SE 167/267A: Signal Processing and Spectral Analysis, WI 2024

Mayank Chadha

Assignment 3

Due Date: March 01, 2024

DTFS: Discrete Time Fourier Series

Consider the same analog signal that we had considered in the first assignment:

$$f(t) = \cos 2000\pi t + 5\sin 6000\pi t + 10\cos 12000\pi t + 20\cos 15000\,\pi t + 10\sin 20000\,\pi t$$

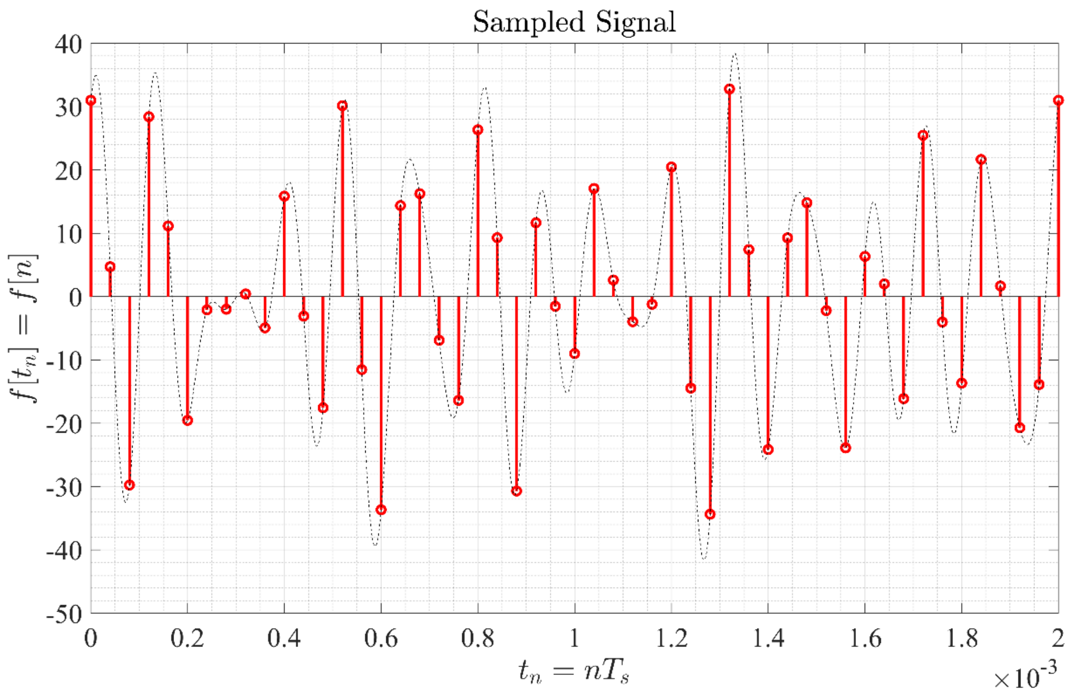
This function is periodic with fundamental period $T_p = 0.002\ seconds$, or $f_p = 500\ Hz$. The maximum frequency content in this signal is $f_{max} = 10000\ Hz$. Therefore, we would want to sample it at a sampling rate $f_s > 2f_{max}$.

Problem 1: Sample this signal between $t \in [0,0.002]$ seconds with sampling rate of $f_s = 25000\ Hz$ (greater than the Nyquist rate). First get the vector of sampled time $t_n \in [0,0.002]$ and then obtain the sequence $f[n] = f(t = t_n)$. Visualize the result by plotting the original analog signal using dotted black line and samples using red dots. Let L denote the number of samples in the signal $f[n]$. What is L?

