EOS 423 // EOS 518

Lab 2.1: Practice with Fourier Analysis

Due: 1:30 pm March 7, 2025

You have one week to complete this assignment and upload your responses as a PDF to Brightspace. You are not excluded from working with others (pairs are recommended), but each person will submit their own copy of the assignment. **Responses to questions should be typed, using complete sentences and standard grammar.** Double check that your image resolution is high enough to read. If you write your response in a word processor, please export to .PDF before submitting your response.

A pair of students will be randomly selected on Monday to initiate and lead a discussion of the assignment. Be prepared to show your progress and discuss any challenges you still face.

Name: ______________________________

Question	1	2	**Total:**
Marks:	14	6	20
Score:			

1 Introduction to Fourier Analysis

In the natural sciences, the procedure to search for periodic signals in datasets goes by various names, including spectral analysis, harmonic analysis or Fourier analysis. The latter name refers to Jean-Baptiste Joseph Fourier (1768 – 1830), the French mathematician who was especially interested in heat transfer and vibrations (he is also generally credited for the discovery of the greenhouse effect). Using his considerable smarts, Fourier realized that any function $x(t)$, so long as it was single-valued (i.e., one value of $x(t)$ for every value of t), could be represented as the sum of a series of cosine waves taking this form:

$$x_k(\theta) = A_k \cos(k\theta - \phi_k) \tag{1}$$

where θ is an angular variable for a function $x_k(\theta)$, with a period of 2π radians. Thus, for a given frequency k, the value $x(\theta)$ is determined by A_k (the amplitude of the wave) and the phase angle ϕ_k (if $\phi_k = \pi/2$ radians, or 90°, then ϕ_k could be dropped and the cosine function replaced with the sine function). If we rely on the following trigonometric identity for difference between two angles, R and S:

$$\cos(R - S) = \cos(S)\cos(R) + \sin(S)\sin(R) \tag{2}$$

we can re-write equation 1 as:

$$x(\theta) = A_k \cos(\phi_k)\cos(k\theta) + A_k \sin(\phi_k)\sin(k\theta) \tag{3}$$

and because the phase angle is constant for a given frequency, we can define constants $a_k = A_k \cos(\phi_k)$ and $b_k = A_k \sin(\phi_k)$, and the general formula for a cosine wave becomes:

$$x_k(\theta) = a_k \cos(k\theta) + b_k \sin(k\theta) \tag{4}$$

a sum of both a cosine and a sine function. Returning to Fourier and his realization, we can now write a formula for the *Fourier series*:

$$x(\theta) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(k\theta) + \sum_{k=1}^{\infty} b_k \sin(k\theta) \tag{5}$$

$$\tag{6}$$

where we have simplified for the case when k = 0. This series means that any $x(t)$ could be expressed this way, where $x(t) = x(\theta)$. At this point, it is helpful to express θ as a function of t, our original independent variable, and T, the duration of $x(t)$, which is set to the period of the sinusoid, 2π. Hence, the first non-zero frequency in our Fourier series (also called the *fundamental* or *first harmonic* frequency) is $f_1 = 1/T$. All the other non-zero-frequencies in a Fourier series are harmonics (integer multiples) of this fundamental frequency, i.e. the second harmonic is $f_2 = f_1 + \Delta\text{f} = 2f_1$, third harmonic is $f_3 = f_2 + \Delta\text{f} = 3f_1$. Thus, the k^{th} frequency in our summation is

$f_k = f_{k-1} + \Delta\text{f} = \text{k}f_1$. This enables us to write the general form for the Fourier series of a time function $x(t)$ of period T:

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(\frac{2\pi k}{T}t) + \sum_{k=1}^{\infty} b_k \sin(\frac{2\pi k}{T}t) \tag{7}$$

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(2\pi f_k t) + \sum_{k=1}^{\infty} b_k \sin(2\pi f_k t) \tag{8}$$

If we next rely on Euler's formulas:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta) \tag{9}$$

$$e^{-i\theta} = \cos(\theta) - i\sin(\theta) \tag{10}$$

where e is Euler's number ($e = 2.718281828459....$) and i is the imaginary number $i = \sqrt{-1}$. If one does enough algebra and trigonometry, one can rewrite Equation 8 as a *complex Fourier series*:

