Chapter 2

Fundamentals

The Fourier transform (FT)

- Consider a mono-frequency sinusoidal function, g(t), given by:

$$g(t) = |A_0| \cos(2\pi f_0 t - \phi_0). \quad (2.1)$$

- g(t) can be completely described in terms of the following parameters:
  - The peak amplitude $|A_0|$, which is the highest amplitude of g(t).
  - The frequency f_0, which is the number of cycles per second.
  - The phase ϕ_0, which is a time-shift applied to g(t) with respect to t = 0.
- A time-dependent periodic signal, h(t) with a fundamental (dominant) period T_0, can be analyzed to an infinite series of sinusoids, each having its own peak-amplitude, frequency, and phase; using the Fourier series:

$$h(t) = \sum_{n=0}^{\infty} |A_n| \cos(2\pi f_n t - \phi_n); \quad (2.2)$$

  where A_n, f_n, and ϕ_n are the peak amplitude, frequency, and phase of the n^{th} sinusoidal component and $f_0 = 1/T_0$ is the fundamental (dominant) frequency of h(t). (Examples 1 and 2).
- We use the forward Fourier transform to compute the peak amplitude A_n and phase ϕ_n at every frequency f_n.
- The forward Fourier transform H(f) of the time-function h(t) is given by:

$$H(f) = \int_{-\infty}^{+\infty} h(t) e^{-i2\pi ft} dt. \quad (2.3)$$

- H(f) is generally a complex function that can be represented as:

$$\left.\begin{aligned} H(f) &= |H(f)|e^{i\phi_h(f)} \\ &= |H(f)|\{\cos[\phi_h(f)] + i\sin[\phi_h(f)]\} \\ &= H_r(f) + iH_i(f) \\ |H(f)| &= \sqrt{H_r^2(f) + H_i^2(f)} \\ \phi_h(f) &= \tan^{-1}[H_i(f)/H_r(f)] \end{aligned}\right\}, \quad (2.4)$$

  where $H_r(f)$, $H_i(f)$, $|H(f)|$, and $\phi_h(f)$ are the real part, imaginary part, amplitude spectrum, and phase spectrum of H(f), respectively.

- Practically, when we carry out the forward Fourier transform, it returns the amplitude spectrum $|H(f)|$ and phase spectrum $\phi_h(f)$ as functions of frequency f.

- Conversely, given $|H(f)|$ and $\phi_h(f)$, we can synthesize h(t) using the inverse Fourier transform:

$$h(t) = \int_{-\infty}^{+\infty} H(f)e^{i2\pi ft}df, \quad (2.5)$$

  where H(f) is calculated from $|H(f)|$ and $\phi_h(f)$ using equation (2.4).

- The algorithm used commonly to calculate the Fourier transform numerically is the fast Fourier transform (FFT).
  - The FFT works fast only if the length (i.e., number of samples) of h(t) is a power of 2 (i.e., 2, 4, 8, …).
  - Zero right-padding is usually used for other lengths make them powers of 2. For example, a 100-sample function can be extended by adding 28 zeros after its last sample to make its length = 128 = 2^7 samples.

- The Fourier transform will transform any function p(a) from its domain (a) to a function P(b) in another domain (b) provided that: b = 1/a (e.g., time to frequency).
- In this course, we will denote a Fourier transform pair as: $h(t) \Leftrightarrow |H(f)|$ and we will only calculate the amplitude spectrum of the function. The phase spectrum will only be calculated if needed.
- More Internet resources are on: https://en.wikipedia.org/wiki/Fourier_transform.