Hint: To help you with plot formatting, I'll give you the code for creating this plot. The "stem" function is used to plot the sequence and the "plot" is used to plot the analog signal.

```
%------------------------
% Plot the signal
%------------------------
figure(1)
stem(sampled_time_vector, sampled_signal, 'r', 'LineWidth', 2);
hold on
plot(time_vector,signal,'--k')
grid on
grid minor
set(gca,'FontSize',20);
set(gca, 'FontName', 'Times New Roman')
xlabel('$t_n=nT_s$', 'Interpreter', 'latex');
ylabel('$f[t_n]=f[n]$', 'Interpreter', 'latex');
title('Sampled Signal', 'Interpreter', 'latex');
set(gcf,'PaperUnits','centimeters','PaperPosition',[0 0 32 18])
print(sprintf('Sampled Signal'),'-dpng','-r500');
```

You should get this plot:



Discrete Time Fourier Transform (DTFS):

Consider a periodic sequence $f[n]$ with period of N samples. Therefore, $f[n+N] = f[n]$. Since the frequency range for discrete-time signal is unique for the interval $\omega \in [-\pi, \pi]$ or $\omega \in [0, 2\pi]$. In this assignment, we stick to $\omega \in [0, 2\pi]$. Therefore, the discrete-time periodic signal can consist of frequencies separated by $\frac{2\pi}{N}$, or the k^{th} frequency is given by:

$$\omega_k = \left(\frac{2\pi}{N}\right)k, \quad \text{where}, k = 0,1,2,\ldots,(N-1).$$

Frequency ω is normalized angular frequency representing discrete signal. It is related to frequency in Hz (denoted by $\hat{f}$) by the relationship:

$$\omega = 2\pi\left(\frac{\hat{f}}{f_s}\right)$$

Corresponding to the discretized ω_k, we obtained discretized frequency in Hertz, denoted by $\hat{f}_k$ as:

$$\omega_k = 2\pi\left(\frac{\hat{f}_k}{f_s}\right)$$

For these discretized frequencies, we can obtain the Fourier coefficients called DTFS coefficients as:

$$F_k = \frac{1}{N}\sum_{m=0}^{N-1} f[m]e^{-i\omega_k m}$$

The sequence can be recovered using these Fourier Coefficients as:

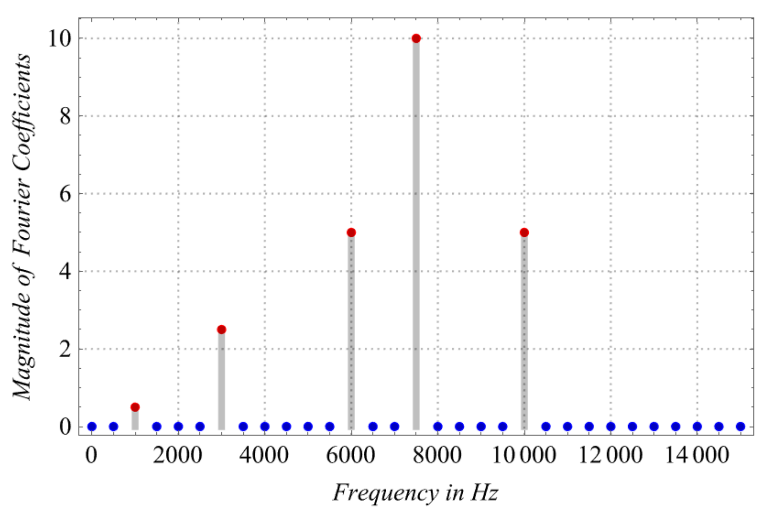
$$f[n] = \sum_{k=0}^{N-1} F_k e^{i\omega_k n}$$

Problem 2: For the sampled signal $f[n]$ in part 1, obtain (and display) the discrete frequencies ω_k and $\hat{f}_k$.