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi i f_k t} \tag{11}$$

The complex coefficients c_k tell you the required *amplitude* of the cosine and sine waves for each frequency f_k needed to reproduce $x(t)$. They relate directly to the real coefficients a_k and b_k according to the following formulas:

$$c_k = \frac{1}{2}(a_k - ib_k) \tag{12}$$

$$c_k = \frac{1}{T}\int_0^T x(t)(\cos(2\pi f_k t) - i\sin(2\pi f_k t))dt \tag{13}$$

$$c_k = \frac{1}{T}\int_0^T x(t)e^{-2\pi i f_k t}dt \tag{14}$$

2 Discrete Fourier Transform

Of course, while mathematicians can deal with functions or time series that stretch to infinity, scientists are always dealing with finite numbers (Fig. 1). Measurements taken from a stratigraphic section, for example, have a beginning and an end (covering some period T). The dataset will consist of a series of values (numbered n_0, n_1, n_2, etc.), making up some total number of samples (N). Each sample is separated by a *sampling interval* (dt, assumed here to be equal - *this is very important!*). Any resulting time series or stratigraphic series (x_n) is always going to be a discrete approximation of some continuous function ($x(t)$, representing, for example, changing sea level with time.

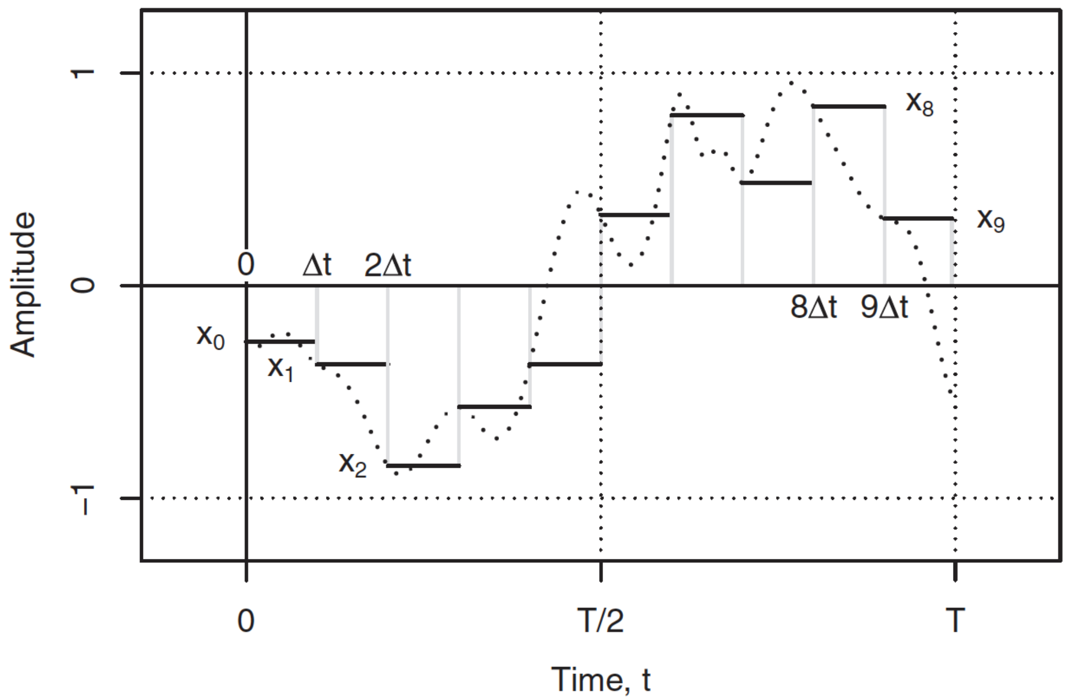

Figure 1: a discrete approximation of a continuous function

Thus, we need an approach for describing the sinusoids needed to describe a *discretized* time series, x_n (Fig. 1). In other words, we need a discrete version of equation of Equation 14, and this equation takes this form:

$$X_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-2\pi i \frac{kn}{N}} \tag{15}$$

where X_k is being used in place of c_k (but is otherwise the same as c_k). Equation 15 is the basic forward discrete Fourier transform. The index k (frequency) also ranges between $0 \leq k \leq N-1$, like n (time). Thus, equation 15 is really a system of linear equations:

$$\underline{\mathbf{X}} = \begin{pmatrix} X_0 \\ X_1 \\ . \\ . \\ . \\ X_k \\ . \\ . \\ . \\ X_{N-1} \end{pmatrix} \text{ and } \underline{\mathbf{x}} = \begin{pmatrix} x_0 \\ x_1 \\ . \\ . \\ . \\ x_n \\ . \\ . \\ . \\ x_{N-1} \end{pmatrix} \tag{16}$$

with Equation 15 becoming:

$$\underline{\mathbf{X}} = \mathbf{F}\underline{\mathbf{x}}$$

and performing a *fast Fourier transform* solves it through matrix algebra, the lowest frequency resolved is $1/T$, where T is the total time represented by the time series (also equal to $N \cdot dt$). If N (the number of data points) is even, the highest resolved is frequency $1/(2dt)$, where dt is the sampling interval. In between these two end-points, each resolved frequency is separated by $1/T$. With these FFT coefficients (X_k), which again tell you the *amplitude* of each cosine and sine wave for each frequency k, one can always exactly recreate the original time series x_n.

When doing Fourier analysis with the aim of identifying periodic signals, it is important to remember this fact: *any time series can be described as a summation of periodic signals*, including pure white noise (Fig. 2). If you have 100 data points that you think records sea level change over the last 100 years, and you discover that it can be perfectly described by 50 sinusoids, you have not discovered that sea level has 50 nested cycles - you have simply demonstrated the Fourier relationship (Equation 1). Instead, what we are looking for with such an analysis is: does a periodic signal explain a significant amount of the variation within a time series? If a significant peak is located, does this

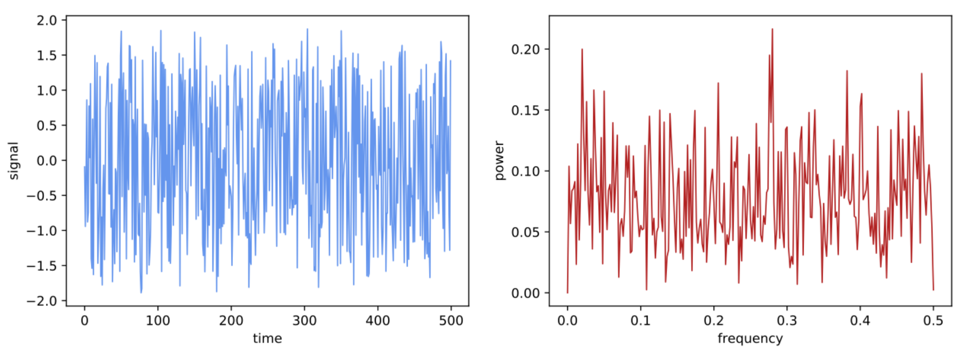

Figure 2: A time series x_n made of white noise (left), and the fast Fourier transform (X_k) on the right. The y-axis of the FFT indicates the relative contribution of each sinusoid frequency in explaining the total signal, and can be used to exactly reproduce the original blue curve.

periodic signal *make sense* as a potential driver of Earth system change (i.e., is it a known or expected frequency of change in climate, etc.)? Keep these concepts in mind as you explore two different datasets below!

Questions

Question 1 (14)

THE LAST 5 MILLION YEARS The file `LR04.csv` contains a re-sampled time series of δ^{18}O values measured from benthic foraminefera calcite, compiled from 57 globally-distributed Ocean Drilling Program deep sea cores. The following questions will have you explore different aspects of this dataset, in both the time domain (i.e., plotting and playing with the time series itself) and the frequency domain (i.e., performing fast Fourier transforms of the data). For questions 1b,c,e,f below, you will be producing a figure with two subplots (one on top, another below), with the variations of the raw data signal in the first subplot (i.e., different portions of the full dataset, or with different kinds of pre-processing). In the second subplot, you will show plot the discrete Fourier transform frequency power of that data slice. Instead of writing your own Fourier transform for these questions, check out methods fft, rfft, fftfreq, and rfftfreq in scipy.fftpack.

