Mini-project in Astronomical Data Analysis

Dr. Michelle Lochner

Instructions

The aim of this mini-project is to ensure your skillset is a good match to the project requirements. I also hope it will be fun! There's no machine learning knowledge required here, instead you will make use of standard mathematical tools to analyse some astronomical data.

You must use python 3 for your coding and should make use of numpy and matplotlib, with the following import convention:

```
import numpy as np
import matplotlib.pyplot as plt
```

You will write a report on your analysis in a single pdf file. Please use this naming convention: `<family name>_<first name>_project.pdf`. You should use latex to create this report. Please do not submit any python files or notebooks. Instead, include your python code in an appendix at the end of your report, using an appropriate latex package to ensure correct formatting and highlighting. Indicate clearly which question each part of the code is addressing.

Report structure

Your report must include the following:

- An introduction of no more than two pages (including figures) written for an audience of fellow students at an Honours level, who have some background in physics but do not have much knowledge of astronomy or periodograms. Your introduction must be fully referenced using a standard referencing style (with a bibliography included at the end of the report) and must include the following topics:
    - The Fourier transform, periodogram and Lomb-Scargle periodogram
    - Variable stars and particularly Cepheid variables
    - Why Cepheid variables are so important in astronomy
    - The importance of the Fourier transform in astronomy and how it is used in variable star analysis
- After the introduction, you must present your results for each of the five questions below. Introduce each set of results with an appropriate sentence that summarises what you did (as if the reader has not read the original question). Ensure any plots are clear and easy to read.
- An appendix containing your python code, clearly indicating which part of the code addresses which question. Ensure all code is included, such as library imports and plotting function calls.

Question 1 - Fourier transform

The Fourier transform is one of the most important tools for any scientist but particularly in astronomy. For this question, you will use python to take the fast Fourier transform (FFT) of a "top hat" signal. **You must use the function `np.fft.fft()` for this question**. Write code to replicate the plots below, which show the top hat signal on the left and its Fourier transform on the right.

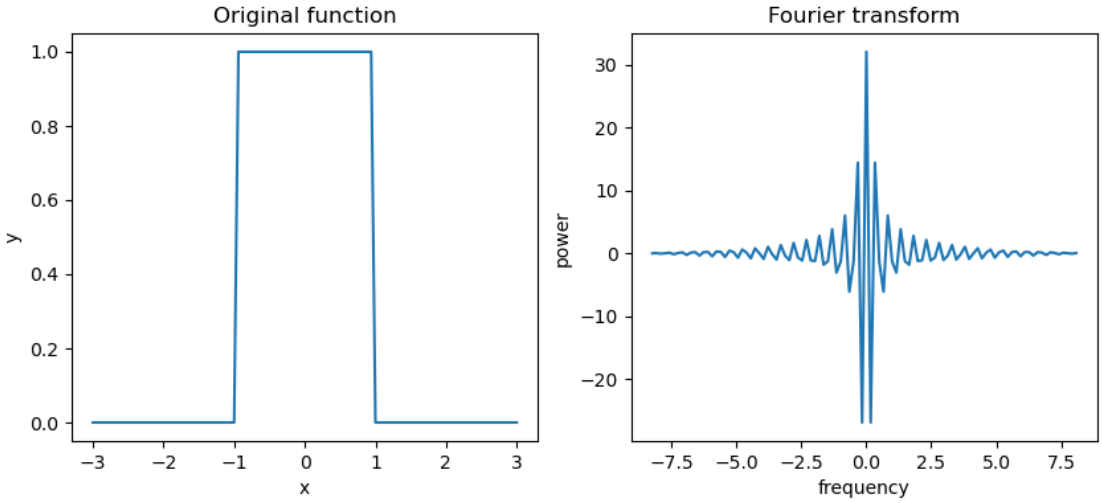


Figure 1: Recreate this plot for this question. The left plot is the input signal, the right plot is the real component of the FFT of the input signal.

Question 2 - Periodograms

A common problem in astronomy and other disciplines is to try to locate repeating or periodic signals in data (often time series data). The raw FFT needs to be converted to something more useful called a periodogram to help extract periodic signals from data. The periodogram can very clearly show if there are dominant periodic frequencies in the signal and what values they have.

Generate data for the following function, where x is in the range of -3 to 3, with 1000 samples:

$$y = \sin(2\pi\omega_1 x) + \sin(2\pi\omega_2 x), \tag{1}$$

where $\omega_1 = 2$ and $\omega_2 = 4$.

Compute the periodogram for y. You must use the `np.fft.fft` function and calculate the periodogram using it[1]. Make a plot with two panels: on the left show the input function and on the right show the periodogram. You should see two very clear peaks at the values of 2 and 4 (i.e. the input frequencies).