The impulse (delta) function

- The impulse function δ(t) is defined practically as:

$$\delta(t) = \begin{Bmatrix} \infty & t = 0 \\ 0 & t \neq 0 \end{Bmatrix}. \qquad (2.6a)$$

  - δ(t) can be thought of as a function that has a very large magnitude at t = 0 and infinitesimally small duration such that the area under the curve is one.
  - For sampling purposes, δ(t) can be treated as a function that has a magnitude of one at t = 0 and zero elsewhere (Figure, Wikipedia, 2013).
  - Generally, for constant k and t_0, the function $k\,\delta(t-t_0)$ has a magnitude of k at $t = t_0$ and zero elsewhere.
  - The most desirable property of δ(t) is its FT: $\delta(t) \Leftrightarrow 1$, $-\infty < f < \infty$. That is, the FT of δ(t) is constant for all frequencies (Figure, www.physforum.com, 2013).
  - Another function that is related to δ(t) is the sampling (Dirac) function D(t) given as:

$$D(t) = \sum_{n=0}^{N} \delta(t - n\Delta t) = \begin{Bmatrix} 1 & t = 0, \Delta t, 2\Delta t, \ldots, N\Delta t \\ 0 & t \neq 0, \Delta t, 2\Delta t, \ldots, N\Delta t \end{Bmatrix}, \qquad (2.6b)$$

where Δt is the sampling interval and N is the number of samples (Figure, Wikipedia, 2013).

- When *D(t)* is multiplied by a continuous function, it extracts the function values at intervals of Δt and discards all other values. This is used for sampling or analog-to-digital (A/D) conversion.
- More Internet resources are on: https://en.wikipedia.org/wiki/Dirac_delta_function.

The sinc function

- The sinc function is defined as:

$$\text{sinc}(t) = \begin{Bmatrix} 1 & t = 0 \\ \dfrac{\sin(2\pi f_0 t)}{2\pi f_0 t} & t \neq 0 \end{Bmatrix}, \qquad (2.7a)$$

where f_0 is the frequency of the sine function.

- The sinc(t) function is basically a mono-frequency sinusoidal function scaled by its own time.
- Again, the most desirable property of sinc(t) is its FT:

$$\text{sinc}(t) \Leftrightarrow \begin{Bmatrix} 1 & |f| \leq f_0/2 \\ 0 & |f| > f_0/2 \end{Bmatrix}. \qquad (2.7b)$$

That is, the FT of sinc(t) has a magnitude of one for frequencies in the interval ($-f_0/2$, $+f_0/2$) and zero elsewhere. It is a square (boxcar) function (Figure, www.physforum.com, 2013).

- The sinc(t) function is commonly used in the frequency domain for filtering (i.e., keeping frequencies in an interval (f_1, f_2) and discarding frequencies outside this range) as we will discuss in frequency filtering.

- More Internet resources are on: https://en.wikipedia.org/wiki/Sinc_function.

Convolution

- The convolution of two time-dependent functions x(t) and h(t) yields the time-dependent function y(t) given by:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau, \quad (2.8)$$

where * denotes the convolution operation.

- See these figures that explain the meaning of equation 2.8: Ex1, Ex2; Brigham, 1988).
- To convolve two sampled functions x(t) and h(t), we:
  1. Fold h(t) (i.e., reverse the order of its samples).
  2. Shift the folded h(t) all the way to the left until its last sample overlaps with the first sample of x(t). Note that this step is not necessary for h(t) that starts at t = 0.
  3. Multiply the first samples of x(t) and h(t). This is the convolution value at this shift.
  4. Shift the shifted and folded h(t) one sample to the right.
  5. Multiply overlapping samples of x(t) and the shifted and folded h(t) and add the result. This is the convolution value at this new shift.
  6. Repeat steps 4-5 until there is no more overlapping samples of x(t) and the shifted and folded h(t).
- For sampled functions x(t) and h(t) that have number of samples N_x and N_h, respectively, the convolution y(t) has a number of samples N_y given by:

$$N_y = N_x + N_h - 1. \quad (2.9)$$

- The convolution of a function x(t) with $\delta(t - t_0)$ is evaluated by simply shifting x(t) so that its vertical axis is positioned at $t = t_0$. That is: $x(t) * \delta(t - t_0) = x(t - t_0)$.

  Examples (Ex1, Ex2; Brigham, 1988).

- The convolution theorem states that convolution of two functions in the time domain is equivalent to multiplying their Fourier transforms in the frequency domain:

$$y(t) = x(t) * h(t) \Leftrightarrow Y(f) = X(f)\, H(f), \qquad (2.10)$$

  where: $Y(f) = |Y(f)|\, e^{i\phi_y(f)}$, $|Y(f)|$ and $\phi_y(f)$ are amplitude and phase spectra of y(t).

