}$$

What about working backwards

$$F'_k = \frac{\sin(\pi(k - p k_L)) \sin(2\pi(k - p k_L)/pN)}{2pN \sin^2(\pi(k - p k_L)/pN)}$$

Simplify angle variables and remove the $\frac{1}{N}$ factor (since it comes from the DFT definition) to concentrate on converting the exponentials to sin functions.

$$F'_k = \frac{\sin(x) \sin(2x/N)}{2 \sin^2(x/N)}$$

Exercise: p.102 DFT over $-L/2$ to $L/2$ of boxcar from $-L/4$ to $L/4$ (...)

$$\begin{aligned}
F_k &= \frac{\sin(\pi k/2) \sin(2\pi k/N)}{2N \sin^2(\pi k/N)} \\
&\cos\left(\frac{\pi k}{2}\right) + 2\Re\left[\frac{1 - \omega_N^{kN/4}}{1 - \omega_N^k}\right] - 1 \\
&\cos\left(\frac{2\pi k N}{4}\right) + 2\Re\left[\frac{1 - e^{i2\pi k N/4}}{1 - e^{i2\pi k/N}}\right] - 1
\end{aligned}$$

SIDE: in thinking about this stuff I wonder if it is necessary to distinguish between Fourier coefficients in continuous v discrete settings...but my hunch is No, because the Fourier coefficients are obtained in the same way in both cases.

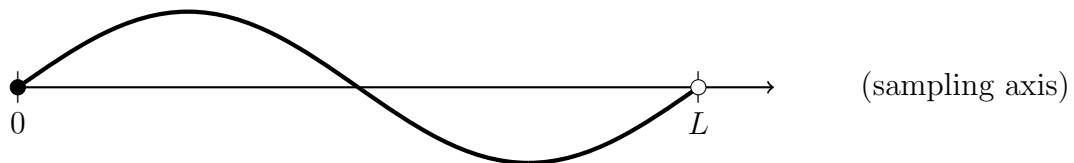


Figure 14: The fundamental mode, aka “one-mode” (red), which determines the grid spacing in the frequency domain $\Delta s = 1/L$. The dots indicate where this frequency is sampled during the present (black) and ensuing (hollow) sampling cycle.

All our knowledge comes from understanding simple examples. The DFT with even N computed over $-L/2, L/2$ of the square pulse defined over $-L/4, L/4$

$$f(x) = \begin{cases} 0 & , L/4 < |x| \leq L/2 \\ 1 & , |x| < L/4 \\ 1/2 & , x = \pm L/4 \end{cases}$$

is given by

$$F_k = \frac{1}{N} \sum_{n=-N/2+1}^{N/2} f_n \omega_N^{-kn}$$

It is easier to dissect the sum with the factor of N on the other side

$$NF_k = \frac{1}{2} \left(\omega_N^{-kN/4} + \omega_N^{kN/4} \right) + \sum_{n=-N/4+1}^{N/4-1} \omega_N^{-kn}$$

Collapsing the exponentials and folding the sum (minus 1 due to double counting the $n = 0$ for $\pm n$)

$$\begin{aligned} &= \cos(\pi k/2) + \sum_{n=0}^{N/4-1} (\omega_N^{-kn} \omega_N^{kn}) - 1 \\ &= \cos(\pi k/2) + 2\text{Re} \sum_{n=0}^{N/4-1} (\omega_N^{-kn} \omega_N^{kn}) - 1 \end{aligned}$$

Recognizing the geometric sum

$$= \cos(\pi k/2) + 2\text{Re} \left(\frac{1 - \omega_N^{kN/4}}{1 - \omega_N^k} \right) - 1$$

Briggs and Henson do not show their work but state from this expression the DFT simplifies to

$$F_k = \frac{\sin(\pi k/2) \sin(2\pi k/N)}{2N \sin^2(\pi k/N)}$$

, which applies for k from $-N/2 + 1$ to $N/2$ but not for $k = 0$ for which the DFT is

$$F_0 = \frac{1}{2}$$

$$\begin{aligned} F_k &= \cos(\pi k/2) + 2\text{Re} \left(\frac{1 - \omega_N^{kN/4}}{1 - \omega_N^k} \right) - 1 \\ F_k &= \cos(\pi k/2) + 2\text{Re} \left(\frac{1 - e^{i2\pi kN/(4N)}}{1 - e^{i2\pi k/N}} \right) - 1 \\ &= \cos(\pi k/2) + 2 \left(\frac{1 - \cos(2\pi kN/4N)}{1 - \cos(2\pi k/N)} \right) - 1 \\ &= \cos(\pi k/2) + 2 \left(\frac{1 - \cos(\pi k/2)}{1 - \cos(2\pi k/N)} \right) - 1 \\ &= \cos(\pi k/2) + 2 \left(\frac{1 - \cos(\pi k/2)}{1 - \cos(2\pi k/N)} \right) - 1 \end{aligned}$$

From here there are multiple ways to proceed. Simply proceeding algebraically without any trigonometry gives

$$\begin{aligned}
&= \cos(\pi k/2)(1 - \cos(2\pi k/N)) + 2 \left(\frac{1 - \cos(\pi k/2)}{1 - \cos(2\pi k/N)} \right) - (1 - \cos(2\pi k/N)) \\
&= \frac{1 - \cos(\pi k/2) + \cos(2\pi k/N) - \cos(\pi k/2) \cos(2\pi k/N)}{1 - \cos(2\pi k/N)} \\
&= \frac{(1 - \cos(\pi k/2))(1 + \cos(2\pi k/N))}{1 - \cos(2\pi k/N)}
\end{aligned}$$

From here it is not obvious how Briggs and Henson arrive at their equation

$$\frac{\sin(\alpha) \sin(2\beta)}{2 \sin^2(\beta)}$$

This actually seems like an abuse of the identity

$$1 - \cos^2 \theta = \sin^2 \theta$$

where they expand terms

$$(1 - \cos \theta)(1 + \cos \theta) = \sin \theta \sin \theta$$

but then simply replace one of θ values

$$(1 - \cos \theta)(1 + \cos \phi) = \sin \theta \sin \phi$$

but this is not math!

Use trigonometric identities

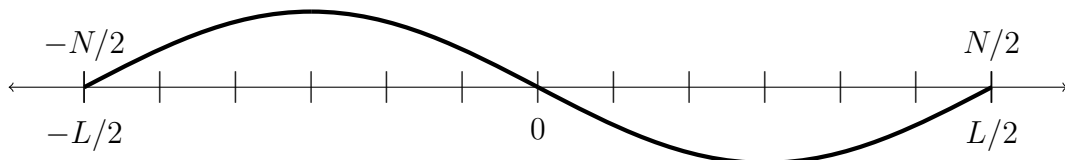
$$\begin{aligned}
&= \frac{(1 - \cos(\alpha))(1 + \cos(2\beta))}{1 - \cos(2\beta)} \\
&= \frac{(\cos^2(\alpha/2) + \sin^2(\alpha/2) - (\cos^2(\alpha/2) - \sin^2(\alpha/2))) (\cos^2(\beta) + \sin^2(\beta) + (\cos^2(\beta) - \sin^2(\beta)))}{\cos^2(\beta) + \sin^2(\beta) - (\cos^2(\beta) - \sin^2(\beta))} \\
&= \frac{(2 \sin^2(\alpha/2)) (2 \cos^2(\beta))}{2 \sin^2(\beta)} \\
&= \frac{2 \sin^2(\alpha/2) \cos^2(\beta)}{\sin^2(\beta)}
\end{aligned}$$

THIS IS WHERE YOU COME IN AND FIX THINGS:

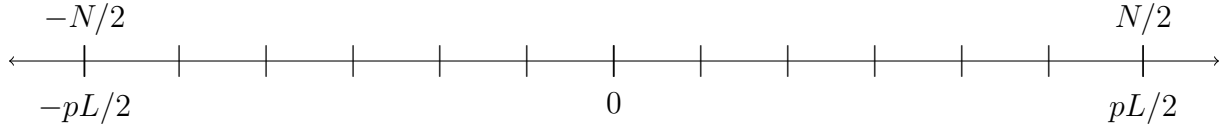
Consider $f(x)$ to be the cyclic complex exponential function (this exercise is based on Briggs and Henson's section on leakage in chapter 6 of The DFT book[5])

$$f(x) = e^{+i2\pi k'x/L}$$

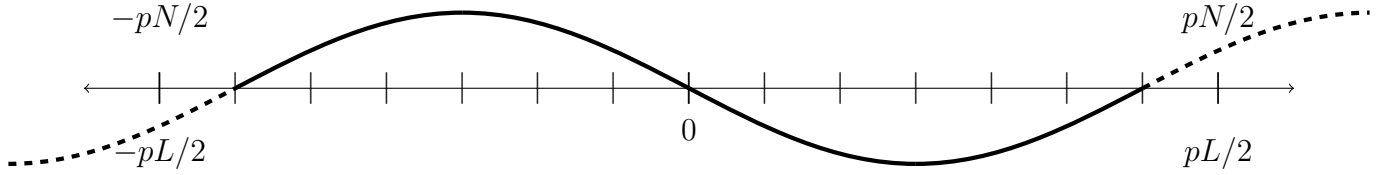
,a single mode in the Fourier series obtained from the N -point DFT, defined on the symmetric grid of width L .



In the case of leakage, the grid is potentially expanded such that the sampling extent is not a multiple of the full period of $f(x)$.



To ensure the frequency associated with $f(x)$ is accessible on the expanded grid the number of sample points must also be scaled (the set of frequencies in the N -point DFT over L and in the pN -point DFT over pL must overlap)



Consider the pN -point DFT of $f(x)$ on the expanded pL grid (ODD SYMMETRIC - simplest conceptual case)

$$\begin{aligned}\mathcal{F}[k] &= \sum_{n=-pN/2+1}^{pN/2} e^{-i2\pi kx/(pL)} \\ (\mathcal{F}f)[k] &= \sum_{n=-pN/2}^{pN/2} e^{i2\pi k'x/L} e^{-i2\pi kx/(pL)} = \sum_{n=-pN/2}^{pN/2} \exp\left(-i2\pi\left(\frac{k}{pL} - \frac{k'}{L}\right)x\right) \\ &= \sum_{n=-pN/2}^{pN/2} \exp\left(-i2\pi\left(\frac{k}{pL} - \frac{pk'}{pL}\right)x\right) = \sum_{n=-pN/2}^{pN/2} e^{-i2\pi(k-pk')x/(pL)}\end{aligned}$$

On the spatial (temporal) grid the position is $x = n\Delta x$, which on the expanded grid becomes $x = n\frac{pL}{pN} = nL/N$ and (((UPDATE THESE FORMULAS...)))

$$\begin{aligned}&= \sum_{n=-pN/2}^{pN/2} e^{-i2\pi(k-pk')\frac{(nL)}{N}\frac{1}{(pL)}} = \sum_{n=-pN/2}^{pN/2} e^{-i2\pi(k-pk')n/(pN)} \\ &= \sum_{n=-pN/2}^{pN/2} \left(e^{-i2\pi(k-pk')/(pN)}\right)^n = \frac{e^{-i2\pi(k-pk')(-pN/2)/(pN)} - e^{-i2\pi(k-pk')(pN/2+1)/(pN)}}{1 - e^{-i2\pi(k-pk')/(pN)}}\end{aligned}$$

Up to this point no conventions regarding end-points have been taken into account. On first attempt, every possible convenience will be made. First, try asserting the end points are equal

$$\begin{aligned}&= \frac{e^{i\pi(k-pk')} + e^{-i\pi(k-pk')}}{2} + \sum_{n=-pN/2+1}^{pN/2-1} \left(e^{-i2\pi(k-pk')/(pN)}\right)^n \\ &= \cos(\pi(k-pk')) + \frac{e^{-i2\pi(k-pk')(-pN/2+1)/(pN)} - e^{-i2\pi(k-pk')(pN/2)/(pN)}}{1 - e^{-i2\pi(k-pk')/(pN)}}\end{aligned}$$

Focusing on the term on the right and factoring the numerator and the denominator gives

$$\frac{e^{-i\pi(k-pk')/(pN)} \left(e^{-i2\pi(k-pk')(-pN/2+1)/(pN)+i\pi(k-pk')/(pN)} - e^{-i2\pi(k-pk')(pN/2)/(pN)+i\pi(K-pk')/(pN)} \right)}{e^{-i\pi(k-pk')/(pN)} \left(e^{i\pi(k-pk')/(pN)} - e^{-i\pi(k-pk')/(pN)} \right)} =$$

$$\frac{e^{-i\frac{2\pi}{(pN)}(k-pk')\left(-\frac{pN}{2}+1\right)+i\frac{\pi}{(pN)}(k-pk')} - e^{-i\frac{2\pi}{(pN)}(k-pk')\left(\frac{pN}{2}\right)+i\frac{\pi}{(pN)}(k-pk')}}{e^{i\pi(k-pk')/(pN)} - e^{-i\pi(k-pk')/(pN)}}$$

The denominator is $(i2) \sin(\pi(k - pk')/(pN))$. Focusing on the numerator

$$e^{i\frac{2\pi}{pN}(k-pk')(pN-2)+i\frac{2\pi}{2pN}(k-pk')} - e^{-i\frac{2\pi}{2pN}(k-pk')(pN)+i\frac{2\pi}{2pN}(k-pk')}$$

$$= e^{i\frac{2\pi}{2pN}(k-pk')(pN-2+1)} - e^{-i\frac{2\pi}{2pN}(k-pk')(pN-1)}$$

$$= e^{i\frac{\pi}{pN}(k-pk')(pN-1)} - e^{-i\frac{\pi}{pN}(k-pk')(pN-1)}$$

$$= (i2) \sin(\pi(k - pk')(pN - 1)/(pN))$$

Collecting these results

$$\mathcal{F}[k] = \cos(\pi(k - pk')) + \frac{\sin(\pi(k - pk')(pN - 1)/(pN))}{\sin(\pi(k - pk')/(pN))}$$

Combining terms

$$= \frac{\cos(\pi(k - pk')) \sin(\pi(k - pk')/(pN)) + \sin\left(\pi(k - pk')\left(1 - \frac{1}{pN}\right)\right)}{\sin(\pi(k - pk')/(pN))}$$

And using the trig identity $\sin(\theta - \phi) = \sin \theta \cos \phi - \cos \theta \sin \phi$

$$= \left(\cos(\pi(k - pk')) \sin\left(\frac{\pi}{pN}(k - pk')\right) + \sin(\pi(k - pk')) \cos\left(\frac{\pi}{pN}(k - pk')\right) \right. \\ \left. - \cos(\pi(k - pk')) \sin\left(\frac{\pi}{pN}(k - pk')\right) \right) / \sin\left(\frac{\pi}{pN}(k - pk')\right)$$

Say in words what this thing is... the DFT of the complex exponential defined (with fundamental mode of wavelength L) over the extent L computed over the extent pL (with modes with a fundamental wavelength of length pL)...say this better...

$$\boxed{\mathcal{F}[k] = \frac{\sin(\pi(k - pk')) \cos(\pi(k - pk')/(pN))}{\sin(\pi(k - pk')/(pN))}}$$

AHEAD:

an approach to deminishing leakage error is by adopting a different window, Hanning, Gaussian, etc. (p.609 [18]), to name a few.

Jerri puts a “conservative choice” for the sampling duration to $10 \times$ the wavelength of the fundamental frequency, or equivalently $10 \times$ the bandwidth(recall $\Omega L = N$)[18](4.H Deciding the Period for the Discrete Fourier Transform). Also: Chu also has advice on setting sampling duration (p.263, sec.8.6.2)[26].

Jerri NOTES[18]: Jerri defines a bandlimited function as being bandlimited between $(-a, a)$, and the "band limit" to be a (p.631 [18]). Note: Jerri points out that if the sampling duration is far from an integral multiple of the signal component period then the Hanning window helps less and less... ...Jerri's approach to dealing with leakage is to apply super-Gaussian (Jerri likes order equal to 6) windows (p.634 [18]), but even this approach is only moderately better. Also: Weaver p.134 on Super-Gaussian windows

Brigham NOTES[25]: sec.9.2 techniques for reducing leakage.

... AHEAD: adapt this for the digital context 1. consider of expanded grid pL to have expanded indices pN

Consider how the indexes change when expanding the axis and the grid points along with it... Consider the function $f(x) = e^{i2\pi k'x/L}$ And consider the DFT $\mathcal{F} = \sum_{n=0}^{N-1} e^{-i2\pi kn/N}$ Indices k' , k , n , and L or N need to be adjusted to formulate

$$\mathcal{F}f$$

15 Coding Preliminaries

Issues of Error, Accuracy, and Stability: floating-point, roundoff error, truncation error, stability, (Numerical Recipes section 1.1) math is hard ... ref[?].

Integer math is exact, division throws away the remainder.