Question 3 - Cepheid variables

I have given you some real Cepheid variable data in the form of a text file labeled `variable.dat`. Each row represents a new observation of the cepheid variable, taken with an optical telescope (in V band). The first column is the time of the observation in days (actually technically Heliocentric Julian Day minus 2450000), the second column is the magnitude of the observation and the third column is the uncertainty on the magnitude.

Install the powerful astronomy package `astropy` and use the `LombScargle` package to compute the Lomb-Scargle periodogram for your data. You will make use of this periodogram to learn the intrinsic period of this variable star. You can then use this period to perform "light curve folding" to recover the true structure of the variable star's light curve. Produce the following for this question:

1. Plot the Lomb-Scargle periodogram for your data
2. Write down the period of your Cepheid, which corresponds to the peak of the periodogram.
3. Plot the original data.
4. Plot the folded light curve as a function of phase.
5. Write a short (roughly one paragraph) discussion on why the original data looks so messy and how light curve folding works to recover a clean signal.

[1] For future reference, the scipy.signal.periodogram function is an easier way to do this, but you must demonstrate your understanding by doing this the hard way with np.fft.fft.

Question 4 - Calculating distances

The modern version of the period-luminosity that Henrietta Leavitt discovered is:

$$M = -2.76(log_{10}(P) - 1) - 4.16, \tag{2}$$

where P is the period in days. To convert this to distance, a standard distance equation can be used:

$$d = 10^{(m-M+5)/5}, \tag{3}$$

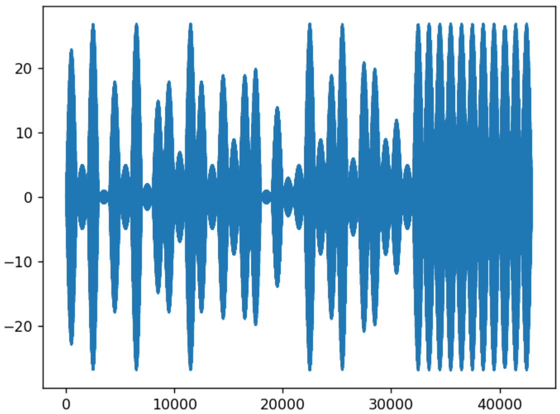
where m is the average apparent magnitude of the light curve, M is the absolute magnitude as defined above and d is the distance, measured in parsecs (a commonly used distance measure).

Compute the distance to your cepheid variable star. Write it in your answer sheet in parsecs and also in kilometres.

In the early 1900's, it was thought that the Milky Way galaxy was the entire universe and that all the galaxies that could be seen in the night sky (referred to as "nebulae") were within our galaxy. Write a short paragraph about the role that Cepheid variables played in settling this important scientific debate. Comment on whether or not this distance measurement that you just calculated could have helped settle the debate.

Question 5 - a strange signal

For this question, we will imagine that you are a successful astronomer, observing with the MeerKAT telescope. You are analysing your data (`signal.txt`) when you notice a strange signal that doesn't look like any astronomical source that you've ever seen[2].



Upon careful study, you notice that every 1000 samples, the signal seems to change. While scientists have been fooled before[3] you can't help wondering, could this signal be a message from intelligent life somewhere out there in space?

By analysing the message, you are able to work out a translation from amplitude to a character (see table at the end of this assigment - note that amplitude 27 corresponds to a space character). Write python code to determine the maximum of the signal in 1000 sample chunks and work out the message from the aliens.

Hidden message

Whilst working on this, an extremely bright physics student notices that the frequency *also* changes every 1000 samples, even though it's less obvious than the amplitude. Use the skills you've developed to decode the second secret message hidden in the frequency modulation.

[2]This question is heavily inspired by the creative brain of my husband and the excellent book and movie Contact (https://en.wikipedia.org/wiki/Contact_(1997_American_film))

[3]https://www.space.com/38916-pulsar-discovery-little-green-men.html

Character	Amplitude	Frequency
a	1	50
b	2	55
c	3	60
d	4	65
e	5	70
f	6	75
g	7	80
h	8	85
i	9	90
j	10	95
k	11	100
l	12	105
m	13	110
n	14	115
o	15	120
p	16	125
q	17	130
r	18	135
s	19	140
t	20	145
u	21	150
v	22	155
w	23	160
x	24	165
y	25	170
z	26	175
	27	180
:	28	185
/	29	190
=	30	195
?	31	200
.	32	205

Character	Amplitude	Frequency
0	33	210
1	34	215
2	35	220
3	36	225
4	37	230
5	38	235
6	39	240
7	40	245
8	41	250
9	42	255
A	43	260
B	44	265
C	45	270
D	46	275
E	47	280
F	48	285
G	49	290
H	50	295
I	51	300
J	52	305
K	53	310
L	54	315
M	55	320
N	56	325
O	57	330
P	58	335
Q	59	340
R	60	345
S	61	350
T	62	355
U	63	360
V	64	365
W	65	370
X	66	375
Y	67	380
Z	68	385