Problem 3: Obtain the DTFS F_k, its magnitude $|F_k|$, and phase $\angle F_k$. Create the following four plots (using "stem" function) with proper labeling of x and y axis:

1. Plot magnitude $|F_k|$ in y-axis with ω_k in x-axis. (use "abs" function in MATLAB)
2. Plot phase $\angle F_k$ in y-axis with ω_k in x-axis. (use "angle" function in MATLAB and "round" up to three digits)
3. Plot magnitude $|F_k|$ in y-axis with $\hat{f}_k$ in x-axis. (use "abs" function in MATLAB). Also mark the fundamental harmonic frequencies $\{0,1000,3000,6000,7500,10000\}Hz$ using vertical blue lines in your plot.
4. Plot phase $\angle F_k$ in y-axis with $\hat{f}_k$ in x-axis. (use "angle" function in MATLAB and "round" up to three digits)

Problem 4: Compare plot 3 obtained in Problem 3 with the similar plot obtained using the analog periodic signal in your previous assignment (see figure below). Carefully observe plot 3 and comment on why the DTFS magnitude $|F_k|$ vs. $\hat{f}_k$ plot obtained using *discrete signal* is not similar to the plot obtained using *analog signal* (It has only 5 frequencies for which magnitude $|F_k|$ was non-zero).



Problem 5: Using F_k obtained in part 3, reconstruct the original signal using inverse DTFS. Visualize the result by plotting the original analog signal using dotted black line and reconstructed samples using red dots. Also plot the reconstruction error rounded up to 3 digits in a separate plot.

DTFT: Discrete Time Fourier Transform

An aperiodic sequence $f[n]$ can be transformed into the "Fourier Transform" $F(\omega)$, were,

$$F(\omega) = \sum_{k=-\infty}^{\infty} f[k]e^{-i\omega k}$$

Clearly, $F(\omega)$ is a continuous function in $\omega \in [0,2\pi]$. The original discrete sequence can be obtained as:

$$f[n] = \frac{1}{2\pi}\int_0^{2\pi} F(\omega)e^{i\omega n}d\omega$$

Problem 6: Assume that the samples obtained in Problem 1 is for aperiodic function (since we had restricted sampling to $t \in [0,0.002]$). Obtain the Discrete Time Fourier Transform (DTFT) $F(\omega)$. Visualize the result by plotting the following:

1. $|F(\omega)|$ vs. $\omega \in [0,2\pi]$
2. $|F(\hat{f})|$ vs. $\hat{f}$. Here, $\hat{f}$ is frequency in Hz defined by $\omega = 2\pi\left(\frac{\hat{f}}{f_s}\right)$.

DFT: Discrete Fourier Transform

(Sampling in Frequency Domain)

Problem 7: Let L define number of samples in the sequence (see Problem 1). We can obtain DFT by sampling DTFT $F(\omega)$ with N samples, where $N \geq L$ (this avoids aliasing in "time domain"). Do the following:

1. Assume $N = L$. Obtain DFT, denoted by $F_k \equiv F[\omega_k] = f(\omega = \omega_k)$, by sampling DTFT at $\omega_k = \left(\frac{2\pi}{N}\right)k$, where, $k = 0,1,\ldots,(N-1)$. Visualize this result by plotting the following:
   - Absolute value of DTFT using solid blue line (continuous function $|F(\omega)|$). Using "stem" function, also plot the magnitude of DFT (magnitude $|F_k| \equiv |F[\omega_k]|$).
   - Phase of DTFT using solid blue line (continuous function $\angle F(\omega)$). Using "stem" function, plot the phase of DFT (phase of $\angle F_k \equiv \angle F[\omega_k]$).
2. Using "tic toc" function in MATLAB, evaluate and report the time required to obtain the samples F_k.

Hint: Use symbolic toolbox in MATLAB for this part.

FFT: Fast Fourier Transform

(Fast Algorithm to sample DTFT in frequency domain and obtain DFT)

Problem 8: Use "fft" function in MATLAB to obtain N-point DFT for $N = L$. Note that FFT is just an algorithm that obtains DFT (obtained in part 7) in a computationally efficient manner. For N-point DFT, the FFT algorithm essentially samples the DTFT at $\omega_k = \left(\frac{2\pi}{N}\right)k$, where $k = 0,1,\dots,(N-1)$. That is, the FFT result obtained in this part must be the same as the result obtained in Problem 7. Visualize this result by plotting the following:

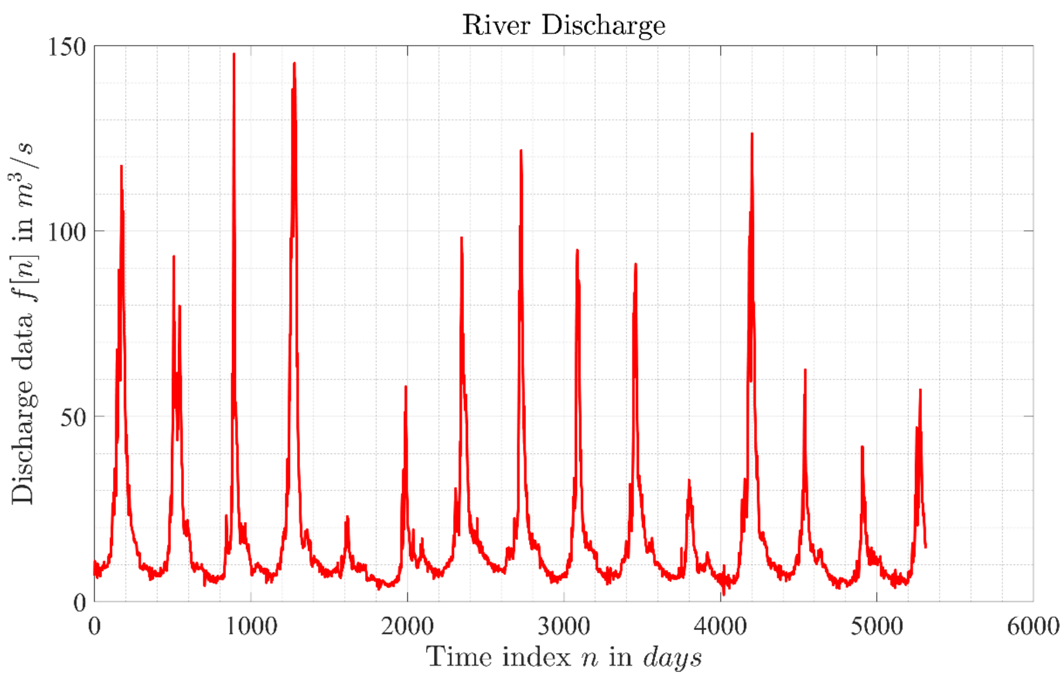
1. Absolute value of DTFT using solid blue line (continuous function $|F(\omega)|$). Using "stem" function, plot the magnitude of FFT on top of the DTFT magnitude plot.
2. Phase of DTFT using solid blue line (continuous function $\angle F(\omega)$). Using "stem" function, plot the phase of FFT on top of the DFT phase plot.

Problem 9: Using "tic toc" function in MATLAB, evaluate and report the time required to obtain the FFT. Based on the time obtained in obtaining DFT in Problem 7, how fast is FFT relative to brute force sampling?

Problem 10: What is the relationship to the DFT or FFT magnitude plot with the DTFS magnitude plot obtained in Problem 3. What is the significance of this relationship in sampling in the frequency domain and equivalently aliasing in time domain?

River Discharge Data Analysis

I am providing you with the excel sheet "Observations.csv" that has data for discharge of a river over $L = 5312\ days$. One value of data is collected per day, i.e., $f_s = 1\ sample/day$. The plot below shows the signal.



Problem 11: Since there are L number of data points in the time series $f[n]$, our goal will be to obtain N-point DFT with $N \geq L$. For this case, we assume $N = L$. Obtain the set of discrete normalized frequencies ω_k and the set of discrete frequencies $\hat{f}_k$ for $k = 0,1,\cdots,N$. Note that the unit of $\hat{f}_k$ is day^{-1}. Considering 365 days per year and by utilizing $\hat{f}_k$, obtain the discrete frequencies $\tilde{f}_k$ that has units in $year^{-1}$.