  - $|Y(f)| = |X(f)|\, |H(f)|$, $|X(f)|$ and $|H(f)|$ are amplitude spectra of x(t) and h(t).
  - $\phi_y(f) = \phi_x(f) + \phi_h(f)$, $\phi_x(f)$ and $\phi_h(f)$ are phase spectra of x(t) and h(t) (Ex1, Ex2; Brigham, 1988).

- The inverse of the convolution theorem is also true (Example; Brigham, 1988):

$$x(t)\, h(t) \Leftrightarrow X(f) * H(f). \qquad (2.11)$$

- Convolution is commutative. That is: $x(t) * h(t) = h(t) * x(t)$. This means that it doesn't matter if you fold and shift x(t) or h(t). Examples (Ex1, Ex2; Brigham, 1988).
- It is usually faster to perform convolution in the frequency domain for functions of long durations (i.e., N_x & $N_h > 32$) if the FFT is used.
- Numerical example of convolution.
- More Internet resources are on: https://en.wikipedia.org/wiki/Convolution.

Correlation

- The correlation of two time-dependent functions x(t) and h(t) yields the time-dependent function y(t) given by:

$$y(t) = x(t) \otimes h(t) = \int_{-\infty}^{\infty} x(\tau) h(t+\tau) d\tau, \qquad (2.12)$$

where ⊗ denotes the correlation operation. Examples (Ex1, Ex2; Brigham, 1988).

- To correlate two sampled functions x(t) and h(t), we:
  1. Shift h(t) all the way to the left until its last sample overlaps with the first sample of x(t).
  2. Multiply x(t) first sample with h(t) last sample and add the result. This is the correlation value at this shift.
  3. Shift the shifted h(t) one sample to the right.
  4. Multiply overlapping samples of x(t) and the shifted h(t) and add the result. This is the convolution value at this new shift.
  5. Repeat steps 3-4 until there is no more overlapping samples of x(t) and the shifted h(t).
- For sampled functions x(t) and h(t) that have number of samples N_x and N_h, respectively, the correlation y(t) has a number of samples N_y given by:

$$N_y = N_x + N_h - 1. \qquad (2.13)$$

- The correlation of a function with another different function is called crosscorrelation while the correlation of a function with itself is called autocorrelation.
- Applying the convolution theorem on the correlation yields:

$$y(t) = x(t) \otimes h(t) = x(t) * h(-t) \Leftrightarrow Y(f) = X(f)\, H^*(f) \qquad (2.14)$$

where $H^*(f) = H_r(f) - i\, H_i(f) = |H(f)|e^{-i\Phi h(f)}$ is the complex conjugate of $H(f) = H_r(f) + i\, H_i(f) = |H(f)|e^{i\Phi h(f)}$; $Y(f) = |Y(f)|\, e^{i\phi_y(f)}$; $\boxed{|Y(f)| = |X(f)|\,|H(f)|}$; $\boxed{\phi_y(f) = \phi_x(f) - \phi_h(f)}$.

- Note that correlation is NOT commutative. That is: $x(t) \otimes h(t) \neq h(t) \otimes x(t)$. This means that, given $x(t) \otimes h(t)$, you must shift $h(t)$.
- Numerical examples of cross-correlation and auto-correlation.
- More Internet resources are on: https://en.wikipedia.org/wiki/Cross-correlation and https://en.wikipedia.org/wiki/Autocorrelation.

The z-transform

- The z-transform of a sampled function $x(t) = (x_{-M}, x_{-M+1}, \ldots, x_{-1}, x_0, x_1, \ldots, x_{N-1}, x_N)$ is defined as:

$$X(z) = \sum_{k=-M}^{N} x_k z^k = x_{-M} z^{-M} + x_{-M+1} z^{-M+1} + \ldots + x_{-1} z^{-1} + x_0 + x_1 z + \ldots + x_{N-1} z^{N-1} + x_N z^N \quad (2.15a)$$