Machines implement floating-point in their own way processor floating-point representation specified in standards IEEE 754-1985, IEEE 854-1987, IEEE 754-2008,

15.1 Coding Preliminaries: Floating-point Math

Resources: <https://floating-point-gui.de/> "What Every Programmer Should Know About Floating-Point Arithmetic!?! or Why don't my numbers add up?"

rubinhlандаu, Oregon State Univ., YouTube channel on Computational Physics single precision (32 bit): 6-7 decimal places of precision double precision (64 bit): 15-16 decimal places of precision

START: John D. Cook's blog - Anatomy of a floating point number (-doesn't use hyphen- not written as floating-point)

Normalized scientific notation in base 10 is when the **significand** is between $[1, 10)$, ie: $1.0 \times 10^2 = 100$. Normalized scientific notation in base 2 is when the **significand** is between $[1, 2)$. The 64 bit IEEE 754 double precision floating-point standard definition is

$$\begin{aligned} & (\pm)(1.\text{fraction}) \times (2)^{\text{exponent}} \\ & = (-1^{(s)})(2 - 2^{-(f)}) \times (2)^{(e)-1023} \end{aligned}$$

Indexing from the left, bit 1 (s) sets the sign (\pm), bits 2-12 (e) provide an exponent, and the remaining 52 bits (f) set the precision. Given the significand for normalized scientific notation always starts with 1, this bit (pun intended) of information does not need to be stored! Just the fractional portion of the significand, (1.fraction), is represented in storage. As a result, 52 bits can effectively be used to encompass 53 bits worth of precision (can you demonstrate this?).

$$x + 0 == x$$

$$\text{if}(x == y), \text{ then } x - y == 0$$

$$\frac{x}{\sqrt{x^2 + y^2}} \leq 1$$

15.2 Floating-point Error Correction

Section(5.6) in Numerical Recipes for how to handle round off error (ie: this is important for using the quadratic formula)

15.3 Floating-point Optimization: Julia

We want to write an implementation of the FFT for an embedded system...But first, some notes on floating point numbers and how operations on them are implemented by computers.

Notes from “Optimizing Floating Point Math in Julia” by Oscar Smith in JuliaCon 2022 YouTube video. Computers store numbers in binary scientific notation

$$x = s \cdot 2^e \cdot m$$

where s is the sign, the exponent e is a positive or negative integer, and m is a *mantisa* between 1 and 2. This format was standardized in IEEE 754 (in 1985). The two main sizes

- 32 bit: 4 billion values, “relative tolerance of $10e^{-7}$ ”(???)
- 64 bit: 16 quadrillion values, “about 16 decimal places of accuracy”

CPUs provide operations, ie: $+, -, *, /$ A **C** library exists called **Libm** that provides access to other functions. Every operating system has its own implementation of the math library. Linux version is slow, Windows version is broken, Apple version is good. Julia version has better cross platform compatibility *consider if computer has fused multiply add operations*

What the hard-ware provides is what is available... Avoid **division** and **if/else** (although you’ll probably need 1 or 2), because they’re slow. Core routine efficiently approximates functions and is fast on a modern computer. Polynomials are one approach, which can be evaluated quickly by Horner’s rule

$$a_0 + a_1x + a_2x^2 + a_3x^3 = a_0 + x(a_1 + x(a_2 + a_3x))$$

where an n -degree polynomial can be re-written as only n additions and n multiplications. Often only a few (relatively) terms of the polynomial are needed for approximation. Usually have exponential-ish convergence.

Taylor Series

$$f(x) \approx f(x_0) + xf'(x_0)^2 + \dots$$

Easy to compute coefficients (how?).

To get the optimal polynomial use Minmax polynomials, which are the polynomials with the minimum maximal error. They can be generated using Remez algorithm provided by Remez.jl package. Significantly more accurate than Taylor polynomials. Not as fast (don’t calculate at runtime). Let’s you chose the range over which you’re approximating.

Rational approximations generally have better worst-case error (a measure of accuracy) than the polynomial approximation of the same degree, but require division (which is slow) so the rational approximation generally needs to be good at about 3 - 4 degrees lower than the polynomial approximation to be worth the computational overhead, and generally this is not the case. But to approximate over a wide range the rational approximation may be more accurate. Remez.jl also provides rational functions, but proceed with caution.

Fused-multiply-add `fma` is your friend, computes $\text{fma}(a, b, c) = a \cdot b + c$, as fast as just a multiplication on most CPUs, and more accurately by only performing one rounding operation at the end. ...Really good for polynomials. `fma` is also useful for error correction/ implementing more accurate arithmetic, to compensate for all the error in your calculations, ie: $\text{fma}(a, -b, a \cdot b) = \text{the error in computing } a \cdot b$

Julia has `fma`, and `muladd`, which uses processor specific compiler heuristics chosen at initialization decide whether to fuse or not. To get the error properties of `fma` use `fma`. But if you only care about the performance then `muladd`. Using the `MuladdMacro.jl` package the compiler will turn all additions and multiplications into `muladd`'s, so you don't have to!

Next topic: Range reduction ...

16 Practical Implementation of the DFT - The FFT Algorithm

"The Fast Fourier Transform [is] the most valuable numerical algorithm in our lifetime" (Gil Strang, p.495, Linear Algebra Ed.3) Gil Strang uses the $w_N = e^{i2\pi/N}$ notation, so I should use it too!

when evaluating the algorithm think of time as the number of arithmetic operations, assume it takes constant time to multiply or add two real numbers.

It is fair to say that any good FFT algorithm I write *should* use N equal to some power of 2. This is almost sure to include some zero-padding...come back!

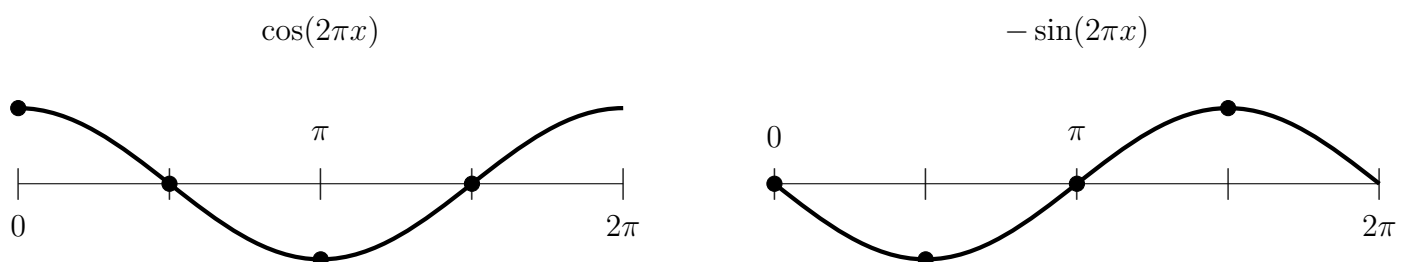
...This should be like a warm up, early encounter with the beast you're trying to tame...

The DFT matrix for $N = 4$ is a good starting example because it is easy to evaluate the matrix elements.

...building up the DFT matrix using

$$w = e^{-i2\pi} = \cos(2\pi) - i \sin(2\pi)$$

is evaluated at the points



with cycles repeating every 2π interval such that the $N = 4$ DFT matrix is

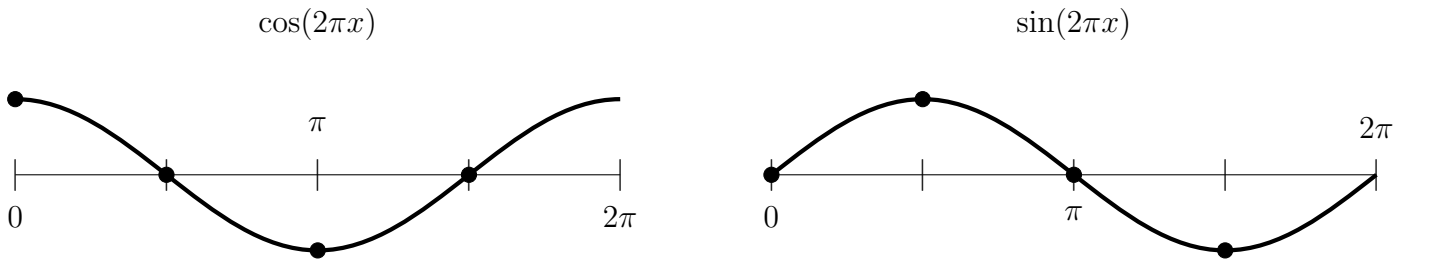
$$\begin{aligned}
W_4 &= \begin{bmatrix} w^{-(0)(0)/4} & w^{-(0)(1)/4} & w^{-(0)(2)/4} & w^{-(0)(3)/4} \\ w^{-(1)(0)/4} & w^{-(1)(1)/4} & w^{-(1)(2)/4} & w^{-(1)(3)/4} \\ w^{-(2)(0)/4} & w^{-(2)(1)/4} & w^{-(2)(2)/4} & w^{-(2)(3)/4} \\ w^{-(3)(0)/4} & w^{-(3)(1)/4} & w^{-(3)(2)/4} & w^{-(3)(3)/4} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w^{-1/4} & w^{-2/4} & w^{-3/4} \\ 1 & w^{-2/4} & w^{-4/4} & w^{-6/4} \\ 1 & w^{-3/4} & w^{-6/4} & w^{-9/4} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w^{-1/4} & w^{-1/2} & w^{-3/4} \\ 1 & w^{-1/2} & w^{-1} & w^{-(1+2/4)} \\ 1 & w^{-3/4} & w^{-(1+2/4)} & w^{-(2+1/4)} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w^{-1/4} & w^{-1/2} & w^{-3/4} \\ 1 & w^{-1/2} & w^{-1} & w^{-1}w^{-1/2} \\ 1 & w^{-3/4} & w^{-1}w^{-1/2} & w^{-2}w^{-1/4} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w^{-1/4} & w^{-1/2} & w^{-3/4} \\ 1 & w^{-1/2} & w^{-1} & w^{-1/2} \\ 1 & w^{-3/4} & w^{-1/2} & w^{-1/4} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix}
\end{aligned}$$

A check that this is indeed the correct matrix, note the columns should be linearly independent. The dot product of any two vectors should equal zero. Recall, for complex vectors the dot product is $\bar{a}^T b = a^H b$, as in a -transpose-complex-conjugate, or a -Hermitian dotted with b .

Sadly the definition for the DFT that I have used is not the one that is used in modern implementation... For reasons that will become clear in what follows, the DFT matrix is built up by the elements

$$w^{kn/N} = e^{i2\pi kn/N} = \cos\left(\frac{2\pi kn}{N}\right) + i \sin\left(\frac{2\pi kn}{N}\right)$$

For $N = 4$, the expression $w^{kn/N}$ gets evaluated at the grid points $0, \pi/2, \pi, 3\pi/2$



with cycles repeating every 2π interval such that the $N = 4$ DFT matrix is

$$\begin{aligned}
W_4 &= \begin{bmatrix} w^{(0)(0)/4} & w^{(0)(1)/4} & w^{(0)(2)/4} & w^{(0)(3)/4} \\ w^{(1)(0)/4} & w^{(1)(1)/4} & w^{(1)(2)/4} & w^{(1)(3)/4} \\ w^{(2)(0)/4} & w^{(2)(1)/4} & w^{(2)(2)/4} & w^{(2)(3)/4} \\ w^{(3)(0)/4} & w^{(3)(1)/4} & w^{(3)(2)/4} & w^{(3)(3)/4} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w^{1/4} & w^{2/4} & w^{3/4} \\ 1 & w^{2/4} & w^{4/4} & w^{6/4} \\ 1 & w^{3/4} & w^{6/4} & w^{9/4} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & w^{1/4} & w^{1/2} & w^{3/4} \\ 1 & w^{1/2} & w^1 & w^{1/2} \\ 1 & w^{3/4} & w^{1/2} & w^{1/4} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}
\end{aligned}$$

In modern implementations this is the DFT in its entirety. Note, in the notation of Numerical Methods, the definition of the DFT (1) does not have a $(-)$ negative sign associated with the complex exponential, (2) lacks a factor of N^{-1} , (and a minor detail is it) uses the index j (I use the index n). The fundamental step involved in defining the FFT algorithm, the Danielson and Lanczos *lemma* (Numerical Recipes[29]), is based on the recursive call to the DFT

$$F[k, n, N] = \sum_{n=0}^{N-1} f[n] e^{i2\pi kn/N} \quad \text{where } k, n \in \{0, 1, 2, \dots, N-1\}$$

, which can be broken up into two DFTs of half the length by separating the function/signal along the even and odd indices (after all, no DFT samples all points)

$$= \sum_{n=0}^{N/2-1} f[2n] e^{i2\pi k(2n)/N} + \sum_{n=0}^{N/2-1} f[2n+1] e^{i2\pi k(2n+1)/N}$$

, where the n -index, and concomitantly the k -index both run over the restricted domain, from ie: $[0, 1, 2, \dots, N/2-1]$ - indices involved in “splitting” the DFT are primed here to distinguish them from the originals, whilst referring to the same respective grids

$$\begin{aligned}
&= \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k'(2n')/N} + e^{i2\pi k'/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k'(2n')/N} \\
&= \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k' n'/(N/2)} + e^{i2\pi k'/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k' n'/(N/2)}
\end{aligned}$$

The k' values of the $N/2$ -point DFT

$$\{k' \mid 0, \dots, N/2-1\}$$

can be extended to recover the latter half of the k values

$$\{k \mid N/2, \dots, N-1\}$$

of the original N -point DFT via

$$k = k' + N/2$$

Continuing the DFT computation for $\{k \mid N/2, \dots, N-1\}$ with $k = k' + N/2$

$$F_{k'+N/2} = \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi(k'+N/2)n'/(N/2)} + e^{i2\pi(k'+N/2)/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi(k'+N/2)n'/(N/2)}$$

The complex exponentials of $N/2$ -point DFTs are $N/2$ -periodic in k'

$$= \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k' n'/(N/2)} + e^{i2\pi(k'+N/2)/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k' n'/(N/2)}$$

The complex exponential coefficient, $w^{i2\pi/N}$, on the odd DFT is N -periodic, so the $N/2$ additional rotation returns -1

$$\begin{aligned} &= \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k' n'/(N/2)} + e^{i2\pi(N/2)/N} e^{i2\pi k'/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k' n'/(N/2)} \\ &= \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k' n'/(N/2)} + e^{i\pi} e^{i2\pi k'/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k' n'/(N/2)} \\ &= \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k' n'/(N/2)} - e^{i2\pi k'/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k' n'/(N/2)} \end{aligned}$$

where $\{k' \mid 0, \dots, N/2-1\}$.

In summary, if the number of samples, N , is even, and n is indexed starting from 0 (Using the definition of the DFT given in Numerical recipes, lacks the $1/N$ factor, and the complex exponential is positive rather than negative (probably all for reasons related to computer computation))

$$(\mathcal{F}f)[k] = \begin{cases} \text{If } k = k' \in [0, 1, 2, \dots, N/2-1] \\ \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k' n'/(N/2)} + e^{i2\pi k'/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k' n'/(N/2)} \\ \text{If } k = k' + N/2 \in [N/2, \dots, N-1] \\ \sum_{n'=0}^{N/2-1} f[2n'] e^{i2\pi k' n'/(N/2)} - e^{i2\pi k'/N} \sum_{n'=0}^{N/2-1} f[2n'+1] e^{i2\pi k' n'/(N/2)} \end{cases}$$

where primed indices, $\{k', n' \mid 0, 1, 2, \dots, N/2-1\}$, distinguish those involved in the splitting process from those, $\{k, n \mid 0, 1, 2, \dots, N-1\}$, of the original DFT.

The splitting process continues recursively until the resulting series are of length 1, at which point we are done, “since the DFT of a sequence of length 1 is itself” (p.382 The DFT Book, []) (...understand this final statement...maybe do $N=4$).

The summation formulation makes the splitting process clear and intuitive, but is cumbersome to proceed with as the algorithm continues recursively. The fundamental step of splitting the DFT summation along even and odd index values can be written via factorization of, W , the DFT matrix

$$W_N f = \left(B_N \begin{bmatrix} W_{N/2} & 0 \\ 0 & W_{N/2} \end{bmatrix} P_N \right) f$$

And P is a permutation matrix that partitions the signal along even and odd indices

$$Pf = \begin{bmatrix} f_{\text{even}} \\ f_{\text{odd}} \end{bmatrix}$$

For example

$$P_4 f = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} a \\ c \\ b \\ d \end{bmatrix}$$

And B is often referred to as a butterfly matrix

$$B_N = \begin{bmatrix} I_{N/2} & \Omega_{N/2} \\ I_{N/2} & -\Omega_{N/2} \end{bmatrix}$$

where $\Omega_{N/2}$ is the, $N/2 \times N/2$, square matrix with the elements, ie: $\{w^{n/N}\}_{n=0}^{n=N-1}$, along the diagonal.