- Given $\omega_k \in [0,2\pi]$, what will be the lower and upper limit of $\hat{f}_k$ and $\hat{f}_k$?
- Plot the samples of ω_k, $\hat{f}_k$, and $\tilde{f}_k$ (use plot function) and appropriately label the x and y axis with correct units. Your x axis should represent the index k.

Problem 12: Obtain 5312-point FFT for the signal $f[n]$. You'll get the Fourier coefficients F_k corresponding to discrete frequencies of ω_k or $\hat{f}_k$ or $\tilde{f}_k$ (all of these are frequencies with different units). Plot the following:

- Plot $|F_k|^2$ vs. $\tilde{f}_k$. You'll realize that most of the dominant frequencies are low (concentrated near the zeros).
- Plot $|F_k|^2$ vs. $\tilde{f}_k$ but **restrict** the value of $\tilde{f}_k \leq 12$. This is because the frequencies $\tilde{f}_k$ has units in $year^{-1}$. Which means that the period corresponding to k^{th} frequency is $\tilde{f}_k^{-1}$. By restricting $\tilde{f}_k$ to 12, we are considering periodicity of up to 1 year. That is, for example, for $\tilde{f}_k = 1\ year^{-1}$, the period is $\tilde{f}_k^{-1} = 1\ year$; for a period of $\tilde{f}_k = 2\ year^{-1}$, the period is $\tilde{f}_k^{-1} = 0.5\ year = 6\ months$, and so on. Using the plot, report the first 3 dominant frequencies in the signal and interpret each of these frequencies and comment on why you think that these are the most dominant frequencies for the given river discharge signal.
- In a single plot, using "subplot" in MATLAB consisting of 3 rows, plot the following measures of FFT's magnitude (appropriately label the axes):
  - First row: Plot $|F_k|$ vs. $\tilde{f}_k$
  - Second row: Plot $|F_k|^2$ vs. $\tilde{f}_k$
  - Third row: Plot $10\log_{10}|F_k|$ vs. $\tilde{f}_k$

  What do these three measures of the FFT represent, and what are the situations where these magnitude measures of the FFT are suitable?

Problem 13: For window length of 365×4 (4 years), overlap sample size of $365 \times 4 - 30$ (30 days short of the window size), and number of FFT samples equal to the window length of 365×4 (minimum requirement to avoid time aliasing), obtain the STFT coefficient matrix S, frequency vector F, and time vector T using "spectrogram" function in MATLAB:

```
[S, F, T] = spectrogram(x,window,noverlap,nfft,fs)
```

The frequency vector `F` that MATLAB yields has units of day^{-1} (same as the units of the sampling frequency `fs`). Transform `F` in $year^{-1}$ and:

- Plot the spectrogram **restricted** to frequency of 12 $year^{-1}$ (we know that most of the frequences are closer to zero). Make sure to label all the axes appropriately, display the color bar with appropriate label and the frequencies have units in $year^{-1}$. Create three spectrogram plots (use 3 rows of "subfigure" in MATLAB) corresponding to:

- First row: Plot the spectrogram corresponding to the magnitude of the coefficient $|S|$.
- Second row: Plot the spectrogram corresponding to the energy density $|S|^2$.
- Third row: Plot the spectrogram corresponding to $10 \log_{10}|S|$.

For each of these plots, make sure to show the color bar and label what the "colorbar" represents. A "colorbar" is very important to correctly represent the spectrogram.

Hint: Use "surf" function in MATLAB. For example, see the screenshot for the plotting the spectrogram corresponding to the magnitude of the coefficient $|S|$. Of course you should use subplots to plot all three spectrogram.

```
figure(17)
surf(T, F, abs(S), 'EdgeColor', 'none');
axis tight;
view(0, 90);
xlabel('Time (days)', 'Interpreter', 'latex');
ylabel('Frequency $\tilde{f}_k$ ($year^{-1}$)', 'Interpreter', 'latex');
title('Spectrogram');
colorbar;
ylabel(colorbar, 'Magnitude of STFT $|\mathcal{G}|$', 'Interpreter', 'latex');
set(gca,'FontSize',14);
set(gca, 'FontName', 'Times New Roman')
set(gcf,'PaperUnits','centimeters','PaperPosition',[0 0 32 18])
print(sprintf('River_Spectrogram '),'-dpng','-r500');
```

- Plot the waterfall plot for the same data (with frequency restricted up to 12 $year^{-1}$) using the energy density $|S|^2$. Make sure to label all three axes appropriately. Hint: Use "waterfall" function in MATLAB.
- Interpret the spectrogram and the waterfall plot and report your observations.

