Project 2: Radio Data Analysis

Contact: zarif.kader@mail.mcgill.ca

Office Hours: Scheduled on Request with min 18 hrs notice

Submission Format: All code, and accompanying text submitted in .pdf, .py, .ipynb. All questions should be answered and code should be well commented along with a brief qualitative description of any outcome. Figures should have proper axis labels and all files should clearly indicate which problems they are associated with as a comment on the top of the file and in the file name.

Code Software: For this assignment you are expected to write the bulk of your code in python. This project has been designed with sigpyproc in mind, which can be installed as a python module.

Goal: The main goal of this project is to introduce you to how calibration is done, how to understand noise and noise statistics, and how to search for signals and to try and optimize them.

We will use data from CHIME, a radio telescope observing between 400-800 MHz located near Penticton, B.C.. The data provided is from the CHIME pulsar project.

Part 1: Calibration

Due Date: February 20th

We have raw radio data that we want to calibrate to a known source. We have pointed our telescope towards this source and now need to calibrate our raw data, i.e. measured in ADC counts, to physical units, i.e. measured in Jy flux units.

1. Our calibrator source is 3C 129. It is a point source At 400 MHz the spectral flux is $S_{400} = 8.684$ Jy with a spectral index of $\alpha = -0.6$. Plot the spectrum of this source over the CHIME frequency band.

2. Load in the file *calibrator_source.fil*. Obtain the average spectrum of this calibrator source and plot it. Compare it to spectrum of the source, are there any odd features?

3. Create a transfer function, converting the ADC counts to Jy units. Apply it to the calibrator source data and create a waterfall plot. How does the data look, are there any odd features?

4. Load in the file *blank_sky.fil*. Apply the calibration transfer function to this dataset. Plot the average spectrum and create a waterfall plot. How does the data look, are there any odd features?

5. Calculate the mean and standard deviation per channel. Normalize each channel and create a waterfall plot of the dataset. What type of statistics describe the noise of this dataset?

Part 2: Search For a Pulsar

Due Date: March 11th

With our telescope calibrated, we can now search for signals. We've been told there's a pulsar in our telescope's field of view and we want to see if we can detect it.

1. Somehow, we do not know the period of the pulsar but we do know the spectrum. It has a spectral density at 400 MHz of $S_{400} = 0.15$ Jy and a spectral index of -1.5. We have measured the noise of the blank sky. Assume it is the same in the direction of this pulsar. Note the signal-to-noise ratio (S/N) is proportional to the ratio of source flux density to the system equivalent flux density (which we obtained as the average spectrum of the blank sky data from the previous part) and the total number of samples being integrated,

$$S/N \propto \frac{S_\nu}{S_{sys}}\sqrt{N}.$$

   What is the expected S/N ratio of 1 single pulse at? How many times should we then need to fold or average over the pulsar period to observe the pulsar with $S/N \sim 2$?

2. Load in the file *pulsardata.fil*. You should find the S/N for 1 pulse to be quite small, therefore ensure RFI contaminated channels are flagged. Normalize the data per channel to standardize the statistics across all channels.

3. Radio waves propagating through the ISM are dispersed by the intervening plasma and have an arrival time delay depending on the observing frequency. The arrival time delay is given by,

$$\tau(f, DM) = k_{DM} DM \left(\frac{1}{f^2} - \frac{1}{f_{ref}^2} \right).$$

   Our two parameters to search over are the DM and the period. If the channels are properly normalized then we can average across frequencies to further increase our S/N. Create a grid of dms to search over using the function *dmt_transform()* to obtain the S/N as a function of DM and time, $S/N(DM, t)$.

4. To search for the periodic signal, we will do it in frequency space. Take the Fourier transform of $S/N(DM, t)$ over time to obtain $S/N(DM, f)$. Plot a waterfall plot of this function. Obtain the S/N maximizing DM and period.

5. With the S/N maximizing parameters use the *fold()* function to fold the dataset into 1024 channels and 1 single pulse. Plot the waterfall for the folded pulse profile. Average over frequency to obtain the frequency averaged pulse profile and plot it.

Part 3: Search For Unknown Signals

Due Date: March 25th

We have searched for a known signal, so now let us search for an unknown signal. As discussed in class, we will focus on searching for signals with a negative dispersion measure, that is the observed value of DM is $DM < 0$, to try to detect signs of extraterrestrial intelligence. DM is a quantity set by electron density ISM and is correlated to the distance of propagation. To understand typical DM, n_e, and galactic distances, here are two commonly cited models for the electron density content of our galaxy are NE2001 and YMW16. Suppose extraterrestrial intelligence has knowledge of this. We could then infer they might transmit signals with a negative DM to counteract the propagation effect for other extraterrestrial intelligent life to detect. For this part, you are tasked to search for signatures by:

1. Creating a basic model for this hypothetical signal.
2. Creating and applying a search method to find this signal in the radio data provided.
3. Characterizing the detection limits of your search applied to the data.

Your search should incorporate negative DM as a parameter search. Any other parameters to include in the search to identify a signal is up to your discretion and will depend on the type of signal to want to search for.

Every student will have a data file labeled with their student id of ∼10 s of a different exposure to the blank sky. You may include other student's files into your search but ensure your file is a part of your search. You may use the blank sky calibration from part 1 to convert to units of Jy. You are encouraged to discuss with your peers about how to code your search and how to claim a signal detection and conversely set detection limits. To guide you and provide a starting point, make sure you think about and incorporate these elements into your search:

1. You need to construct some model for the signal you are trying to detect. For example, if you want to search for signals from extraterrestrial intelligence, do

you expect the signal be repeating? Will it be broadband? Will it be of a certain radio luminosity or spectral index? What are the key observables that you can use to claim a detection? Consider how, as an astronomer, you might consider a signal to be alien in nature or, conversely, how you might send a signal for extraterrestrial life to detect.

2. In order to claim a signal detection, you need to have some criteria that must be met in your data. You should then apply these criteria to your search. From your model of the signal, you should have some selection cuts that will be applied in the search parameter space. For example, a S/N threshold cut or a $DM < 0$ cut.

3. It is important for any search to characterize the limits of your search, i.e. your detection limit. One method to check your search for potential false positives and understanding your noise statistics to apply your search to regions of data with no signal in it. For example, applying it to *blank_sky.fil*, a separate/independent observation of the blank sky, as a data set where no signals are expected.

4. The most difficult part of searching for unknown signals is to set constraints or exclusion limits on your search parameter space. You should have a null hypothesis that you can use to set these constraints. A full statistical treatment of your model is not expected, however you should be able make a statement based on your search. For example, a statement like "For a signal of type **X**, we report no detections above a S/N threshold of **X** which implies no signals originating from $<$ **X** pc with a flux $>$ **X** Jy.".