The DFT decomposition

$$W_N f = \left(B_N \begin{bmatrix} W_{N/2} & 0 \\ 0 & W_{N/2} \end{bmatrix} P_N \right) f$$

can be recursively applied to the DFT matrices of the middle box diagonal matrix of DFTs. Let's look at an example before going for something more general. For example, $N = 4$

$$\begin{aligned} W_4 f &= B_4 \begin{bmatrix} W_2 & 0 \\ 0 & W_2 \end{bmatrix} P_4 \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix} = B_4 \begin{bmatrix} \left(B_2 \begin{bmatrix} W_1 & 0 \\ 0 & W_1 \end{bmatrix} P_2 \right) & 0 \\ 0 & \left(B_2 \begin{bmatrix} W_1 & 0 \\ 0 & W_1 \end{bmatrix} P_2 \right) \end{bmatrix} \begin{bmatrix} f_0 \\ f_2 \\ f_1 \\ f_3 \end{bmatrix} \\ &= B_4 \begin{bmatrix} B_2 & 0 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} W_1 & 0 & 0 & 0 \\ 0 & W_1 & 0 & 0 \\ 0 & 0 & W_1 & 0 \\ 0 & 0 & 0 & W_1 \end{bmatrix} \begin{bmatrix} P_2 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} f_0 \\ f_2 \\ f_1 \\ f_3 \end{bmatrix} \\ &= B_4 \begin{bmatrix} B_2 & 0 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} W_1 & 0 & 0 & 0 \\ 0 & W_1 & 0 & 0 \\ 0 & 0 & W_1 & 0 \\ 0 & 0 & 0 & W_1 \end{bmatrix} \begin{bmatrix} f_0 \\ f_2 \\ f_1 \\ f_3 \end{bmatrix} \end{aligned}$$

The Fourier transform of length 1 - regardless of the DFT definition used - is the identity

$$\sum_{n=0}^{N-1} f[n] e^{i2\pi kn/N} = \sum_{n=0}^{1-1} f[n] e^{i2\pi k0/N} = \sum_{n=0}^0 f[n](1) = f[0](1)$$

Here, the signal/function is evaluated at an arbitrary point, $[n = 0]$ simply refers to the index given. The $N = 4$ DFT matrix decomposition simplified, or $N = 4$ FFT, is

$$= B_4 \begin{bmatrix} B_2 & 0 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} f_0 \\ f_2 \\ f_1 \\ f_3 \end{bmatrix}$$

Following this procedure, more generally

$$W_N f = B_N \begin{bmatrix} W_{N/2} & 0 \\ 0 & W_{N/2} \end{bmatrix} P_N f = B_N \begin{bmatrix} W_{N/2} & 0 \\ 0 & W_{N/2} \end{bmatrix} \begin{bmatrix} f_{\text{even}} \\ f_{\text{odd}} \end{bmatrix} = B_N \begin{bmatrix} F_{\text{even}} \\ F_{\text{odd}} \end{bmatrix}$$

Taking this further - even/odd gets abbreviated e/o - the DFT decomposition is applied to the DFTs along the signal/function at even and odd indices, respectively

$$F_e = B_{N/2} \begin{bmatrix} W_{N/4} & 0 \\ 0 & W_{N/4} \end{bmatrix} P_{N/2} f_e = B_{N/2} \begin{bmatrix} W_{N/4} & 0 \\ 0 & W_{N/4} \end{bmatrix} \begin{bmatrix} f_{ee} \\ f_{eo} \end{bmatrix} = B_{N/2} \begin{bmatrix} F_{ee} \\ F_{eo} \end{bmatrix}$$

$$F_o = B_{N/2} \begin{bmatrix} W_{N/4} & 0 \\ 0 & W_{N/4} \end{bmatrix} P_{N/2} f_o = B_{N/2} \begin{bmatrix} W_{N/4} & 0 \\ 0 & W_{N/4} \end{bmatrix} \begin{bmatrix} f_{oe} \\ f_{oo} \end{bmatrix} = B_{N/2} \begin{bmatrix} F_{oe} \\ F_{oo} \end{bmatrix}$$

Using these formulas to perform the factorization

$$\begin{aligned} &= B_N \begin{bmatrix} B_{N/2} & 0 \\ 0 & B_{N/2} \end{bmatrix} \begin{bmatrix} W_{N/4} & 0 & 0 & 0 \\ 0 & W_{N/4} & 0 & 0 \\ 0 & 0 & W_{N/4} & 0 \\ 0 & 0 & 0 & W_{N/4} \end{bmatrix} \begin{bmatrix} P_{N/2} & 0 \\ 0 & P_{N/2} \end{bmatrix} P_N f \\ &= B_N \begin{bmatrix} B_{N/2} & 0 \\ 0 & B_{N/2} \end{bmatrix} \begin{bmatrix} W_{N/4} & 0 & 0 & 0 \\ 0 & W_{N/4} & 0 & 0 \\ 0 & 0 & W_{N/4} & 0 \\ 0 & 0 & 0 & W_{N/4} \end{bmatrix} \begin{bmatrix} P_{N/2} & 0 \\ 0 & P_{N/2} \end{bmatrix} \begin{bmatrix} f_e \\ f_o \end{bmatrix} \\ &= B_N \begin{bmatrix} B_{N/2} & 0 \\ 0 & B_{N/2} \end{bmatrix} \begin{bmatrix} W_{N/4} & 0 & 0 & 0 \\ 0 & W_{N/4} & 0 & 0 \\ 0 & 0 & W_{N/4} & 0 \\ 0 & 0 & 0 & W_{N/4} \end{bmatrix} \begin{bmatrix} f_{ee} \\ f_{eo} \\ f_{oe} \\ f_{oo} \end{bmatrix} \\ &= B_N \begin{bmatrix} B_{N/2} & 0 \\ 0 & B_{N/2} \end{bmatrix} \begin{bmatrix} F_{ee} \\ F_{eo} \\ F_{oe} \\ F_{oo} \end{bmatrix} \end{aligned}$$

This procedure proceeds until the middle box diagonal matrix of DFTs is the identity matrix and all that remains is a sequence of butterfly matrices to the left, and permutation matrices on the right. The complete DFT factorization that leads to the FFT is (notation for the function/signal is omitted to fit the page)

$$W_N = B_N \begin{bmatrix} B_{N/2} & 0 \\ 0 & B_{N/2} \end{bmatrix} \cdots \begin{bmatrix} B_2 & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ \vdots & 0 & B_2 & 0 & \vdots \\ 0 & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \cdots & B_2 \end{bmatrix} I_N \begin{bmatrix} P_2 & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ \vdots & 0 & P_2 & 0 & \vdots \\ 0 & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \cdots & P_2 \end{bmatrix} \cdots \begin{bmatrix} P_{N/2} & 0 \\ 0 & P_{N/2} \end{bmatrix} P_N$$

Note, P_2 on a vector of length 2 leaves it unchanged, so the block diagonal matrix of P_2 's is redundant. (...) This is the linear algebra behind the radix-2 FFT algorithm. But what does Radix-2 refer to???

In pseudo code the algorithm can be broadly described in two stages, sometimes referred to as split and recombine

```
# Stage 1
x ← PTx
# Stage 2
for t = 1 : T
    x ← Atx
end
```


The splitting stage involves sorting the signal in preparation for matrix multiplication in the next stage.

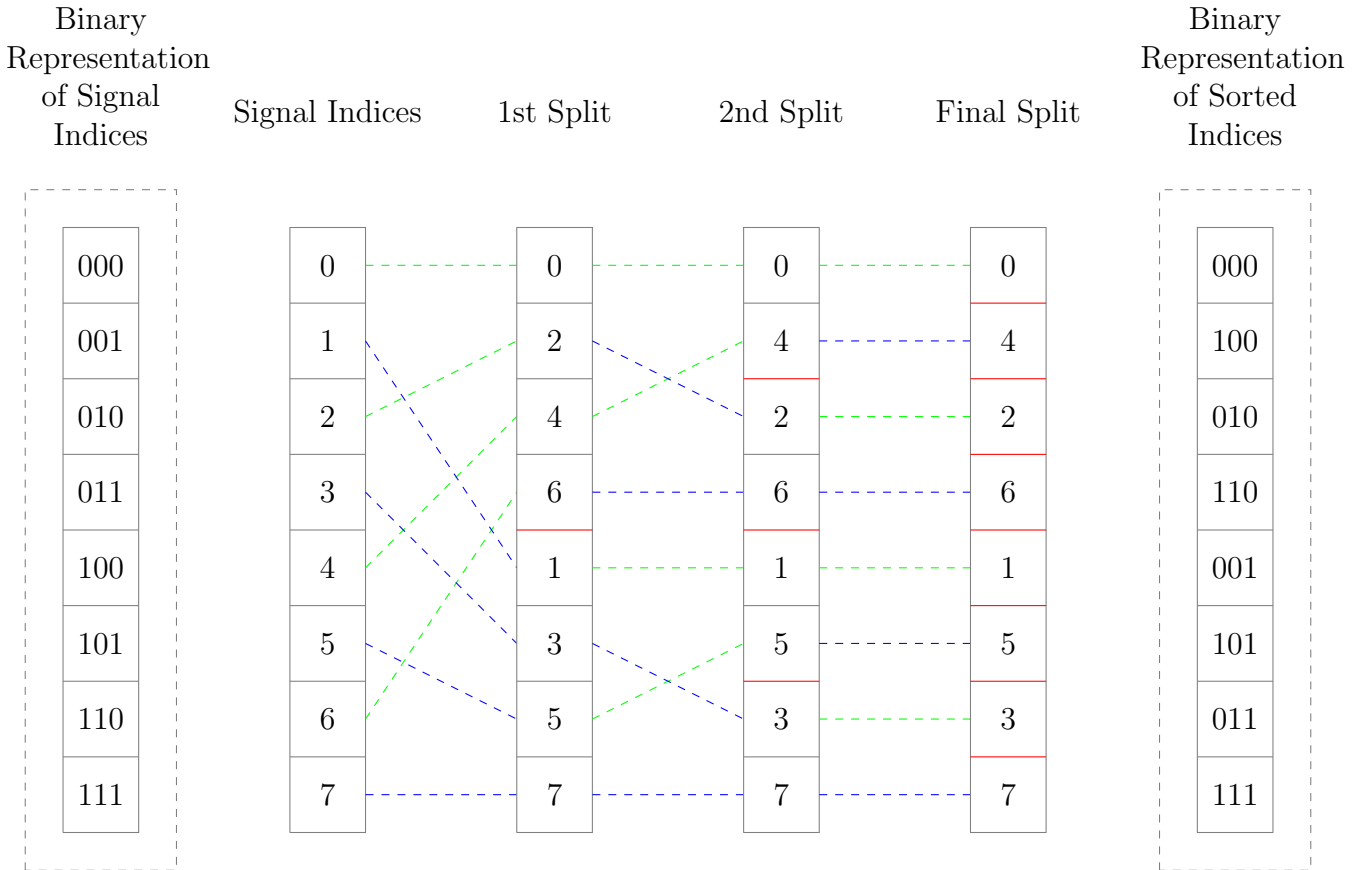


Figure 15: Depiction of the complete splitting process, from left to right, applied to a signal/function with $N = 8$, and $\{n \mid 0, \dots, N - 1\}$. The vertical columns represent the signal with samples organized from top to bottom. The n -index value is specified in each bin. **Red** lines partition the signal into even (above) and odd (below) index sequences. **Green** dashed lines indicate grouping of even indices. **Blue** dashed lines indicate grouping of odd indices. Notice, the third and final splitting step is redundant. Binary representation of the original, and final sorted indices are shown at the borders, in dashed boxes.

The fundamental splitting process amounts to multiplication by a permutation matrix, which is repeated until the signal remains unchanged by further splitting. This is performed a total of (T) times, where (T) is given by $2^T = N$, but the last splitting step is redundant.

In practice, the sorting process - successive grouping of even and odd sequences, preserving relative order - is efficiently handled by indexing samples in binary and using a procedure sometimes called *bit-reversal*. I'm not going to prove this, it's just amazing! Bit-reversal amounts to writing the binary representation with the reverse order of 1's and 0's. In figure 15, the indices when $N = 8$ are written in binary in the dashed box on the left, as well as for the sorted (post splitting) indices in the dashed box on the right. AHEAD: ref. Petr Rosel, Timing of some bit reversal algorithms, Signal Processing, Vol18, Issue4, 1989, 425-433.

Implementing bit-reversal in Julia. In Julia use `bitstring()` to get the full 64bit binary representation of a number. To perform bit reversal each binary index needs the same number of digits. The index value N determines the left leading zero padding for the index values less than N . Remove the leading zeros from the `bitstring()` representation so that all binary indices have the same number of digits.

```
function bitrev(N)
    skip = leading_zeros(N-1)+1 # no. left zero padding in bit-representation to omit
```

```

for i in 0:N-1
    println( parse{Int, reverse(bitstring(i)[skip : end])}, base=2) )
end
end

```

The combine stage of the FFT algorithm performs the sequence of butterfly matrix multiplications

$$B_N \begin{bmatrix} B_{N/2} & 0 \\ 0 & B_{N/2} \end{bmatrix} \cdots \begin{bmatrix} B_2 & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ \vdots & 0 & B_2 & 0 & \vdots \\ 0 & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \cdots & B_2 \end{bmatrix}$$

on the bit-reverse sorted sequence. Recall the butterfly matrix is

$$B_N = \begin{bmatrix} I_{N/2} & \Omega_{N/2} \\ I_{N/2} & -\Omega_{N/2} \end{bmatrix}$$

where $\Omega_{N/2}$ is the, $N/2 \times N/2$, square matrix with the elements, $\{w^{n/N}\}_n$, ie: $\{w^{n/N}\}_{n=0}^{n=N-1}$, along the diagonal.

The `LinearAlgebra` package provides, `I`, the identity matrix, and `Diagonal()` for constructing the butterfly matrix.

```

function butterfly(N)
    D = cispi.(2*[0:N/2-1...] / (N/2) )
    D = Diagonal(D)
    [I D; I -D]
end

```

To assemble the kaleidoscope (a group of butterflies) matrices

16.1 Discrete Spectral Decomposition

Recall (sec. 2.1, Fourier Series Using Complex Exponentials), both k and $-k$ are needed to describe the k -th mode.

$$\begin{aligned}
 f(x) &= \frac{A_0}{2} + \sum_{k=1}^{\infty} \left(A_k \cos(kx) + B_k \sin(kx) \right) \\
 &= \frac{A_0}{2} + \sum_{k=1}^{\infty} \left(A_k \frac{e^{ikx} + e^{-ikx}}{2} + B_k \frac{e^{ikx} - e^{-ikx}}{2i} \right) \\
 &= \frac{A_0}{2} + \frac{1}{2} \sum_{k=1}^{\infty} \left(A_k - iB_k \right) e^{ikx} + \frac{1}{2} \sum_{k=-1}^{-\infty} \left(A_{-k} + iB_{-k} \right) e^{ikx} \\
 &= \sum_{k=-\infty}^{\infty} C_k e^{ikx} \quad , \text{ where } \begin{cases} C_0 = A_0/2 & , \text{ if } k = 0 \\ C_k = \frac{1}{2}(A_k - iB_k) & , \text{ if } k > 0 \\ C_k = \frac{1}{2}(A_{-k} + iB_{-k}) & , \text{ if } k < 0 \end{cases}
 \end{aligned}$$

In the case where the index runs from 0 to N , the negative indices can be identified by the N -periodicity of the DFT. The index $k = N - 1$ corresponds to $k = -1$, which pertains to $k = 1$. The index $k = N - 2$ corresponds to $k = -2$, which pertains to $k = 2$. And so on.

16.2 Further notes on FFT implementation

done in-place - what does it mean for the FFT to be done in-place? ans. without the need for to instantiate additional arrays requiring additional storage. - this ultimately should avoid the reordering stage? - 1.6 The Cooley-Tukey Framework and 1.6.1 An In-Place Formulation

(description of decimation in frequency...amenable to more general values of N , and can also be done in place, with same computational cost as the Cooley-Tukey FFT)p.384-385 The DFT Book[5] - Gentleman and Sande 1966 ref64 The DFT Book - Stockham ref38, Glassman ref66 The DFT Book

HOW TO DO THIS??? p.386 - if you are calculating both the DFT and the IDFT then you can avoid the rearrangement step....! - by using decimation-in-frequency the for forward transform and decimation-in-time for the inverse transform

Is Van Loan[23] the best resource for the FFT? There is some linear algebra notation that gets introduced in the literature, the **Kronecker product** between matrices A_{pq} and B_{rs} , which produces the

$$pr \times qs$$

matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \cdots & a_{pq}B \end{pmatrix}$$

This notation allows us to write the DFT of the signal x as

$$W_N x = (I_1 \otimes B_N)(I_2 \otimes B_{N/2}) \cdots (I_{N/2} \otimes B_2) P^T x$$

where $N = 2^T$, P^T is the permutation (sorting/bit-reversal) matrix, and B is a butterfly matrix.

The “chirp transform algorithm” by Oppenheim and Schaffer 2010 is a FFT implementation for arbitrary N , ie: not a power of 2...But honestly not the place to start...zero padding and N -of some power of 2 is more practical implementation to begin with.

16.3 Zero Padding

COMEBACK: Jeremy Howard’s ai tutorials on youtube...

PRE: computer architecture bit widths, ie: 32, 64

On 64 bit architecture the largest number that should be possible to represent is 2^{64} . Let’s check this in Julia

```
julia> 2^64
0
julia> 2^63
-9223372036854775808
```

Julia uses machine integers(Two's Complement)/overflow/wrap-around/native arithmetic, because it's fast; and `SafeIntegers.jl` is one package for working around this.