Problem 14: Obtain wavelet transform using the "cwt" function in MATLAB. There is no need to limit the frequency value in this case.

```
[cfs,frq,scales] = cwt(river_data,fs);
```

- Obtain the corresponding sampled time vector using $t_n = nT_s$. The sampled time vector should have a length of 5312.
- Transform the frequency vector `frq` into $year^{-1}$.
- Create three wavelet plots (use 3 rows of "subfigure" in MATLAB) corresponding to:
  - First row: Plot the wavelet intensity corresponding to the magnitude of the coefficient $|W|$.
  - Second row: Plot the wavelet intensity corresponding to the energy density $|W|^2$.
  - Third row: Plot the wavelet intensity corresponding to $10 \log_{10}|W|$.

For each of these plots, make sure to show the color bar and label what the "colorbar" represents and set y-axis to log scale. A "colorbar" is very important to correctly represent the wavelet.

Hint: Use "surface" function in MATLAB. For example, see the screenshot for the plotting the spectrogram corresponding to the energy density $|W|^2$. Of course, you should use subplots to plot wavelet intensity using all three metrics.

```
figure(20)
surface(sampled_time_vector2,frq,abs(cfs).^2)
axis tight
shading flat
xlabel('Time (days)', 'Interpreter', 'latex');
ylabel('Frequency ($year^-1$)', 'Interpreter', 'latex');
title('Wavelet Transform Coefficients');
set(gca,'FontSize',14);
set(gca, 'FontName', 'Times New Roman')
colorbar;
ylabel(colorbar, '$|\mathcal{W}|^2$', 'Interpreter', 'latex');
set(gca,"yscale","log")
set(gcf,'PaperUnits','centimeters','PaperPosition',[0 0 32 18])
print(sprintf('River_Wavelet_Energy_Density'),'-dpng','-r500');
```

- Plot the waterfall plot for the same data using the energy density $|W|^2$. Make sure to label all three axes appropriately. Use the code below for plotting:

```
figure(22)
waterfall(frq,sampled_time_vector2,abs(cfs)'.^2)
set(gca,XDir="reverse",View=[30 50])
xlabel('Frequency ($year^-1$)', 'Interpreter', 'latex');
ylabel('Time (days)', 'Interpreter', 'latex');
zlabel('Energy Spectral Density $|S|^2$', 'Interpreter', 'latex');
set(gcf,'PaperUnits','centimeters','PaperPosition',[0 0 32 18])
set(gca,'FontSize',14);
set(gca, 'FontName', 'Times New Roman')
print(sprintf('WaterFall Plot Wavelet'),'-dpng','-r500');
```

- Interpret the wavelet and the waterfall plot and report your observations. Comment about the uncertainty principle based on the waterfall plot obtained for the spectrogram and wavelet analysis. What is the primary difference between the STFT and wavelet analysis in terms of tackling the time-frequency resolution. Comment in detail.

Problem 15: Download an mp3 file of your favorite song. Using the .mp3 file, extract the data and the sampling frequency using a code similar to the following code:

```
[Y,FS] = audioread('MJ.mp3');
duration_seconds = size(Y, 1) / FS;
duration_time_vector= linspace(0, duration_seconds, size(Y, 1));
player = audioplayer(Y(:,1), FS);
play(player);
```

Notice that your data "Y" may have multiple columns corresponding to different channels used in recording. You can play and use one of the channels (the code above plays and utilizes channel 1). Obtain the spectrogram, and the waterfall plot. Comment on your observations.