(a) (2 points) Make a plot of all the data, with δ^{18}O on the y-axis and age (in ka., or thousands of years before present) on the x-axis. On the same figure, plot again the data from the two time intervals that you will be analyzing separately below (*interval 1:* between 0 and 599.5 ka., *interval 2:* between 1500 and 2998 ka.). Your plot should have three colors: one to indicate interval 1, another for interval 2 and a third for the rest of the data. What is the time resolution (dt) for each interval?

(b) (2 points) Perform a discrete Fourier transform of the raw data from interval 1 (*i.e., do not process or modify in any way*). Make a figure with two subplots, with the upper subplot showing δ^{18}O vs. age and the lower sub-plot of showing its DFT frequency spectrum (frequency on the x-axis and amplitude on the y-axis). What is the dominant peak, and what feature of the raw data does this peak correspond to?

(c) (2 points) Now, subtract away the mean of the interval 1 dataset (so `newdata = data - np.mean(data)`), and then perform a discrete Fourier transform of this "zero-mean" data. Make a figure with two subplots, with the upper sub-plot showing δ^{18}O (zero-mean) vs. age and the lower sub-plot of showing its DFT frequency spectrum (frequency on the x-axis and amplitude on the y-axis). What has changed about the DFT?

(d) (2 points) What are the significant peaks in the DFT of the zero-mean, interval 1 dataset? Using your knowledge of Earth system science, what do these peaks correspond to? When considering "significance," look for peaks that are local maxima in the spectra. Are some peaks more clear than others?

(e) (2 points) Let us turn to interval 2 now. As you did in question 1c, first remove the mean from the interval 2 data, and then perform a discrete Fourier transform. Make a figure with two subplots, with the upper sub-plot showing δ^{18}O (zero-mean) from interval 2 vs. age, and the lower sub-plot showing its DFT frequency spectrum (frequency on the x-axis and amplitude on the y-axis). Which frequency explains the most variance in the analyzed data, and how do you account for this peak?

(f) (2 points) Now, for interval 2, remove both mean of the data and the linear trend before performing the DFT. A nifty way of doing this in Python uses the `signal` submodule from `scipy` (`from scipy import signal`). Once imported, typing `newdata = signal.detrend(data)` will remove BOTH the mean and any linear trend in the dataset. Make a figure with two subplots, with the upper sub-plot showing δ^{18}O (de-trended and zero-mean) vs. age from interval 2, and the lower sub-plot showing its DFT frequency spectrum (frequency on the x-axis and amplitude on the y-axis). How does this spectra compared to the spectra for interval 1 that you produced in question 1d?

(g) (2 points) Taking a step back to your figure from 1a, what do these analyses suggest about the important controls on climate for the last 5 million years? Why is this dataset an appropriate one to ask this question?

Question 2 (6)

A MYSTERY COUPLET DATASET The file `mystery.csv` is a list of lamina thicknesses (N = 208) in units of millimeters. The first row has the oldest lamina, and the last is the youngest. Each lamina is defined by a coarser, sandy layer at the base and a fine mud cap at its top (a sand-mud couplet).

(a) (3 points) For the data in `mystery.csv`, first remove both the mean and any linear trend in the couplet thickness data. Make a figure with two subplots, with the upper sub-plot showing couplet thickness (de-trended and zero-mean) on the y-axis vs. couplet number on the x-axis, and the lower sub-plot showing its DFT frequency spectrum (frequency on the x-axis, and amplitude on the y-axis). What is the dominant frequency? Given the units for couplet thickness (these couplets are teeny-tiny!) and this frequency, what signal do you think is being recorded here?

(b) (3 points) A high frequency signal dominates the DFT of this couplet dataset. To explore whether lower frequencies signals are also present and significant, load the datafile `smoothed-mystery.csv`, which is a 7-point running mean of the couplet dataset. Remove both the mean and any linear trend from this smoothed dataset, and make a figure with two subplots. The upper sub-plot will show couplet thickness (smoothed, de-trended and zero-mean) on the y-axis vs. couplet number on the x-axis, and the lower sub-plot will show its DFT frequency spectrum (frequency on the x-axis, and amplitude on the y-axis). What is the dominant frequency? Given your answer to question 2a, what signal do you think is being recorded here?