Julia's out the box uses machine integers aka Two's complement: binary digit with greatest place value indicate positive or negative, this is called the "most significant bit" when 1 indicates negative. because it's fast. To see the binary representation of a number in Julia

[illegible]

Overflow - going beyond the max value - wraps around

```
julia> 2^63
-9223372036854775808
```

```
julia> -2^63
-9223372036854775808
```

[illegible]

To use this technique you must find the next highest power of 2. Consider: how your choice of definition of DFT affects this...

... perhaps one more thing to consider about the signal before we try to compute its DFT via the FFT is zero-padding, s.t. N is a power of 2.

the thing about zero-padding is that by modifying the signal you consequently modify the frequency grid also.

zero padding is done at the end of a sequence[4].

Check out section13p1p1 in Numerical Methods[?]

zero padding on the asymmetric grid without $1/N$ in def of DFT (p.464, sec7p11p4 Osgood) zero padding on the symmetric grid with $1/N$ in def of DFT (p.90, The DFT Book), here you can see why the $1/N$ factor makes things complicated...

17 Julia Code

Why use Julia: math looks like math, real code is better than pseudo code

- core ideas:
- the correct way to represent the signal is a vector
- the correct way to represent the DFT is a matrix

random signal

- use random number generator
- set N - length of vector
- return vector of data points

ie: N=10 signal of 32bit integers
 signal = rand(Int32, 10)

box function of given width

- set N - total no. data points
- set N - no. of data points for box
- set ON/OFF - set to 1/2 at discontinuities?
- set position - default center
- return a vector of data points.

The boxcar function

- N - set the number of samples
- b - the width of the box
- default to a centered box (update later)
- use fill() - equivalent to rep(0, 10) in R
- use semicolon for vertical concatenation, ie: [10; 15; 21]

```
function boxcar(N, b)
    [fill(0, div(N-b), 2); fill(1, b); fill(0, div(N-b), 2)]
end
# ...more simply
boxcar(N, b) = [fill(0, div(N-b), 2); fill(1, b); fill(0, div(N-b), 2)]
```

But...this is not good enough: cannot handle N odd

```
julia> boxcar(11, 2)
10-element Vector{Int64}:
```

```
0
0
0
0
1
1
0
0
0
0
```

this picks a convention that is aligned with the index range [N/2+1: N/2]

```
julia> boxcar(N, b) = [fill(0, div(N-b, 2, RoundDown)); fill(1, b); fill(0, div(N-b, 2, RoundUp))]
boxcar (generic function with 1 method)
```

```
julia> boxcar(11, 2)
11-element Vector{Int64}:
```

```
0
0
0
0
1
1
0
0
0
0
0
```

```

# *** MUST UPDATE *** boxcar() to use ones() and zeros() functions instead of fill ...

#####
# Calc DFT of cos(a*x) #
#####
N = 4

# generate N samples of the function cos() along the interval 2pi: method 1
julia> samples = [range(start=0, stop=3pi/2, length=4)...]
4-element Vector{Float64}:
 0.0
 1.5707963267948966
 3.141592653589793
 4.71238898038469
# notice the floating-point error (not exactly 0 where it should be at samples 2 and 4)
julia> cos.(samples)
4-element Vector{Float64}:
 1.0
 6.123233995736766e-17
 -1.0
 -1.8369701987210297e-16

# Notice: Julia struggles most when the answer should be 0
julia> x = [range(start=0, stop=3pi/2, length=N)...]
4-element Vector{Float64}:
 0.0
 1.5707963267948966
 3.141592653589793
 4.71238898038469

julia> cos.(2x)
4-element Vector{Float64}:
 1.0
 -1.0
 1.0
 -1.0

# To generate the function samples without the floating-point errors
#... using the axis indices
julia> x = [0:N-1...]
4-element Vector{Int64}:
 0
 1
 2
 3
#...scale the cospi() function as needed
julia> cospi.(2x/N)
4-element Vector{Float64}:
 1.0
 0.0

```

```

-1.0
 0.0
# Just to demonstrate this for cos() with a x-coefficient
julia> x = [0:N-1...]
4-element Vector{Int64}:
 0
 1
 2
 3
julia> cospi.(2*2*x/N)
4-element Vector{Float64}:
 1.0
-1.0
 1.0
-1.0
# NOTE: Perhaps the most disconcerting thing about floating-point errors for 0 is not being

# The DFT Matrix
# The complex exponential on pi intervals cispi() broadcast over the (k)x(n) matrix of indi
# Note the use of 0-based indexing convention
function dft_matrix(N)
    cispi.(2/N*[0:N-1...]*[0:N-1...])
end

# Compute the DFT of a Signal
function dft_compute(f)
    dft_matrix(length(f))*f
end

#####
# (OLD) Calc DFT of cos() #
#####
# set N
julia> N = 5
5
# def signal
julia> s = cos.(2*pi/N*[0:N-1...])
5-element Vector{Float64}:
 1.0
 0.30901699437494745
-0.8090169943749473
-0.8090169943749475
 0.30901699437494723
# define DFT k,n-indices matrix
julia> W = [0:N-1...] * [0:N-1...]'
5x5 Matrix{Int64}:
 0  0  0  0  0
 0  1  2  3  4
 0  2  4  6  8
 0  3  6  9 12
 0  4  8 12 16
# define DFT exponents from indices matrix
julia> W*=-im*2*pi/N

```

```

5x5 Matrix{ComplexF64}:
 0.0-0.0im  0.0-0.0im      0.0-0.0im      0.0-0.0im      0.0-0.0im
 0.0-0.0im  0.0-1.25664im  0.0-2.51327im  0.0-3.76991im  0.0-5.02655im
 0.0-0.0im  0.0-2.51327im  0.0-5.02655im  0.0-7.53982im  0.0-10.0531im
 0.0-0.0im  0.0-3.76991im  0.0-7.53982im  0.0-11.3097im  0.0-15.0796im
 0.0-0.0im  0.0-5.02655im  0.0-10.0531im  0.0-15.0796im  0.0-20.1062im
# define DFT matrix by exponentiation of each element
julia> W=exp.(W)
5x5 Matrix{ComplexF64}:
 1.0-0.0im      1.0-0.0im      1.0-0.0im      1.0-0.0im      1.0-0.0im
 1.0-0.0im  0.309017-0.951057im -0.809017-0.587785im -0.809017+0.587785im  0.309017+0.9
 1.0-0.0im -0.809017-0.587785im  0.309017+0.951057im  0.309017-0.951057im -0.809017+0.5
 1.0-0.0im -0.809017+0.587785im  0.309017-0.951057im  0.309017+0.951057im -0.809017-0.5
 1.0-0.0im  0.309017+0.951057im -0.809017+0.587785im -0.809017-0.587785im  0.309017-0.9
# compute DFT of signal
julia> F_k = W*s
5-element Vector{ComplexF64}:
 -1.1102230246251565e-16 + 0.0im
           2.5 - 5.551115123125783e-17im
 5.551115123125783e-17 - 3.0531133177191805e-16im
 3.608224830031759e-16 - 3.608224830031759e-16im
 2.4999999999999996 + 9.992007221626409e-16im
# select individual DFT component
julia> F_k[2]
2.5 - 5.551115123125783e-17im

```

This can be simplified using the builtin complex exponential function that has been optimized for evaluation at π intervals.

```

# The DFT matrix
function dft_matrix(N)
    W = [0:N-1...]*[0:N-1...]'
    cispi.(2*W/N)
end

```

That was the DFT implementation, which is impractical for industrial settings.

The FFT algorithm: understand how to "break it down, then build it back up to construct the DFT"

Julia code for decomposing the signal into even and odd sequences (once! the big question is how to use recursion?)...BUT is this really necessary, can the computation be done on the signal "in place"?
 select odd indices f[1:2:end] select even indices f[2:2:end]

NOTE: the secret sauce in the FFTW package by Steven G Johnson to cater to the hardware internals.

18 The Discrete Time Fourier Transform (DTFT)

The DTFT transforms a discrete time signal to a continuous frequency representation (p.479 prob.7.27 [4])

19 Multi-taper Method

The multitaper method is a smoothing method

20 Properties of the DFT

Introduce a compact notation that emphasizes important properties. View the complex exponentials as values from a discrete signal, the discrete (vector) complex exponential

$$\omega = (1, e^{2\pi i/N}, e^{2\pi i 2/N}, \dots, e^{2\pi i(N-1)/N})$$

where the m^{th} component is

$$\omega[m] = e^{2\pi i m/N}$$

Define powers of ω as

$$\begin{aligned}\omega^{(n)} &= \omega^n = (1, e^{2\pi i n/N}, e^{2\pi i 2n/N}, \dots, e^{2\pi i(N-1)n/N}) \\ \omega^{(-n)} &= \omega^{-n} = (1, e^{-2\pi i n/N}, e^{-2\pi i 2n/N}, \dots, e^{-2\pi i(N-1)n/N})\end{aligned}$$

So the DFT is

$$\mathcal{F}f = \sum_{n=0}^{N-1} f[n] \omega^{(-n)}$$

which operates on the index, m

$$\mathcal{F}f[m] = \sum_{n=0}^{N-1} f[n] \omega^{(-n)}[m]$$

Properties of the DFT

1. Periodicity of the inputs and outputs (Lect. 21) - difference between continuous and discrete cases.
 - “The definition of the DFT compels you to regard the input f and the output F as not defined only on the interval $[0, N-1]$ but as periodic signals of period N ”.
 - Note: ω is naturally a periodic signal with period N .
2. Orthogonality of (powers of) the complex exponentials, $\omega^{(k)} \perp \omega^{(\ell)}$

$$\omega^{(k)} \cdot \omega^{(\ell)} = \begin{cases} 0 & , k \neq \ell \\ N & , k = \ell \end{cases}$$

- Most interesting properties of DFT are related to this.
- Proof:

The inner product of $\omega^{(k)}$ and $\omega^{(\ell)}$ is

$$\begin{aligned}\omega^{(k)} \cdot \omega^{(\ell)} &= \sum_{n=0}^{N-1} \omega^{(k)}[n] \overline{\omega^{(\ell)}[n]} \\ &= \sum_{n=0}^{N-1} e^{2\pi i k n/N} \overline{e^{2\pi i \ell n/N}} \\ &= \sum_{n=0}^{N-1} \left(e^{2\pi i (k-\ell)/N} \right)^n\end{aligned}$$

This is a geometric series, and if $k \neq \ell$

$$\begin{aligned} &= \frac{1 - \left(e^{2\pi i(k-\ell)/N} \right)^N}{1 - e^{2\pi i(k-\ell)/N}} \\ &= \frac{1 - e^{2\pi i(k-\ell)}}{1 - e^{2\pi i(k-\ell)/N}} = 0 \end{aligned}$$

If $k = \ell$

$$\omega^{(k)} \cdot \omega^{(\ell)} = N$$

3. The complex exponentials are orthogonal but not orthonormal

$$\|\omega^{(n)}\|^2 = \omega^{(n)} \cdot \omega^{(n)} = N$$

- Source of a N or $1/N$ factor in many formulas related to the DFT.
4. The inverse DFT (IDFT) has a factor of $1/N$, which arises because the vector complex exponentials are not orthonormal, which distinguishes it from its continuous analog

$$\begin{aligned} \mathcal{F}^{-1}f[m] &= \frac{1}{N} \sum_{n=1}^{N-1} f[n] e^{2\pi i n m / N} \\ &= \frac{1}{N} \sum_{n=1}^{N-1} f[n] \omega^{(n)}[m] \end{aligned}$$

$$\boxed{\mathcal{F}^{-1}f = \frac{1}{N} \sum_{n=1}^{N-1} f[n] \omega^{(n)}}$$

- Proof (show that, $\mathcal{F}^{-1}\mathcal{F}f[m] = f[m]$):

$$\begin{aligned} \mathcal{F}^{-1}\mathcal{F}f[m] &= \frac{1}{N} \sum_{n=0}^{N-1} \mathcal{F}f[n] e^{2\pi i n m / N} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \left(\sum_{k=0}^{N-1} f[k] e^{-2\pi i k n / N} \right) e^{2\pi i n m / N} \end{aligned}$$

swap order of summation

$$= \frac{1}{N} \sum_{k=0}^{N-1} f[k] \sum_{n=0}^{N-1} e^{-2\pi i k n / N} e^{2\pi i n m / N}$$

Recall

$$\sum_{n=0}^{N-1} e^{-2\pi i k n / N} e^{2\pi i n m / N} = \begin{cases} 0 & , k \neq m \\ N & , k = m \end{cases}$$

20.1 Special Cases

: pick up here for updating **boldsymbol** notation

There are cases where the DFTs is analagous to the continuous Fourier transform.

The Fourier transform of f at $m = 0$

$$F[0] = \sum_{n=0}^{N-1} f[n] \omega^{-1}[0] = \sum_{n=0}^{N-1} f[n]$$

which is analagous to the continuous case, where $s = 0$

$$(\mathcal{F}f)(0) = \int_{-\infty}^{\infty} f(t) dt$$

Two special discrete signals:

The constant value function

$$\mathbf{1} = (1, \dots, 1)$$

The discrete δ -functions are given by

$$\begin{aligned}\delta_0 &= (1, 0, 0, 0, \dots) \\ \delta_1 &= (0, 1, 0, 0, \dots) \\ \delta_2 &= (0, 0, 1, 0, \dots) \\ &\vdots\end{aligned}$$

The DFT of the delta-function at 0, δ_0 , is

$$\mathcal{F}\delta_0 = \sum_{n=0}^{N-1} \delta_0[n] \omega^{(-n)}$$

The δ_0 -function at 0 pulls out the 0th component of $\omega^{(-n)}$, which is always equal to 1 regardless of the power of ω . This happens for every element in the summation, so

$$\begin{aligned}&= (1) \omega^0 \\ &= (1)(1, 1, \dots, 1) \\ &= (1, 1, \dots, 1) = \mathbf{1}\end{aligned}$$

analagous to the continuous case.

The DFT of the delta-function at k , δ_k , is

$$\mathcal{F}\delta_k = \sum_{n=0}^{N-1} \delta_k[n] \omega^{-n}$$

Again, δ_k pulls out the k^{th} component of $\omega^{(-n)}$; performed for every element in the summation gives

$$\begin{aligned}&= (1) \omega^{(-k)} \\ &= \omega^{(-k)}\end{aligned}$$

analagous to the continuous case.

The DFT of the vector of complex exponentials (has a factor of N) is

$$\boxed{\mathcal{F}\omega^{(k)} = N\delta_k}$$

Proof:

$$\begin{aligned}
\mathcal{F}\omega^{(k)} &= \sum_{n=0}^{N-1} \omega^{(k)}[n] \omega^{(-n)} \\
&= \sum_{n=0}^{N-1} \omega^{(k)}[n] \omega^{(-n)}[m] \\
&= \sum_{n=0}^{N-1} e^{2\pi i n k / N} e^{-2\pi i m n / N}
\end{aligned}$$

this is the inner product of $\omega^{(k)}$ and $\omega^{(m)}$, which is

$$\begin{aligned}
&= \begin{cases} 0 & , k \neq m \\ N & , k = m \end{cases} \\
&= N\delta_k
\end{aligned}$$

The DFT can be viewed as an $N \times N$ matrix, which operates on the signal vector. Define the N^{th} root of 1 (a scalar) as

$$\omega = e^{2\pi i / N}$$

to simplify references to matrix elements. Using this notation, the DFT becomes

$$\mathcal{F}f[m] = \sum_{n=0}^{N-1} \omega^{-nm} f[n]$$

which suggests matrix multiplication

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega^{-1} & \dots & \omega^{-(N-1)} \\ 1 & \omega^{-2} & \dots & \omega^{-2(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{-(N-1)} & \dots & \omega^{-(N-1)^2} \end{pmatrix} \begin{pmatrix} f[0] \\ f[1] \\ \vdots \\ f[N-1] \end{pmatrix} = \begin{pmatrix} \mathcal{F}f[0] \\ \mathcal{F}f[1] \\ \vdots \\ \mathcal{F}f[N-1] \end{pmatrix}$$

A usefull mnemonic is for the DFT matrix is

$$(\mathcal{F})_{nm} = (\omega^{-nm})$$

Note the DFT is symmetric but not unitary, the adjoint \mathcal{F}^* is not equal to the conjugate transpose of \mathcal{F}

$$\mathcal{F}^* \mathcal{F} = \mathcal{F} \mathcal{F}^* = N\mathbb{I}_{NN} \neq \mathbb{I}$$

This demonstrates that the IDFT is

$$\mathcal{F}^{-1} = \frac{1}{N} \mathcal{F}^*$$

To compute $\mathcal{F}f$ requires N^2 operations, $\mathcal{O}(N^2)$. The Fast Fourier Transform (FFT) reduces this to $\mathcal{O}(N \log(N))$.

(More/Back to general properties of the DFT)

Periodicity

An important difference between the Fourier transform and the DFT is both the inputs and the outputs must be periodic, because the fundamental building block of the DFT, the complex exponential, is periodic.

$$\begin{aligned}\omega[m] &= e^{2\pi im/N} \\ \omega[m+N] &= e^{2\pi i(m+N)/N} \\ &= e^{2\pi im/N} e^{2\pi iN/N}\end{aligned}$$

DON'T FOLLOW THIS...

$$= \omega[m]$$

Likewise, powers of $\omega^{(k)}$ are also periodic of period N .