- Examples:
  - Given $x(t) = x(t=0,t=1,t=2,t=3) = (1, -1/2, 2, -1)$, its $X(z) = 1 - (½)z + 2z^2 - z^3$.
  - Given $x(t) = x(t=-1,t=0,t=1,t=2) = (3,-2, 4, -1)$, its $X(z) = 3/z-2+4z-z^2$.
- In this course, we will assume that time functions start at t=0 unless specified.
- The inverse z-transform is found by collecting the coefficients of the z-transform polynomial.
- For example, the inverse z-transform of $Y(z) = 2 + z - 4z^2 + z^3$ is $y(t) = (2, 1, -4, 1)$.
- Another example of Z-transform.
- The convolution theorem applies to the z-transform: $x(t) * h(t) = X(z) H(z)$.
- The z-transform is related to the Fourier transform through this relation:

$$z = e^{i2\pi f} \quad (2.15b)$$

  - For example, for x(t)=(2,-1):

$$
\begin{gathered}
X(z) = 2 - z \\
X(f) = 2 - e^{i2\pi f} \\
X(f) = 2 - \cos(2\pi f) - i\, sin(2\pi f) \\
\therefore X_r(f) = 2 - \cos(2\pi f)\ and\ X_i(f) = -\sin(2\pi f) \cdot \\
|X(f)| = \sqrt{[2 - cos(2\pi f)]^2 + [sin(2\pi f)]^2} \\
\emptyset_x(f) = \tan^{-1}\{[2 - \cos(2\pi f)]/[-\sin(2\pi f)]\}
\end{gathered}
\qquad (2.15c)
$$

- More Internet resources are on: https://en.wikipedia.org/wiki/Z-transform.

Phase considerations

- A wavelet is a time-domain signal that has a start and end times.
- A minimum-phase wavelet has its energy concentrated at its start time.
- A maximum-phase wavelet has its energy concentrated at its end time.
- A mixed-phase wavelet has its energy concentrated (or distributed) between its start and end times.
- A zero-phase wavelet is a non-realizable wavelet that starts from negative time and ends at positive time with its energy symmetric about t = 0. A non-realizable wavelet is one that cannot be produced directly by real sources.
- Examples.
- The wavelet shape and position with respect to t = 0 can be modified by changing the phase spectrum. (Example; Yilmaz, 2001)
- For processing, it is desirable to have a minimum-phase wavelet because deconvolution requires it; but the wavelet peak will be shifted further in time from the actual reflection time.

- For interpretation, it is desirable to have a zero-phase wavelet since the wavelet peak will occur exactly at the actual reflection time.
- It is common to do processing using minimum-phase wavelets and transform them to zero-phase wavelets before handing the data to the interpreter.

Frequency filtering

- Frequency filtering of a function x(t) means that we apply an operator g(t) to the function so that the output is a function y(t) that contains only frequencies that lie within our desired frequency range [f_i, f_f].
- Filtering can be done in the time-domain by convolving x(t) with g(t) or in the frequency domain by multiplying their Fourier transforms (the convolution theorem).
- Since we want to change only the amplitude spectrum of x(t), we must select a zero-phase g(t).
- The desired amplitude spectrum of G(f) is a square function that has a value of one in the range [f_i, f_f], and zero elsewhere.
- Because the FT exists only for continuous functions and G(f) is discontinuous at f_i and f_f, we must multiply G(f) by a taper (smoothing) function S(f) to eliminate this problem. Examples of taper functions include Hanning and Hamming.
- Furthermore, the sharp slopes in G(f) at f_i and f_f causes the output y(t) and Y(f) to be oscillatory because of Gibbs phenomenon. Therefore, relatively gentle slopes are used at the edges of the square function. (Example; Yilmaz, 2001)
- With these modifications, the final shape of G(f) will be a trapezoid with smooth edges.

- The trapezoid G(f) will have a gentle slope from $f_i - \Delta f_i$ to f_i, one between f_i and f_f, a gentle slope from f_f to $f_f + \Delta f_f$, and zero elsewhere.
- The slope at the high-frequency end should be gentler than that at the low-frequency end (i. e., $\Delta f_f > \Delta f_i$) because Gibbs phenomenon increases slightly with frequency.
- Shifting the band-pass range $[f_i, f_f]$ to higher frequencies will not improve the vertical resolution of the whole seismic trace. Instead, we have to do this by increasing the width of the pass-band frequency filter. (Examples 1 and 2; Yilmaz, 2001)
- Because of absorption of high frequencies by the Earth, a constant pass-band filter might not be sufficient and a time-variant filter (TVF) might be needed.
- TVF means that we segment the seismic record to 3 or 4 segments and design a filter for each segment separately.
- Real-data and numerical examples.
- More internet resources on tapering functions.