The DFT is a linear combination of periodic signals all with period N . The output of the DFT is also periodic with period N .

$$\sum_{n=0}^{N-1} f[n] \omega^{(-n)}$$

The output of the IDFT is also periodic with period N .

$$\frac{1}{N} \sum_{n=0}^{N-1} f[n] \omega^{(n)}$$

Therefore the *signal must be regarded as periodic* with period N . This is a feature of the DFT, which differs from the continuous Fourier transform. Furthermore, asserting periodicity where it may not be present is an issue in practice...(PICK UP). A consequence of periodicity is any set of consecutive N indices will suffice, ie:

$$\mathcal{F}f = \sum_{n=1}^N f[n] \omega^{(-n)}$$

in this case, since

$$f[N] = f[0]$$

Negative indices can be used provided the cyclic reference is clear, ie:

$$f[-1] = f[N-1]$$

$$f[-2] = f[N-2]$$

Another common convention for N even is

$$\mathcal{F}f = \sum_{n=\frac{-1}{N}+1}^{\frac{N}{2}} f[n] \omega^{(-n)}$$

for N odd is

$$\mathcal{F}f = \sum_{n=\frac{-1}{N}}^{\frac{N}{2}} f[n] \omega^{(-n)}$$

which puts the 0th component in the middle of the spectrum.

Duality results using the reverse signal

$$f^{-}[n] = f[-n]$$

Duality for the DFT

$$\mathcal{F}f^{-} = (\mathcal{F}f)^{-}$$

Duality for the DFT which differs from the continuous case

$$\mathcal{F}\mathcal{F}f = Nf^*$$

Do convolution of DFT...(not done in lectures!)

21 The Discrete Fourier Transform (DFT), Pt. II

There are two things you can control, the length / duration (space or time) to sample, (I want to just call this L but it's not standard), and N , the number of samples.

The reciprocity relations provide a conceptual reference when considering the DFT. It is useful to start from the temporal/spacial and frequency grids as a foundation for thinking about the DFT. These grids suggest just how intimately the DFT is connected with how much time or space can be captured as a single frame, and the number of times you can sample within that domain. (Would it be useful to use T , or L , as the period of time, or length of space, despite notation conventions, for clarity of concepts?)

There are multiple routes to arrive at the DFT but rather than focus on one let's consider multiple to see what insights different approaches reveal.

21.1 Preliminaries: The Number of (Sub-) Intervals v. (Sample) Indices

Subtle (important) issue regarding the number of subintervals used to subdivide the spatial/temporal interval and the number of data points sampled.

No. intervals	No. data points
N	$N + 1$
$N-1$	N

For N intervals

$$N \cdot \Delta x = L$$

the width of the frequency interval is

$$\Delta\omega = \frac{1}{L}$$

but there are $N + 1$ sample points (x_n), each with an associated frequency (ω_k). Each frequency considered in the DFT is obtained by

$$n \cdot \Delta\omega = \omega_n$$

so there are more data points than frequencies to match. ***The way to avoid this ambiguity in the meaning of N is to assume/assert that the first and last data points are equal, ie: that the signal is periodic. By this construction the first data point technically the last data point, the one associated with the lowest frequency.

21.2 Preliminaries: The Symmetric v. Asymmetric Interval

Briggs and Henson like the symmetric interval $x \in [-L/2, L/2]$; specifically, the case where the number of subintervals (N) is even, and the indices run from $n : [-N/2 + 1 : N/2] \in \mathbb{Z}$ because the range of

frequencies is $[-\frac{\Omega}{2+1}, \frac{\Omega}{2}]$ necessarily (not by construction). Now there is no ambiguity when specifying the associated $\omega_m = (m)(\Delta\omega)$ as to what is meant by a negative frequency since it is the index not the frequency that carries the negative sign.

While this convention alleviates the issue of “negative frequencies” in the spatial domain, this convention introduces the concept of “negative time” when translating the spatial domain into the temporal domain.

21.3 The Fast Fourier Transform (FFT)

There are a variety of FFT algorithms that are case specific. The basics of treatment by Tukey proceeds as follows. The DFT as an $N \times N$ matrix is

$$\mathcal{F} = (\omega^{-nm})$$

where

$$\omega = e^{2\pi i/N}$$

indexed from 0 to $N - 1$ which operates on a signal regarded as a column vector. This algorithm is of order $\mathcal{O}(N^2)$.

Side: the popular approach is to use the symmetry of the matrix and properties of complex exponential to factor the DFT matrix into simpler matrices that require fewer multiplications.

Proceeding simply based on the DFT (Notation is becoming ambiguous, note that $\omega^{(-N)}[m] = \omega^{-nm}$)

$$\mathcal{F}f[m] = \sum_{n=0}^{N-1} f[n] \omega^{-nm}$$

Write the complete sum over even and odd indices, to write the order N DFT as a combination of two DFTs of order $N/2$, and then iterate. To start requires N is even. To iterate requires N is a power of 2 (zero padding is a technique for dealing with when this assumption is not met). Introduce notation to keep track of powers of the complex exponential

$$\omega[p, q] = e^{2\pi i q/p}$$

such that

$$\begin{aligned} \omega[p, q_1 + q_2] &= e^{2\pi i(q_1 + q_2)/p} \\ &= \omega[p, q_1] \omega[p, q_2] \\ \omega[p, -nm] &= \omega^{-nm} = e^{2\pi i(q_1 + q_2)/p} \end{aligned}$$

For even powers

$$\begin{aligned} \omega[N/2, -nm] &= e^{-2\pi i nm/(N/2)} \\ &= e^{-2\pi i(2nm)/N} \\ &= \omega[N, -2nm] \end{aligned}$$

For odd powers

$$\begin{aligned} \omega[N, -(2n+1)m] &= \omega[N, -2nm - m] \\ &= \omega[N, -2nm] \omega[N, -m] \\ &= \omega[N/2, -nm] \omega[N, -m] \end{aligned}$$

Including these results into the DFT

$$\mathcal{F}f[m] = \sum_{n=0}^{N-1} f[n] \omega[N, -nm]$$

write the sum over even and odd indices

$$\begin{aligned} &= \sum_{n=0}^{(N/2)-1} f[2n] \omega[N, -2nm] + \sum_{n=0}^{(N/2)-1} f[2n+1] \omega[N, -(2n+1)m] \\ &= \sum_{n=0}^{(N/2)-1} f[2n] \omega[N/2, -nm] + \sum_{n=0}^{(N/2)-1} f[2n+1] \omega[N/2, -nm] \omega[N, -m] \\ &= \sum_{n=0}^{(N/2)-1} f[2n] \omega[N/2, -nm] + \omega[N, -m] \sum_{n=0}^{(N/2)-1} f[2n+1] \omega[N/2, -nm] \end{aligned}$$

which is “almost”

$$\mathcal{F}_N f = \mathcal{F}_{N/2} f_{\text{even}} + \omega[N, -m] \mathcal{F}_{N/2} f_{\text{odd}}$$

however, the right hand side consists of $N/2$ -tuples... This formula holds for indices 0 to $N/2 - 1$. To get the remaining indices $N/2$ to $N - 1$, use the indexing scheme

$$m + \frac{N}{2} \quad \text{where } m = (0, 1, \dots, \frac{N}{2} - 1)$$

$$\mathcal{F}_N f[m + N/2] = \sum_{n=0}^{N/2-1} f[2n] \omega[N/2, -n(m + N/2)] + \omega[N, -(m + N/2)] \sum_{n=0}^{N/2-1} f[2n+1] \omega[N/2, -(m + N/2)n]$$

(min 30) come back if you dare!...

(Lect 22: min 46) ...for indices $m = (0, \dots, N/2 - 1)$

$$\mathcal{F}_N f[m] = \mathcal{F}_{N/2} f_{\text{even}}[m] + \omega[N, m] \mathcal{F}_{N/2} f_{\text{odd}}[m]$$

...for indices $m = (0, \dots, N/2 - 1)$

$$\mathcal{F}_N f[m + N/2] = \mathcal{F}_{N/2} f_{\text{even}}[m] - \omega[N, -m] \mathcal{F}_{N/2} f_{\text{odd}}[m]$$

Now Iterate! How to see the reduction in computational complexity $\mathcal{O}(N \log(N))$?

22 (Introduction to) Linear Systems

Major Topics: impulse response, and the transfer function, complex exponentials as eigen functions of linear time invariant systems, *analysis of linear systems in terms of the Fourier transform,

Def. of linear system: a mapping/function from inputs to outputs that satisfies the principles of superposition

$$\begin{aligned} L(v_1 + v_2) &= Lv_1 + Lv_2 \\ L(\alpha v) &= \alpha Lv \end{aligned}$$

The extension to infinite sums (and to integrals)

$$L\left(\sum_{n=0}^{\infty} v_n\right) = \sum_{n=0}^{\infty} Lv_n$$

$$L\left(\sum_{n=0}^{\infty} \alpha v_n\right) = \sum_{n=0}^{\infty} L(\alpha_n v_n) = \sum_{n=0}^{\infty} \alpha_n L(v_n)$$

generally requires additional assumptions on the operator (ie: assume “continuity” of L).

The quintessential (only! (B. Osgood)) example of a linear system is direct proportionality

$$\begin{aligned} L(v_1 + v_2) &= \alpha(v_1 + v_2) \\ &= \alpha v_1 + \alpha v_2 \\ &= L v_1 + L v_2 \\ L(\alpha v) &= \alpha L v \end{aligned}$$

(all linear systems are in some way similar to this case). Note, α need not be a constant; multiplication is a linear system.

$$Lv(t) = \alpha(t)v(t)$$

eg: flipping on a switch for a duration a

$$Lv(t) = \Pi_a(t)v(t)$$

eg: sampling

$$Lv(t) = \text{III}_p(t)v(t)$$

A (slight but important) generalization of “direct proportionality” is to include summing up the results, ie: matrix multiplication

$$\begin{aligned} A_{m \times n} v_n &= w_m \\ (Ax)_i &= \sum_{j=1}^n a_{i,j} v_j = w_i \end{aligned}$$

Linearity of matrix multiplication

$$\begin{aligned} A(v + w) &= Av + Aw \\ A(\alpha v) &= \alpha Av \end{aligned}$$

Common problem

$$\dot{x} = Av$$

Initial conditions

$$x(0) = v$$

Solution

$$x(t) = e^{tA}v$$

special cases: linear systems with special properties derive from special properties of the matrix. Square matrices...

Symmetrix (self adjoint) system

$$A = A^T$$

Hermitian (the complex case)

$$A = A^*$$

where A^* is the conjugate transpose.

Unitary

$$A^*A = \mathbb{I}$$

Often look for eigen vectors and eigen values of A . A vector, $v \neq 0$, is an eigen vector if

$$Av = \lambda v$$

If have eigen vectors

$$v_1, \dots, v_n$$

with corresponding eigen values

$$\lambda_1, \dots, \lambda_n$$

, that form a basis for all the inputs, then can analyze A easily, any input v can be written as

$$\begin{aligned} v &= \sum_{i=1}^n \alpha_i v_i \\ Av &= \sum_{i=1}^n A(\alpha_i v_i) \\ Av &= \sum_{i=1}^n \alpha_i Av_i \\ &= \sum_{i=1}^n \alpha_i \lambda_i v_i \end{aligned}$$

When do linear systems have a basis of eigen vectors (can use special properties)? The spectral theorem in finite dimensions for matrices states if A is Hermitian then there is an orthonormal basis of eigen vectors.

(One more property) “It’s not just that (matrix) multiplication is a good and natural example of finite dimensional linear systems, it’s the only example”. Any finite dimensional linear system can be written as matrix multiplication.

Ex: Inputs are polynomials of degree $\leq n$

$$a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

And the linear system is given by

$$L = \frac{d}{dx}$$

Then find L (hint: can be written as an $(n+1) \times (n+1)$ matrix).

“There is an analagous statement for the infinite dimensional continuous case”... The linear system that generalizes matrix multiplication is integration against a kernel. The inputs are functions $v(x)$. The kernel is a function $k(x, y)$, which defines the linear system. Integration against a kernel is

$$Lv(x) = \int_{-\infty}^{\infty} k(x, y)v(y) dy$$

is linear because integration is linear. This is the infinite dimensional continuous analog of matrix multiplication. Think of v as an infinite column vector. Think of $k(x, y)$ as a matrix with x the row index and y the column index. And integration is the summation.

Special linear systems arise by extra assumptions on the kernel $k(x, y)$. The analog for the kernel of the transpose of a matrix is the symmetry (self adjoint) condition

$$k(x, y) = k(y, x)$$

(self adjoint linear system). Hermitian symmetry in the case of a complex kernel is

$$k(x, y) = \overline{k(y, x)}$$

Orthogonality...

Examples of linear systems given by integration against a kernel, ie: Ex: the Fourier transform,

$$\mathcal{F}f(s) = \int_{-\infty}^{\infty} e^{-2\pi i s t} f(t) dt$$

where

$$k(s, t) = e^{-2\pi i s t}$$

is symmetric...

Ex: convolution

$$Lv = h * v$$

$$Lv(x) = \int_{-\infty}^{\infty} h(x - y)v(y) dy$$

Note, for a linear system given by convolution the kernel does not depend on x and y separately but on their difference. This property leads to convolution being a linear shift/time invariant system, ie:

$$x \longrightarrow x - a$$

$$y \longrightarrow y - a$$

$$x - y \longrightarrow x - a - (y - a) = x - y$$

“It’s not just that integration against a kernel is a good example of continuous infinite dimensional linear systems, it’s the only example!”...requires some assumptions...That is, any linear system can be written as integration against a kernel.

linear systems continued in the discrete finite dimensional case any linear system is given by a matrix. in the continuous infinite dimensional case, where the inputs and outputs are functions, any linear system is give by integration against a kernel.

$$Lv(x) = \int_{-\infty}^{\infty} k(x, y)v(y) dy$$

Digression on cascading(engineering)/composing(math) linear system



If L and M are each linear then

$$MLv = w$$

so is their cascade.

When L and M are given by integration against a kernel

$$Lv(x) = \int_{-\infty}^{\infty} k(x, y)v(y) dy$$

Note, subscript x on M_x suggests how M operates on k

$$MLv(x) = \int_{-\infty}^{\infty} M_x k(x, y) v(y) dy$$

Proof:

$$\begin{aligned} \int_{-\infty}^{\infty} k(x, y) v(y) dy &\approx \sum_i k(x, y_i) v(y_i) \Delta y_i \\ M \left(\sum_i k(x, y_i) v(y_i) \Delta y_i \right) \\ &= \sum_i M \left(k(x, y_i) v(y_i) \Delta y_i \right) \end{aligned}$$

Note, for M operates on the x variable while y is fixed.

$$\begin{aligned} &= \sum_i M_x (k(x, y_i)) v(y_i) \Delta y_i \\ &\approx \int_{-\infty}^{\infty} M_x (k(x, y)) v(y) dy \\ MLv(x) &= \int_{-\infty}^{\infty} M(k(x, y)) v(y) dy \end{aligned}$$

A rigorous proof requires more work. This is further support for the idea that any linear system can be represented as integration against some kernel.

Continuous analog of expressing a vector as the sum of its components

$$v(x) = \int_{-\infty}^{\infty} \delta(x - y) v(y) dy$$

If L is a linear system then Lv is given by applying L to the integral

$$\begin{aligned} Lv(x) &= L \left(\int_{-\infty}^{\infty} \delta(x - y) v(y) dy \right) \\ &= \int_{-\infty}^{\infty} L_x (\delta(x - y)) v(y) dy \end{aligned}$$

Set

$$k(x, y) = L_x (\delta(x - y))$$

then

$$Lv(x) = \int_{-\infty}^{\infty} k(x, y) v(y) dy$$

δ is called an “impulse” when you feed an impulse into a system the system response, so $k(x, y)$ is called the “impulse response”. The system L responds to an impulsive input namely $\delta(x - y)$. Note: h is typically used instead of k .

The Schwartz Kernel Theorem (The deepest fact in the theory of distributions is used daily by every electrical engineers) L is a linear operator on distributions (assumptions on continuity and boundedness) then there is a unique kernel k which is another distribution such that

$$Lv = \langle k, v \rangle$$

Examples:

Ex 1: What is the impulse response of the Fourier transform viewed as a linear system? Multiple approaches. (PickUP: Lect24: 24min)

$$\mathcal{F}(\delta(x - y)) = e^{2\pi ixy}$$

In addition

$$\mathcal{F}f(x) = \int_{-\infty}^{\infty} e^{-2\pi ixy} f(y) dy$$

the Fourier transform as a linear system can be viewed as integration against a kernel (h , traditional notation)

$$h(x, y) = e^{-2\pi ixy}$$

Schwartz Theorem says this kernel (the “impulse response”) is unique. ...Important point: If you can express a linear operator as integration you have found the impulse response.

Ex 2: (Prove this to yourself) ...What about the finite dimensional discrete case? Any finite dimensional linear system is given by multiplication by a matrix. What is the impulse response? The impulse response is the system itself, ie: the matrix.

Ex 3: The switch

$$Lv = \Pi v$$

What is the impulse response? L operating on the x variable.

$$L(\delta(x - y)) = \Pi(x)\delta(x - y)$$

using the sampling property of the δ function

$$= \Pi(y)\delta(x - y)$$

Then the impulse response is

$$h(x, y) = \Pi(y)\delta(x - y)$$

Check this

$$\begin{aligned} & \int_{-\infty}^{\infty} h(x, y)v(y) dy \\ &= \int_{-\infty}^{\infty} \Pi(y)\delta(x - y)v(y) dy \\ &= \int_{-\infty}^{\infty} \delta(x - y)\Pi(y)v(y) dy \end{aligned}$$

$$= \Pi(x)v(x)$$

which is the system.

Special case where the system is given by convolution

$$Lv = h * v$$

where h is fixed.

$$Lv(x) = Lv(x) = \int_{-\infty}^{\infty} h(x-y)v(y) dy$$

Here, can appeal to the Schwartz kernel theorem: the system is given by integration against a kernel then the impulse response is the kernel. Conclusion:

$$L(\delta(x-y)) = h(x-y)$$

Can also check this directly from the uniqueness aspect of the definition of the Schwartz kernel Theorem.

Recall, there is a special property of convolution that makes it especially important for linear systems. There is a (simple) relationship between convolution and delay (shift). Write the delay operator as

$$\tau_z v(x) = v(x-a)$$

Is the delay operator a linear operator? What is its impulse response? (do this, Show:) Convolution of a delayed signal is the delay of the convolution of the signal, ie:

$$h * (\tau_a v) = \tau_a(h * v)$$

Reinterpreted in terms of linear systems, this says if the system is given by convolution

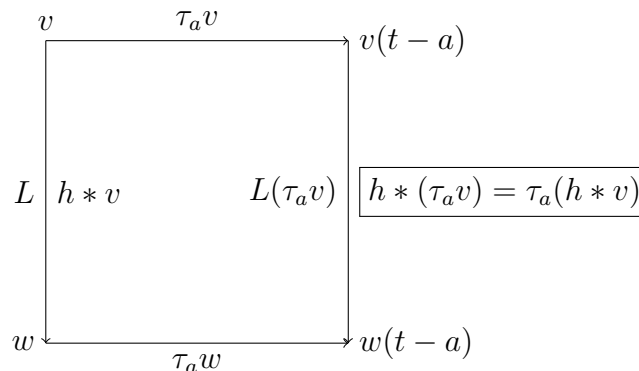
$$Lv = h * v$$

$$w = Lv = h * v$$

Delay of an input

$$v \longrightarrow v(x-a)$$

causes an identical delay of the output.



Convolution is a time/shift invariant system

$$w = Lv$$

$$w(x-a) = L(v(x-a))$$

$$L(\tau_a v) = \tau_a(Lv)$$

Or L commutes with shifts

$$L\tau_a = \tau_a L$$

The converse is also true! If the system is time invariant then it must be given by convolution. (Note: “the switch” is not time invariant.) ...But, why is this true? We know that for any linear system

$$Lv(x) = \int_{-\infty}^{\infty} L(\delta(x-y))v(y) dy$$

is given by the impulse response integrated against the function. What is $L(\delta(x-y))$? Let

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

then

$$L\delta(x-y) = L(\tau_y\delta)$$

But for a time invariant system the shifted input goes to the shifted output, so

$$\tau_y h(x) = h(x-y)$$

So

$$\begin{aligned} Lv(x) &= \int_{-\infty}^{\infty} h(x-y)v(y) dy \\ h(x) &= L(\delta(x)) \\ h(x-y) &= L(\delta(x-y)) \end{aligned}$$

Back to “the switch”

$$\begin{aligned} w &= \Pi v \\ Lv &= \Pi v \\ h(x, y) &= \Pi(y)\delta(x-y) \end{aligned}$$

which is not of the form $f(x-y)$ alone.

Recap: 1. any linear system is integration against a kernel, and the kernel is the impulse response. - special case: a time invariant system is also integration against a kernel but the integration is given by convolution, the “impulse response” is given by the difference of the variables not the variables independently. 2. a system is time invariant iff it is given by convolution. - “any time the convolution comes into the picture, the Fourier transform cannot be far behind” (BOsgood)

22.1 Linear Time Invariant (LTI) Systems and the Fourier Transform

General structure of a linear system L in terms of the impulse response

$$h(x, y) = L(\delta(x-y))$$

The Schwartz Kernel Theorem: the output of a linear system is given by integrating the input against the impulse response, ie: if

$$w(x) = Lv(x)$$

then

$$w(x) = \int_{-\infty}^{\infty} h(x, y)v(y) dy$$

In the special case of an LTI system: a delay in the input results in a corresponding delay of the output, ie: if

$$w(x) = Lv(x)$$

then

$$w(x - y) = L(v(x - y))$$

Furthermore, a linear system is time invariant *iff* if it is given by convolution. The impulse response

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

and by time invariance

$$h(x - y) = L(\delta(x - y))$$

and the action of the system is given by convolution

$$w(x) = \int_{-\infty}^{\infty} h(x - y)v(y) dy$$

The same considerations hold for linear finite discrete systems

$$Lv(x) = Av$$

given by matrix multiplication. L is LTI iff

$$w = h * v$$

where h is the impulse response, v is the input vector, w is the output vector.

$$h = L\delta$$

$$h[n - m] = L\delta_m$$

If

$$w = Av$$

then A has a special form for time invariant systems eg:

$$h = (1, 2, 3, 4) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$w = Av = h * v$$

What is A ? Columns of A are

$$A\delta_0 = h * \delta_0$$

$$A\delta_1 = h * \delta_1$$

$$A\delta_2 = h * \delta_2$$

$$A\delta_3 = h * \delta_3$$

$$(h * \delta_1)[m] = h[m - 1]$$

$$(h * \delta_1) = \begin{pmatrix} 4 \\ 1 \\ 2 \\ 3 \end{pmatrix} = (h * \delta_1)[0] = h[-1] = h[3] = 4$$

$$(h * \delta_2) = \begin{pmatrix} 3 \\ 4 \\ 1 \\ 2 \end{pmatrix}$$

$$(h * \delta_3) = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 1 \end{pmatrix}$$

So the matrix A is

$$A = \begin{pmatrix} 1 & 4 & 3 & 2 \\ 2 & 1 & 4 & 3 \\ 3 & 2 & 1 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

is a “circulant” (triplet) matrix. Notice, 1. the elements along the diagonal are constant, and 2. each column is obtained from the previous column by a shift.

The Fourier transform for LTI systems.

For a LTI system...

$$w = h * v$$

Take the Fourier transform and use the convolution theorem

$$\mathcal{F}w = (\mathcal{F}h)(\mathcal{F}v)$$

$$W(s) = H(s)V(s)$$

The “impulse response” is h ($h(x)$ or $h(x-y)$, depending on situation). The “transfer/system function” is $H(s)$. In the frequency domain the system is given by direct proportion (multiplication). In the time domain the system is ... convolution. For LTI systems the complex exponentials are eigen functions... (...) Input is

$$v(x) = e^{2\pi i \nu x}$$

What is $Lv(x)$ (sometimes called the “frequency response”)? What is the F.T.? Work in the frequency domain

$$\mathcal{F}(e^{2\pi i \nu x}) = \delta(x - \nu)$$

Then the output is

$$W(s) = H(s)\delta(s - \nu)$$

$$W(s) = H(\nu)\delta(s - \nu)$$

Go back to the time domain (take the inverse Fourier transform)

$$w(x) = H(\nu)e^{2\pi i \nu x}$$

ie:

$$L(e^{2\pi i \nu x}) = H(\nu)e^{2\pi i \nu x}$$

This says, $e^{2\pi i\nu x}$ is an eigen function(/vector) for *any* LTI system with eigen value $H(\nu)$. The complex exponentials form a basis of eigen functions...can diagonalize the operator associated with a LTI system. Note, however, in general, $\sin(x)$ and $\cos(x)$ themselves are not eigen functions of LTI systems, ie: Given

$$v(x) = \cos(2\pi\nu x)$$

What is $Lv(x)$?

$$\begin{aligned} L(\cos(2\pi\nu x)) &= L\left(\frac{e^{2\pi i\nu x} + e^{-2\pi i\nu x}}{2}\right) \\ &= \frac{1}{2}\left(L(e^{2\pi i\nu x}) + L(e^{-2\pi i\nu x})\right) \\ &= \frac{1}{2}\left(H(\nu)e^{2\pi i\nu x} + H(-\nu)e^{-2\pi i\nu x}\right) \end{aligned}$$

Without further assumptions this cannot be simplified. If h is real, then h has symmetry

$$H(-\nu) = \overline{H(\nu)}$$

$$\begin{aligned} &L(\cos(2\pi\nu x)) \\ &= \frac{1}{2}\left(H(\nu)e^{2\pi i\nu x} + \overline{H(\nu)}e^{2\pi i\nu x}\right) \end{aligned}$$

which is the real part

$$\text{Re}(H(\nu)e^{2\pi i\nu x})$$

still stuck, this is not an eigen function. Writing $H(\nu)$ in terms of the magnitude and the phase

$$H(\nu) = |H(\nu)|e^{i\phi}$$

then

$$\begin{aligned} H(\nu)e^{2\pi i\nu x} &= |H(\nu)|e^{2\pi i(\nu x + \phi)} \\ \text{Re}(H(\nu)e^{2\pi i\nu x}) &= |H(\nu)|\cos(2\pi\nu x + \phi) \end{aligned}$$

which is still not an eigen function but as close as you can get. With additional assumptions (which?) can these be made eigen functions?

The discrete version:

$$w = Lv = h * v$$

$$W[m] = H[m]V[m]$$

Discrete complex exponentials are eigen vectors. ie:

$$v = \omega^k$$

$$\mathcal{F}(\omega^k) = N\delta_k$$

What is $L(\omega^k)$? In the frequency domain

$$\begin{aligned} W &= HN\delta_k \\ &= H[k]N\delta_k \end{aligned}$$

In the time domain

$$\begin{aligned} w &= H[k]\omega^k \\ L(\omega^k) &= H[k]\omega^k \end{aligned}$$

Recall, powers of ω^k cycle. So

$$(\mathbf{1}, \omega, \omega^2, \dots, \omega^{N-1})$$

form a basis (not orthonormal) of eigen vectors for *any* LTI system.

Ex:

$$w = h * v$$

$$h = (1, 2, 3, 4)$$

Given by

$$w = Av$$

where

$$A = \begin{pmatrix} 1 & 4 & 3 & 2 \\ 2 & 1 & 4 & 3 \\ 3 & 2 & 1 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

So eigen vectors of the system are eigen vectors of A The eigen values are given by the values of the transfer function

$$H[0], H[1], H[2], H[3]$$

$$H = \mathcal{F}h = \sum_{k=0}^3 h[k]\omega^{(-k)}$$

$$= \sum_{k=0}^3 (k+1)\omega^{(-k)}$$

write this out...you get

$$\begin{pmatrix} 10 \\ -2 + 2i \\ -2 \\ -2 - 2i \end{pmatrix}$$

So the eigen values of A are $(10, -2 + 2i, -2, -2 - 2i)$.

23 Higher Dimensional Fourier Transform

ie: Fourier transforms of functions of several variables (x_1, x_2, \dots) . Applications: spectral analysis/signal processing for imaging (where, ie: $f(x_1, x_2)$ is intensity from black to white at point (x_1, x_2) is the is the spatial description of an image)

2D Fourier transform component parts are the 2D complex exponentials. think in terms of vectors.

Recall, the 1d Fourier transform

$$\mathcal{F}f(s) = \int_{-\infty}^{\infty} e^{-2\pi i s t} f(t) dt$$

where f is a function of t and $\mathcal{F}f$ is a function of s .

Then in 2d

$$x = (x_1, x_2)$$

spatial variables.

$$\xi = (\xi_1, \xi_2)$$

frequency variables. The function is

$$f(x) = f(x_1, x_2)$$

. The Fourier transform is

$$\mathcal{F}f(\xi) = \mathcal{F}f(\xi_1, \xi_2)$$

. The st in the complex exponential becomes $x\dot{\xi} = x_1\xi_1 + x_2\xi_2$, ie:

$$e^{-2\pi i(x\dot{\xi})} = e^{-2\pi i(x_1\xi_1 + x_2\xi_2)}$$

The Fourier transform is

$$\begin{aligned}\mathcal{F}f(\xi) &= \int_{\mathbb{R}^2} e^{-2\pi i(x\dot{\xi})} f(x) dx \\ \mathcal{F}f(\xi_1, \xi_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i(x_1\xi_1 + x_2\xi_2)} f(x_1, x_2) dx_1 dx_2\end{aligned}$$

In n -dimensions (skipping the mathematical motivation)

$$x = (x_1, \dots, x_n)$$

$$\xi = (\xi_1, \dots, \xi_n)$$

$$x \cdot \xi = x_1\xi_1 + \dots + x_n\xi_n$$

x has units of length, so the ξ_i have dimension $1/\text{length}$. The Fourier transform is

$$\mathcal{F}f(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i(x\cdot\xi)} f(x) dx$$

The inverse Fourier transform is

$$\mathcal{F}^{-1}g(\xi) = \int_{\mathbb{R}^n} e^{2\pi i(x\cdot\xi)} g(\xi) d\xi$$

Again there is a reciprocal relationship between the spatial domain and the frequency domain.

To consider

$$e^{\pm 2\pi i(x\cdot\xi)}$$

geometrically

first consider the 1d case

$$e^{2\pi i s t}$$

fixing the frequency s , then it is periodic with period $1/s$

$$\begin{aligned}e^{2\pi i s(t + \frac{1}{s})} \\ &= e^{2\pi i s t} e^{2\pi i s \frac{1}{s}} \\ &= e^{2\pi i s t}\end{aligned}$$

will be equal to 1 at the points $t = 0, \pm \frac{1}{s}, \pm \frac{2}{s}, \dots$ spaced $\frac{1}{s}$ apart.

now consider the 2d case fixing ξ_1 and ξ_2 when is the function

$$e^{-2\pi i(x_1\xi_1 + x_2\xi_2)}$$

equal to 1?

$$x_1\xi_1 + x_2\xi_2 = n$$

$$n = 0, \pm 1, \pm 2, \dots$$

Gives a family of evenly spaced parallel lines spaced $1/||\xi||$ apart. ie:

$$x_1\xi_1 + x_2\xi_2 = 0$$

Is (recall analytic vector geometry) a line through the origin with normal vector ξ .

$$x_1\xi_1 + x_2\xi_2 = 1$$

Is a line not through the origin, parallel to the previous, with normal vector ξ . How far apart are these lines?

$$\begin{aligned} ||x|| \cos \theta \\ x \cdot \xi = ||x|| ||\xi|| \cos \theta \\ ||x|| \cos \theta &= \frac{x \cdot \xi}{||\xi||} \\ &= \frac{1}{||\xi||} \end{aligned}$$

(Figure at min 45 Lect26) so

$$e^{-2\pi i(x_1\xi_1+x_2\xi_2)} = 1$$

along each of these lines, is periodic in the direction ξ with period $1/||\xi||$. High frequency (fast oscillation of the complex exponential (in a specific direction)) means the magnitude of $||\xi||$ is large, so $1/||\xi||$ is small. Low frequency (slow oscillation) means the magnitude of $||\xi||$ is small, so $1/||\xi||$ is large.

23.1 The Fourier Transform of Separable Functions

Special case(s): Class of examples for which a higher dimensional transform can be computed via lower (typically 1d) dimensional transforms, these are separable functions.

$$\begin{aligned} f(x_1, \dots, x_n) &= f(x_1) \cdots \cdots f(x_n) \\ \mathcal{F}f(\xi_1, \dots, \xi_n) &= \mathcal{F}f(\xi_1) \cdots \cdots \mathcal{F}f(\xi_n) \end{aligned}$$

Example: The 2d boxcar function $\Pi(x_1, x_2)$ (a box of height 1 coming out of the plane)

$$\Pi(x_1, x_2) = \begin{cases} 1 & , |x_1| < 1/2 \\ 1 & , |x_2| < 1/2 \\ 0 & \text{otherwise} \end{cases}$$

which can be written as the product of 1d boxcar functions

$$\begin{aligned} &\Pi(x_1)\Pi(x_2) \\ \mathcal{F}f(\xi_1, \xi_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i(x_1\xi_1+x_2\xi_2)} \Pi(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i x_1 \xi_1} e^{-2\pi i x_2 \xi_2} \Pi(x_1)\Pi(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} e^{-2\pi i x_1 \xi_1} \Pi(x_1) dx_1 \int_{-\infty}^{\infty} e^{-2\pi i x_2 \xi_2} \Pi(x_2) dx_2 \end{aligned}$$

$$\text{sinc}(\xi_1) \text{sinc}(\xi_2)$$

Example: 2d Gaussian

$$g(x_1, x_2) = e^{-\pi(x_1^2 + x_2^2)}$$

is separable

$$= e^{-\pi x_1^2} e^{-\pi x_2^2}$$

The Fourier transform is

$$\begin{aligned} \mathcal{F}g(\xi_1, \xi_2) &= \mathcal{F}g(\xi_1) \mathcal{F}g(\xi_2) \\ &= e^{-\pi(\xi_1^2 + \xi_2^2)} \end{aligned}$$

The central limit theorem also holds in higher dimensions.