Frequency aliasing

- When sampling a continuous time-dependent signal, we must choose a sampling rate (sampling interval), Δt.
- When we sample a function using a sampling rate Δt, the highest frequency that can be retrieved is the **Nyquist frequency: $f_N = 1/(2\Delta t)$**.
- Common sampling rates used in seismic exploration are: 2, 4, and 8 ms. The corresponding Nyquist frequencies are: 250, 125, and 62.5 Hz, respectively. (Example; Yilmaz, 2001)

- Sampling in the time domain means that we multiply the continuous function h(t) by the time sampling function D(t), which is a series of impulses spaced at Δt with each impulse having an amplitude of one.
- Using the convolution theorem, this means that we convolve the amplitude spectra of h(t) and D(t) namely |H(f)| and D(f), where D(f) is also a series of impulses spaced at $\Delta f = 1/\Delta t = 2f_N$ with each impulse having an amplitude of $1/\Delta t$.
- Convolution of |H(f)| and D(f) will replicate |H(f)| at an interval of $2f_N$.
- |H(f)| naturally has a maximum usable frequency f_h beyond which there is no signal.
- If Δt is small such that $\mathbf{f_N > f_h}$; then, |H(f)| replicas will be totally separated and both f_N and f_h can be retrieved. This is the case **with no frequency aliasing**. (Examples 1 and 2; Brigham, 1988)
- If Δt is large such that $\mathbf{f_N < f_h}$; then, |H(f)| replicas will interfere and neither f_N nor f_h can be retrieved. This is the case **with frequency aliasing**. (Example; Brigham, 1988)
- When frequency aliasing occurs, the maximum retrievable frequency is the **alias frequency** given by: $\mathbf{f_a = |2f_N - f_h|}$. (Examples Ex2, Ex3; Yilmaz, 2001)
- For example, if a Vibroseis signal with a frequency range of 10-160 Hz was sampled at a sampling rate of 4 ms; then:
  - $f_N = 1/(2*0.004) = 125$ Hz
  - $f_h = 160$ Hz.
  - Since $f_N < f_h$, aliasing will occur.
  - The alias frequency is: $f_a = |2\times125 - 160| = 90$ Hz.

- This means that we lost all the frequencies above 90 Hz. Note that we could not retrieve neither f_N nor f_h because of aliasing.

- On the other hand, if we sampled the same signal at 2 ms sampling interval, f_N will be 250 Hz while f_h is still 160 Hz. Since $f_N > f_h$, no aliasing will occur and we can easily retrieve both f_h and f_N.
- Practically, we want to sample the data such that we avoid aliasing. In the field, we use an **anti-aliasing analog filter** *before sampling* to insure that $f_N > f_h$. The anti-aliasing filter makes f_h of the analog (continuous) data equal to $f_N/2$ or $3f_N/4$.

Gain Applications

- Example.
- Gain is a time-variant scaling of the amplitudes of the data.
- Gain is done to account for spherical divergence and absorption effects.
- There are two main types of gain methods:

  (1) Physical gain methods, which depend on subsurface velocities.

  (2) Display gain methods, which are done purely for display purposes without requiring velocity information.

(1) Physical Gain Methods:

(a) Single layer:

- A seismic source in a single homogeneous and isotropic layer produces a spherical wavefront whose energy decays as $1/r^2$ and amplitude as $1/r$, where r is the distance from the source.

- Therefore, in order to restore the amplitudes to their original values at the source (A_0), we have to multiply them by the distance traveled by the wave:

$$A_0 = A(r).r = A(t).V.t, \quad (2.16a)$$

where A(r) = A(t) is the amplitude at a distance r (and time t) from the source and V is the layer velocity.