23.2 The Fourier Transform of Radial Functions

Radial functions have symmetry in the radial or polar coordinates, which makes them more natural to work with in polar coordinates, ie: those that only depend on the distance r from the origin. The Fourier transform of a radial function is also a radial function.

The Gaussian is an example of a “radial” function, Introduce polar coordinates

$$r = \sqrt{x_1^2 + x_2^2}$$

$$\theta = \tan^{-1}\left(\frac{x_2}{x_1}\right)$$

$$x_1 = r \cos \theta$$

$$x_2 = r \sin \theta$$

$$e^{-\pi(x_1^2 + x_2^2)} = e^{-\pi r^2}$$

depends only on r not on x_1 and x_2 independently.

Convert the Fourier transform to a “radial” form. Notice, the Fourier transform and the DFT are cartesian by nature. (However there are many cases where the function is naturally described in polar form) Assume that $f(x)$ is a radial function.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i(x_1 \xi_1 + x_2 \xi_2)} f(x_1, x_2) dx_1 dx_2$$

Convert to polar coordinates

$$x_1 = r \cos \theta$$

$$x_2 = r \sin \theta$$

$$\xi_1 = \rho \cos \phi$$

$$\xi_2 = \rho \sin \phi$$

Will see: If f is a function of r then the Fourier transform is a function of ρ .

$$f(x_1, x_2) \longrightarrow f(r)$$

$$dx_1 dx_2 \longrightarrow r dr d\theta$$

How to perform inner product (what is meant by the F.T. being cartesian)?

$$\begin{aligned}
 & x_1 \xi_1 + x_2 \xi_2 \\
 &= r \cos \theta \rho \cos \phi + r \sin \theta \rho \sin \phi \\
 &= r \rho (\cos \theta \cos \phi + \sin \theta \sin \phi) \\
 &= r \rho (\cos(\theta - \phi)) \\
 \mathcal{F}f(x) &= \int_0^\infty \int_0^{2\pi} e^{-2\pi i r \rho \cos(\theta - \phi)} f(r) r \, dr \, d\theta
 \end{aligned}$$

Note:

$$\begin{aligned}
 & \int_0^{2\pi} e^{-2\pi i r \rho \cos(\theta - \phi)} \, d\theta \\
 &= \int_0^{2\pi} e^{-2\pi i r \rho \cos \theta} \, d\theta
 \end{aligned}$$

is independent of ϕ , given the integration is over a complete period. However, this is not an elementary integral, there is no formula for the antiderivative that enables you to evaluate the integral in closed form. ...So define

$$J_0(a) = \frac{1}{2\pi} \int_0^{2\pi} e^{-ia \cos \theta} \, d\theta \quad \text{where } a \text{ is real}$$

is the zeroth order Bessel function of the first kind. Bessel functions always come up in context of radial functions. So we have

$$2\pi J_0(2\pi r \rho)$$

and

$$F(\rho) = 2\pi \int_0^\infty f(r) J_0(2\pi r \rho) r \, dr$$

this is called the zeroth order Hankel transform.

23.3 convolution in n-dim case

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x - y) \, dy$$

In the 2d case

$$(f * g)(x_1, x_2) = \int_{-\infty}^\infty \int_{-\infty}^\infty f(y_1, y_2) g(y_1 - x_1) g(y_2 - x_2) \, dy_1 \, dy_2$$

all formulas, formulas, interpretations continue to apply. Importantly, the convolution theorem still holds

$$\begin{aligned}
 \mathcal{F}(f * g) &= (\mathcal{F}f)(\mathcal{F}g) \\
 \mathcal{F}\mathcal{F}(f * g) &= (\mathcal{F}f) * (\mathcal{F}g)
 \end{aligned}$$

(???) (B.Osgood uses this in the derivation of the sampling theorem, but I rejected this derivation...might have to go back to that...Prove this claim and check above derivation! Try to find derivation of convolution theorem for Fourier transforms)

23.4 Shift Theorem in \mathbb{R}^n

Recall, 1d version

$$\begin{aligned} f(t) &\longleftrightarrow F(s) \\ f(t-b) &\longleftrightarrow e^{-2\pi i s b} F(s) \end{aligned}$$

In 2d, shift each variable independently.

$$\begin{aligned} x_1 &\longrightarrow x_1 - b_1 \\ x_2 &\longrightarrow x_2 - b_2 \\ f(x_1, x_2) &\longrightarrow f(x_1 - b_1, x_2 - b_2) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i(x_1 \xi_1 + x_2 \xi_2)} f(x_1 - b_1, x_2 - b_2) dx_1 dx_2 \end{aligned}$$

(easy) change of variables

$$\begin{aligned} u_1 &= x_1 - b_1 \\ u_2 &= x_2 - b_2 \\ du_1 du_2 &= dx_1 dx_2 \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i((u_1+b_1)\xi_1 + (u_2+b_2)\xi_2)} f(u_1, u_2) du_1 du_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i(u_1+b_1)\xi_1} e^{-2\pi i(u_2+b_2)\xi_2} f(u_1, u_2) du_1 du_2 \\ &= e^{-2\pi i(b_1\xi_1 + b_2\xi_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i u_1 \xi_1} e^{-2\pi i u_2 \xi_2} f(u_1, u_2) du_1 du_2 \\ &= e^{-2\pi i(b_1\xi_1 + b_2\xi_2)} \mathcal{F}f(\xi_1, \xi_2) \\ b &= (b_1, b_2) \\ \xi &= (\xi_1, \xi_2) \\ f(x_1 - b_1, x_2 - b_2) &= f(x - b) \\ f(x) &\longleftrightarrow F(\xi) \\ f(x - b) &\longleftrightarrow e^{-2\pi i \xi \cdot b} F(\xi) \end{aligned}$$

23.5 Scaling Theorem in \mathbb{R}^n

1d case

$$f(t) \longleftrightarrow F(s)$$

reciprocal relationship

$$f(at) \longleftrightarrow \frac{1}{|a|} F\left(\frac{s}{a}\right)$$

2d case

$$f(x_1, x_2)$$

scale x_1 and x_2 independently, reciprocal relationship

$$f(a_1x_1, a_2x_2) \longleftrightarrow \frac{1}{|a_1|} \frac{1}{|a_2|} F\left(\frac{\xi_1}{a_1}, \frac{\xi_2}{a_2}\right)$$

extends naturally to higher dimension.

More general notion of scaling

$$x_1 \rightarrow ax_1 + bx_2$$

$$x_2 \rightarrow cx_1 + dx_2$$

ie: Matrix multiplication

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

So

$$f(x) \rightarrow f(Ax)$$

what happens to the Fourier transform? The Fourier transform of $f(Ax)$ is

$$\mathcal{F}f(Ax) = \int_{\mathbb{R}^n} e^{-2\pi i(x \cdot \xi)} f(Ax) dx$$

change variable

$$u = Ax$$

variables are not independent (are coupled) a linear change of variables the volume scales by the absolute value of the determinant

$$d\vec{u} = |\det(A)| d\vec{x}$$

What about the complex exponential? What happens to $x \cdot \xi$?

$$u = Ax$$

$$x \cdot \xi = (A^{-1}u) \cdot \xi$$

There is a general relationship between matrix multiplication and the inner product.

$$Bx \cdot y = x \cdot B^T y$$

(eg EE263: Reference to another course by B.Osgood) Write the inverse transpose as

$$(A^{-1})^T = A^{-T}$$

$$A^{-1}u \cdot \xi = u \cdot A^{-T}\xi$$

Plug into the Fourier transform

$$\begin{aligned} & \int_{\mathbb{R}^n} e^{-2\pi i(u \cdot A^{-T}\xi)} f(u) \frac{1}{|\det(A)|} du \\ &= \frac{1}{|\det(A)|} \int_{\mathbb{R}^n} e^{-2\pi i(u \cdot A^{-T}\xi)} f(u) du \\ &= \frac{1}{|\det(A)|} \mathcal{F}f(A^{-T}\xi) \end{aligned}$$

The stretch theorem

$$f(x) \rightarrow F(\xi)$$

$$f(ax) \longrightarrow \frac{1}{|\det(A)|} F(A^{-T}\xi)$$

Special cases

Scaling the coordinates

$$x_1 \longrightarrow a_1 x_1$$

$$x_2 \longrightarrow a_2 x_2$$

the matrix is given by

$$A = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$$

$$A^{-1} = \begin{pmatrix} 1/a_1 & 0 \\ 0 & 1/a_2 \end{pmatrix}$$

$$A^{-T} = \begin{pmatrix} 1/a_1 & 0 \\ 0 & 1/a_2 \end{pmatrix}$$

leads to earlier results.

coordinates under rigid rotation in 2d

$$A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

is orthogonal

$$A^T A = I$$

because

$$\begin{aligned} \det(A^T A) &= \det(I) = 1 \\ (\det(A^T))(\det(A)) & \\ (\det(A)^2) & \end{aligned}$$

So

$$|(\det(A))| = 1$$

A preserves volumes. Volumes are distorted under a linear transformation by the absolute value of the determinant.

$$A^T A = I$$

$$A^T = A^{-1}$$

$$(A^{-1})^T = A$$

$$A^{-T} = A$$

Plut into formula

$$\begin{aligned} f(Ax) &\longrightarrow \frac{1}{|\det(A)|} F(A^{-T}\xi) \\ &= F(A\xi) \\ f(Ax) &\longrightarrow F(A\xi) \end{aligned}$$

rotation of the spatial domain corresponds to the same rotation of the frequency domain. Physical interpretation: rotating an image \sim is the same as rotating its spectrum.

$$f(x) \longleftrightarrow F(\xi)$$

$$f(Ax) \longleftrightarrow \frac{1}{|\det(A)|} F(A^{-T}\xi)$$

In higher dimensions, reciprocal relationship is reinterpreted as inverse transpose, A^{-T} .

Higher order delta Functions

$$\langle \delta, \varphi \rangle = \varphi(0)$$

$$\varphi(x_1, \dots, x_n)$$

$$0 = (0, \dots, 0)$$

The Fourier transform is

$$\mathcal{F}\delta = 1$$

Shifted delta function

$$\delta(x - b) = \delta_b$$

$$\langle \delta_b, \varphi \rangle = \varphi(b)$$

$$b = (b_1, \dots, b_n)$$

The Fourier transform is

$$\mathcal{F}\delta_b = e^{-2\pi i b \cdot \xi}$$

$$f\delta = f(0)\delta$$

$$f\delta_b = f(b)\delta_b$$

Scaling properties of Delta The 1-d case

$$\delta(ax) = \frac{1}{|a|}\delta(x)$$

The n-d case

$$\delta(Ax) = \frac{1}{|\det(A)|}\delta(x)$$

A a matrix.

III's, lattices, and crystallography. Physical example of where “reciprocal” means inverse transpose. Transform coordinates by a nonsingular matrix A

$$\mathcal{F}(f(Ax)) = \frac{1}{|\det(A)|}\mathcal{F}f(A^{-T}\xi)$$

23.6 The Higher-Order III Function

Recall, for integer spacing in 1-d

$$\mathcal{F}\text{III} = \text{III}$$

And, for non-integer spacing

$$\text{III}_p(x) = \sum_{k=-\infty}^{\infty} \delta(x - kp)$$

the Fourier transform is

$$\mathcal{F}\text{III}_p = \frac{1}{p}\text{III}_{1/P}$$

important reciprocal relation.

The 2-d III function. For evenly spaced points in \mathbb{R}^2 consider a grid in the plane with points at integer coordinates, aka the integer lattice, \mathbb{Z}^2 .

$$\text{III}_{\mathbb{Z}^2}(x) = \sum_{k \in \mathbb{Z}^2} \delta(x - k)$$

where $k = (k_1, k_2)$ and k_1, k_2 are integers. The Fourier transform follows in the same way as the 1-d case

$$\mathcal{F}\text{III}_{\mathbb{Z}^2} = \text{III}_{\mathbb{Z}^2}$$

For non-integer spacing the 2-d III function is Spacing is modified by applying a (non-singular) 2×2 matrix to the integer lattice. Integer combinations of e_1, e_2 under linear transformation by A to go integer combinations of basis vectors of v_1, v_2 . The result of $A(\mathbb{Z}^2)$ is an oblique lattice

$$\mathfrak{L} = \{k_1 v_1 + k_2 v_2 | k_1, k_2 \in \mathbb{Z}\}$$

where

$$v_1 = A e_1, v_2 = A e_2$$

...speak about the area of the fundamental cell (parallelogram given by the basis vectors) in the lattice.

$$\text{Area}(\mathfrak{L}) = \text{area of parallelogram}$$

The III function on an oblique lattice is

$$\text{III}_{\mathfrak{L}}(x) = \sum_{p \in \mathfrak{L}} \delta(x - p)$$

The Fourier transform is

$$\mathcal{F}\text{III}_{\mathfrak{L}} =$$

$$\mathfrak{L} = A(\mathbb{Z}^2)$$

The reciprocal lattice

$$\mathfrak{L}^* = A^{-T}(\mathbb{Z}^2)$$

The area of \mathfrak{L}^* is the reciprocal of the area of the other lattice.

$$\text{Area}(\mathfrak{L}^*) = \frac{1}{\text{Area}(\mathfrak{L})}$$

The Fourier transform

$$\mathcal{F}\text{III}_{\mathfrak{L}} = \frac{1}{\text{Area}(\mathfrak{L})} \text{III}_{\mathfrak{L}^*}$$

involves the dual lattice

In higher dimensions...(USE THE GENERALIZED STRETCH THEOREM)

$$\mathcal{F}\text{III}_{\mathfrak{L}} = \frac{1}{|\det(?)|} \text{III}_{\mathfrak{L}^*}$$

involves the dual lattice

For a crystal Recall, it is the *electron density* for the atoms in a crystal that perform the measurement in an X-ray diffraction experiment. periodized version via convolve with III

$$\rho_{\mathfrak{L}}(x) = \rho(x) * \text{III}_{\mathfrak{L}}$$

$$= \sum_{p \in \mathfrak{L}} \rho(x - p)$$

The X-ray diffraction experiment produces the (magnitude) of the Fourier transform of ρ the lattice

$$\begin{aligned} \mathcal{F}(\rho_{\mathfrak{L}}) &= (\mathcal{F}\rho)(\mathcal{F}\mathbf{III}_{\mathbb{L}}) \\ &= (\mathcal{F}\rho)\left(\frac{1}{|\det(\mathfrak{L})|}\mathbf{III}_{\mathbb{L}^*}\right) \\ &= \sum_{p^* \in \mathfrak{L}^*} (\mathcal{F}\rho)(s) \delta(s - p^*) \\ &= \sum_{p^* \in \mathfrak{L}^*} (\mathcal{F}\rho)(p^*) \delta(s - p^*) \end{aligned}$$

23.7 Fundamental Problem of Tomography

Application to medical imaging (not MRI).

Density on 2-d surface/slice/whatever

$$\mu(x_1, x_2)$$

Apply X-ray through a region along a line and measure what know the intensity of X-ray upon entry I_0 and measure the intensity at exit I (derived in text)

$$I = I_0 e^{-\int_L \mu}$$

where

$$\int_L \mu$$

is the line integral along the line L . You don't know μ or $\int_L \mu$... Given many measurements one obtains values for all possible lines L . To recover μ consider the measurement values as given by a transform of μ evaluated along the the line L . **The Radon transform** of μ is

$$(\mathcal{R})(\mu)(L) = \int_L \mu$$

get μ by taking the inverse Radon transform. Answer in terms of the 2-d Fourier transform.

23.8 Tomography and inverting the Radon transform

Can you find μ given all values of its transform $\mathcal{R}_\mu(L)$? aka can we invert \mathcal{R}

Introduce coordinates that describe a family of lines in the plane. eg: given the equation of a line

$$y = mx + b$$

(m, b) describes the line So the set of coordinates (m, b)

$$-\infty < b < \infty$$

$$-\infty < m < \infty$$

is no good since it omits the vertical lines. A better set of coordinates is one that describes families of parallel lines, ie: all with the same normal vector

$$\hat{n} = (\cos \phi, \sin \phi)$$

$$0 \leq \phi < \pi$$

where ϕ is the angle wrt the abscissa. So a line is determined by the angle ϕ and the (minimum) signed distance from the origin ρ .

$$-\infty < \rho < \infty$$

So a set of parallel lines is given by a fixed ϕ while ρ may vary (these are not polar coordinates in the plane). The cartesian equation of a particular line in these coordinates is in vector form

$$x \cdot \hat{n} = \rho$$

where

$$n = (\cos \phi, \sin \phi)$$

In coordinate form

$$\rho = x_1 \cos \phi + x_2 \sin \phi$$

Use this to evaluate a line integral

$$\int_L \mu$$

via a δ -function. Fact: consider a δ function along a line parametrized as described as

$$\delta(\rho - x_1 \cos \phi - x_2 \sin \phi) = \begin{cases} 0 & , \text{ off the line} \\ \infty & , \text{ on the line} \end{cases}$$

a line-impulse, a δ function along the line.

$$\int_L \mu = \int_{\mathbb{R}^2} \mu(x_1, x_2) \delta(\rho - x_1 \cos \phi - x_2 \sin \phi) dx_1 dx_2$$

needs some proof but analagous to the 1-d case.

Use this to invert the Radon transform $\mathcal{R}\mu$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x_1, x_2) \delta(\rho - x_1 \cos \phi - x_2 \sin \phi) dx_1 dx_2$$

Consider passing a family of parallel lines to the integral via ρ with ϕ fixed. Then $\mathcal{R}\mu(\rho, \phi)$ becomes a function of ρ with ϕ fixed. ...Take the 1-d Fourier transform wrt ρ

$$\begin{aligned} \mathcal{F}_\rho \mathcal{R}\mu(\rho, \phi) &= \mathcal{F}_\rho(\mathcal{R}\mu(\rho, \phi)) \\ &= \int_{-\infty}^{\infty} e^{-2\pi i r \rho} \mathcal{R}\mu(\rho, \phi) d\rho \\ &= \int_{-\infty}^{\infty} e^{-2\pi i r \rho} \mathcal{R}\mu(\rho, \phi) d\rho \end{aligned}$$

$$= \int_{-\infty}^{\infty} e^{-2\pi i r \rho} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x_1, x_2) \delta(\rho - x_1 \cos \phi - x_2 \sin \phi) dx_1 dx_2 \right) d\rho$$

Change order of integration

$$\begin{aligned} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x_1, x_2) \left(\int_{-\infty}^{\infty} e^{-2\pi i r \rho} \delta(\rho - x_1 \cos \phi - x_2 \sin \phi) d\rho \right) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x_1, x_2) \left(\int_{-\infty}^{\infty} e^{-2\pi i r (x_1 \cos \phi + x_2 \sin \phi)} d\rho \right) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x_1, x_2) \left(\int_{-\infty}^{\infty} e^{-2\pi i (x_1 r \cos \phi + x_2 r \sin \phi)} d\rho \right) dx_1 dx_2 \end{aligned}$$

Introduce dual variables to x_1 and x_2 , $\xi_1 = r \cos \phi$ and $\xi_2 = r \sin \phi$.

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x_1, x_2) \left(\int_{-\infty}^{\infty} e^{-2\pi i (x_1 \xi_1 + x_2 \xi_2)} d\rho \right) dx_1 dx_2$$

In sum

$$\begin{aligned} \mathcal{F}_\rho(\mathcal{R}\mu(\rho, \phi)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x_1, x_2) e^{-2\pi i (x_1 \xi_1 + x_2 \xi_2)} dx_1 dx_2 \\ &= \mathcal{F}\mu(\xi_1, \xi_2) \end{aligned}$$

where this last Fourier transform is a genuine 2-d Fourier transform (recall, ϕ is being held fixed in the initial statement of the Fourier transform wrt ρ).

Recal: Fix ϕ . Let ρ vary. Take 1-d Fourier transform of the Radon transform wrt ρ . This gives the 1-d Fourier transform of μ . Think: measurements return values of

$$\mathcal{F}_\rho \mathcal{R}\mu(\rho, \phi)$$

Making the change of variables

$$\xi_1 = r \cos \phi$$

$$\xi_2 = r \sin \phi$$

results in a function of

$$G(\xi_1, \xi_2) = \mathcal{F}\mu$$

To get μ take the inverse Fourier transform

$$\mu = \mathcal{F}^{-1}G$$

Practical implementation is done via a discrete version. Come back to issues and trade offs of cartesian v. polar representations.

***FIX: turn these into subsections...

24 Distributions Continued

Excerpts from Robert Strichartz, “A Guide to Distribution Theory and Fourier Transforms”.

The thermometer analogy to distributions: Rather than measuring the temperature at a point in space (and time), it is more meaningful to consider

$$\int f(x)\varphi(x) dx$$

where $f(x)$ represents temperature, and $\varphi(x)$ corresponds to the device that performs the measurement, which is “concentrated” within the local vicinity and tends to zero at sufficient distance from the measuring device.

The measurement can be thought of as an average in the sense that $\varphi(x) \geq 0$ and the integral over all spaces is

$$\int \varphi(x) dx = 1$$

Note: this is not a restriction to nonnegative values. The difference of two thermometers with arbitrary coefficients

$$\int f(x)(a_1\varphi_1(x) - a_2\varphi_2(x)) dx$$

can take on any value.

def. *generalized functions*, denoted by

$$\langle f, \varphi \rangle \in \mathbb{R} \text{ or } \mathbb{C}$$

can be real or complex. Require “something akin to $\int f(x)\varphi(x) dx$ exists” for a suitable “test function”.
Linearity

$$a_1\langle f, \varphi_1 \rangle - a_2\langle f, \varphi_2 \rangle = \langle f, a_1\varphi_1 - a_2\varphi_2 \rangle$$

Note if φ_1, φ_2 are test functions then so is $a_1\varphi_1 + a_2\varphi_2$. Remaining questions: what constitutes a test function, and what is the continuity hypothesis?

Denote the space of test functions leading to the theory of distributions as, \mathcal{D} . The “underlying point set” is \mathbb{R}^n (or an open subset, $\Omega \subset \mathbb{R}^n$) with elements, $x = (x_1, \dots, x_n)$, ie: “every point $x \in \Omega$ is surrounded by a ball $y : |x - y| < \epsilon$ contained in Ω , where ϵ depends on x , and”

$$|x - y| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$$

“The class of test functions \mathcal{D} consists of all functions $\varphi(x)$ defined in Ω , vanishing outside a bounded subset of Ω that stays away from the boundary of Ω , and such that all partial derivatives of all orders of φ are continuous.” [20] For $n = 1$, $\Omega = 0 < x < 1$

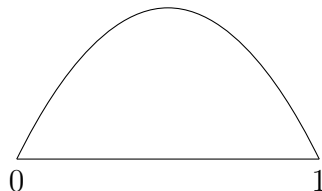


Figure 16: Is not in \mathcal{D} because it does not vanish near the boundaries.

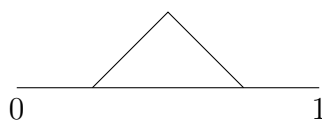


Figure 17: Is not in \mathcal{D} because it is not differentiable at three points.

No analytic function except $\varphi \equiv 0$ can be in \mathcal{D} due to the vanishing requirement. Any formula for φ must be written piecewise, ie: in \mathbb{R}

$$\psi(x) = \begin{cases} e^{-1/x^2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

has continuous derivatives of any order

$$\left(\frac{d}{dx}\right)^k e^{-1/x^2} = \frac{\text{polynomial in } x}{\text{polynomial in } x} e^{-1/x^2}$$

and $\psi(x) \rightarrow 0$ as $x \rightarrow 0$ because e^{-1/x^2} goes to 0 faster than the pole $\frac{1}{\text{polynomial in } x}$. So, any order derivative of $\varphi(x) = \psi(x)\psi(1-x)$ is continuous, $\varphi(x)$ is C^∞ . $\varphi(x)$ vanishes outside $0 < x < 1$, so $\varphi \in \mathcal{D}(a < x < b)$ so long as $a < 0$ and $b > 1$ (why not $a \leq 0$ and $b \geq 1$? DO THIS?) Given one example, others can be generated using normal function modifications

- Horizontal shift $\varphi(x + x_0)$.
- Vertical scaling $a\varphi(x)$.
- Horizontal scaling $\varphi(ax)$.
- Linear combination $a_1\varphi_1(x) + a_2\varphi_2(x)$.
- Products to generate higher dimensional versions $\varphi(x_1)\varphi(x_2)\dots\varphi(x_n) = \varphi_1(x_1, \dots, x_n)$

Def. the class of distributions on Ω is all continuous linear functionals on $\mathcal{D}(\Omega)$, denoted by $\mathcal{D}'(\Omega)$.
Notation for the *functional*

$$\langle f, \varphi \rangle$$

Linear functional

$$a_1\langle f, \varphi_1 \rangle + a_2\langle f, \varphi_2 \rangle = \langle f, a_1\varphi_1 + a_2\varphi_2 \rangle$$

Temporary definition of *continuous*: if φ_1 is close enough to φ then $\langle f, \varphi_1 \rangle$ is close to $\langle f, \varphi \rangle$. “linear functionals tend to be continuous...a certain amount of continuity is built into linearity...all linear functionals on $\mathcal{D}(\Omega)$ will be continuous” (yay!) Continuity argument (not proof):

$$\varphi_1 - \varphi = \varphi_2$$

$$\varphi_1 = \varphi + \varphi_2$$

Move φ_1 closer to φ via $\varphi + t\varphi_2$ By linearity

$$\langle f, \varphi + t\varphi_2 \rangle = \langle f, \varphi \rangle + t\langle f, \varphi_2 \rangle$$

, which as t gets small

$$= \langle f, \varphi \rangle$$

AT: 1.2 Examples of distributions

Any function can specify a distribution

$$\langle f, \varphi \rangle = \int_{\Omega} f(x)\varphi(x) dx$$

provided the integral exists.

25 Begin Wavelets

Strichartz, Robert S "A Guide to Distribution Theory and Fourier Transforms" (2003)

The Gabor transform, using the Gaussian as a prototype, can generate all functions frequency and space via "translations" in frequency and space. Wavelets are a different approach to expand a function in terms of simple component functions that are localized in frequency and space. Translations and "dilations" of

The simplest wavelet is the Haar function

$$\psi(x) = \begin{cases} 1 & \text{if } 0 < x \leq \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} < x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The support of ψ is $[0, 1]$. Dilate ψ by powers of 2. $\psi(2^j x)$ is supported on $[0, 2^{-j}]$ where j is an integer. And translate by 2^{-j} times an integer

$$\psi(2^j(x - 2^{-j}k)) = \psi(2^j x - k)$$

is supported on $[2^{-j}k, 2^{-j}(k+1)]$. With the normalization factor is $2^{j/2}$ the family of functions $2^{j/2}\psi(2^j x - k) = \psi_{j,k}(x)$ is orthonormal

$$\int \psi_{j,k}(x)\psi_{j',k'}(x) dx = \begin{cases} 1 & \text{if } j = j' \text{ and } k = k' \\ 0 & \text{otherwise} \end{cases}$$

If $j = j'$ but $k \neq k'$ then $\psi_{j,k}$ and $\psi_{j',k'}$ have disjoint support, so $(\psi_{j,k})(\psi_{j',k'}) = 0$. If $j \neq j'$ but $k = k'$ then $\psi_{j,k}$ and $\psi_{j',k'}$ may have overlapping support, but the integral of $(\psi_{j,k})(\psi_{j',k'})$ is still 0 via cancelation.

An arbitrary function can be expanded via the *Haar series expansion*

$$f = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k}$$

Define the system as *complete* if the expansion is valid. The usual caveats from Fourier series apply

- The acceptable functions are restricted to those such that

$$\int |f(x)|^2 dx < \infty$$

- The infinite series may not converge in the pointwise sense.

The Haar functions are complete...(p.142)

26 NOTES:

(MOVE THIS...) A lack of smoothness in the function, or in any of its derivatives, being approximated by a Fourier series forces an infinite number of terms. A finite number of terms will be smooth, as it is infinitely differentiable.

Greater generality requires a different point of view. The important condition is integrability of the function. If a function is periodic, and square integrable, $f \in L^2([0, 1])$, if

$$\int_0^1 |f(t)|^2 dt < \infty$$

(f has finite energy).

$$\int_0^1 \left| f(x) - \sum_{k=-n}^n \hat{f}(k) \exp(2\pi i k t) \right|^2 dt \rightarrow 0 \text{ as } n \rightarrow \infty$$

Can only get such convergence results with a generalization of the integral due to Lebesgue (where the L in L^2 comes from, above).

The transition from Fourier series to the Fourier transform involves the transition from periodic phenomena to non-periodic phenomena. View non-periodic function as a limiting case of a periodic function as the period tends to infinity. Recall there are two aspects: analysis and synthesis Analysis: for periodic $f(t)$ then the Fourier coefficients are

$$\hat{f}(K) = \int_0^1 dt \exp(-2\pi i k t) f(t)$$

Synthesis: recover $f(t)$ from its constituent components

$$f(t) = \sum_{k=-\infty}^{\infty} dt \hat{f}(k) \exp(2\pi i k t)$$

Fourier transform is the generalization (limiting case, as the period tends to infinity) of the Fourier coefficients. Inverse Fourier transform is the limiting case of Fourier series. Need a setup when $f(t)$ is periodic of period T and let T tend to infinity. The building blocks for $f(t)$ with period T are

$$\exp(2\pi i k t / T)$$

such that the Fourier series take the form

$$f(t) = \sum_{k=-\infty}^{\infty} C_k \exp(2\pi i k t / T)$$

and the coefficients are got by

$$\begin{aligned} C_k &= \frac{1}{T} \sum_0^T \exp(-2\pi i k t / T) f(t) \\ &= \frac{1}{T} \sum_{-T/2}^{T/2} \exp(-2\pi i k t / T) f(t) \end{aligned}$$

and the spacing between Fourier coefficients (the frequency spectrum) is $1/T$ Note: the reciprocal relationship between the period (time domain) and the frequencies. As the period T tends to infinity the spectrum becomes continuous. Recall we are attempting to allow the period to tend to infinity and transition from periodic to non-periodic phenomena. BUT, assume the function goes to zero within the specified boundary T , then the Fourier coefficients go to zero due to the factor of $1/T!$ ie:

$$C_k \rightarrow 0$$

How to save this? Scale up by T .

$$\mathfrak{f}f\left(\frac{k}{T}\right) = \int_{-T/2}^{T/2} dt \exp(-2\pi i(k/T)t) f(t)$$

Fourier series looks like

$$f(t) = \sum_{k=-\infty}^{\infty} \mathfrak{f}f\left(\frac{k}{T}\right) \exp(2\pi i(k/T)t) \frac{1}{T}$$

References

- [1] Griffiths, D.J. *Introduction to Quantum Mechanics*, Ed. 2. Prentice Hall, New Jersey, 2005.
- [2] Bekefi, G. & Barrett A.H. *Electromagnetic Vibrations, Waves, and Radiation*. MIT Press, Massachusetts 1977.
- [3] Folland, G.B. *Fourier Analysis and its Applications*. Wadsworth & Brooks/ Cole Advanced Books in Software, California, 1992.
- [4] Osgood, B. *Lecture on the Fourier Transform and its Applications* AMS, 2019.
- [5] Briggs, W.L, & Henson V.E. *The DFT, An Owner's Manual* siam, 1995.
- [6] Henson, V.E. *Fourier Methods of Image Reconstruction* PhD thesis, University of Colorado at Denver, 1990.
- [7] Henson, V.E. *DFTs on irregular grids: The antepolated DFT* Technical Report NPS-MA-92-006, Naval Postgraduate School, Monterey, CA, 1992.
- [8] Henson, V.E. *An extended Poisson sum formula for the generalized integral transform and aliasing error bound for the sampling theorem* J. Appl. Anal, 26 (1988) 1988.
- [9] Bracewell, R.N. *The Fourier Transform and its Applications, Ed. 3*. McGraw-Hill, 2000.
- [10] Bracewell, R.N. *Fourier Analysis and Imaging*. Springer, 2003.
- [11] Hamming, R.W. *Digital Filters, Ed. 3*. Dover, 1997.
- [12] Oppenheim, A.V, & Schafer R.E. *Discrete-Time Signal Processing, Ed. 3*. Pearson, 2010.
- [13] Percival, D.B, & Walden A.T. *Spectral Analysis for Physical Applications (Multitaper And Conventional Univariate Techniques)*. Cambridge University Press, 1993.
- [14] Percival, D.B, & Walden A.T. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 2000.
- [15] Percival, D.B, & Walden A.T. *Spectral Analysis for Univariate Time Series*. Cambridge University Press, 2020.
- [16] Karu, Z.Z *Signals and Systems Made Ridiculously Simple*. ZiZi Press, 1995.
- [17] Stein, E.M, & Shakarchi R. *Fourier Analysis An Introduction*. Princeton University Press, 2011.
- [18] Jerri, A.J. *Integral and Discrete Transforms with Applications and Error Analysis*. CRC Press, 1992.

- [19] Jerri, A.J. *Advances in the Gibbs Phenomenon*. Sampling Publishing, 2011.
- [20] Strichartz, R.S. *A Guide to Distribution Theory and Fourier Transforms*. CRC Press, 1994.
- [21] Papoulis, A. & Pillai, S.U. *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, 2002.
- [22] Lighthill, M.J. *An Introduction to Fourier Analysis and Generalised Functions*. Cambridge Monographs on Mechanics, 1958.
- [23] Van Loan, C. *Computational Frameworks for the Fast Fourier Transform*. SIAM, 1992.
- [24] Weaver, H.J. *Applications of Discrete and Continuous Fourier Analysis*. Wiley-Interscience, New York, 1983.
- [25] Brigham, E. *The Fast Fourier Transform and it's Applications*. Prentice Hall, 1988.
- [26] Chu, E. *Discrete and Continuous Fourier Transforms: Analysis, Applications and Fast Algorithms*. CRC Press, 1988.
- [27] Mallet, S.G. *A wavelet tour of signal processing: the sparse way*.
- [28] Wikipedia. *Nyquist–Shannon sampling theorem*, <https://en.wikipedia.org/wiki/Nyquist> accessed 4/15/2022.
- [29] Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery B.P. *Numerical Recipes: The Art of Scientific Computing, 3rd Ed*. Cambridge University Press, 2007.