- However, we usually do not know the distance traveled by the wave in the subsurface nor the medium velocity at this early stage of processing. Therefore, an approximation of the gained amplitude is:

$$A_0 = A(t).t, \quad (2.16b)$$

where A(t) is the amplitude at time t before gain application.

(b) Multiple layers:

- Refraction (ray bending) at layer interfaces increases the distance traveled by the ray to the same point, therefore increasing the amplitude loss due to geometrical spreading.
- If we know the RMS velocities (V_{rms}); then, the gained amplitude is:

$$A_0 = A(t).[V_{rms}^2(t).t]/(V_1^2.t_1), \quad (2.16c)$$

where $V_{rms}(t)$ is the RMS velocity at time t and V_1 and t_1 are the velocity and time at the bottom of the first layer.

- If we do not know the velocities; then, the gained amplitude can be approximated as:

$$A_0 = A(t).t^m, \quad (2.16d)$$

where m is a number between 1 and 2 found by trial and error until amplitudes are balanced equally at all times.

- If absorption effects are considerable, an exponential correction can be used such that the gained amplitude due to absorption only is:

$$AA_0 = A(t).e^{(\eta t)}, \qquad (2.17)$$

where $\eta = 0.01 - 1\ sec^{-1}$.

- Claerbout (1985) suggests using m = 2 in equation (2.16d) to account for both spherical divergence and absorption.

(2) Display Gain Methods:

Automatic Gain Control (AGC)

The most common type of AGC is the instantaneous AGC performed as follows:

(1) The absolute mean ($\overline{A_k} = (1/N)\sum_{i=k}^{k+N-1}|Ai|$) value of trace amplitudes within a time window containing samples k to k+N-1 is computed, where k=1,…,L-N+1, where L is number of samples in the trace.

(2) The trace samples within the time window starting with the k^{th} sample and ending with the $(k+N)^{th}$ sample are divided by the absolute mean amplitude value in that window.

(3) The time window is shifted down by one sample and steps (1)-(2) are repeated until sample L-N+1 is reached.

- In practice, 200 – 500 ms AGC time windows are commonly chosen.
- This method attempts to scale the amplitudes in the trace to 1.0 producing undesirable artifacts at shallow times (Figure).

- AGC should only be used to display the data.

Appendix A

Numerical convolution, cross-, and auto-correlations

x(t) = (2, 1, -2)

h(t) = (1, -1)

Convolution

y(t) = x(t) * h(t) = (2, -1, -3, 2) = h(t) * x(t) (Prove!)

	2	1	-2	
-1	1			y(0) = (2)(1) = 2
	-1	1		y(1) = (2)(-1) + (1)(1) = -2 + 1 = -1
		-1	1	y(2) = (1)(-1) + (-2)(1) = -1 - 2 = -3
			-1	1 y(3) = (-2)(-1) = 2

Back

Cross-correlation

y(t) = x(t) ⊗ h(t) = (-2, 1, 3, -2) ≠ h(t) ⊗ x(t) (Prove!)

	2	1	-2	
1	-1			y(-1) = (2)(-1) = -2
	1	-1		y(0) = (2)(1) + (1)(-1) = 2 - 1 = 1
		1	-1	y(1) = (1)(1) + (-2)(-1) = 1 + 2 = 3
			1	-1 y(2) = (-2)(1) = -2

Back

Auto-correlation

$y(t) = x(t) \otimes x(t) = (-4, 0, 9, 0, -4)$ [Note that: $y(-t) = y(t)$]

		2	1	-2		
2	1	-2				$y(-2) = (2)(-2) = -4$
	2	1	-2			$y(-1) = (2)(1) + (1)(-2) = 2 - 2 = 0$
		2	1	-2		$y(0) = (2)(2) + (1)(1) + (-2)(-2) = 4 + 1 + 4 = 9$
			2	1	-2	$y(1) = (1)(2) + (-2)(1) = 2 - 2 = 0$
				2	1	-2 $y(2) = (-2)(2) = -4$

Back

Z-Transform

$X(z) = (2)z^0 + (1)z^1 + (-2)z^2 = 2 + z - 2\,z^2$

$Y(z) = (1)z^0 + (-1)z^1 = 1 - z$

Back