Initial Data Reduction Tutorial - IPTA student workshop, 2018

Mayuresh Surnis

June 12, 2018

Abstract

In this tutorial, you'll learn to use the PRESTO pulsar data analysis software for doing initial data reduction and quality checks AND LOOKING AT THE DATA.

1 General Introduction

The data coming out of a radio telescope is usually 2 dimensional with time and frequency as the two axes. Although the exact data format depends on the telescope back-end, when one looks at the data, it is typically in either *filterbank* or *psrfits* format. Although both these data formats contain the time-frequency information, it is the header information that is different as well as how the data area arranged and processed. We are going to use one filterbank file `B1907+10.fil` which obviously has an observation of PSR B1907+10 (J1909+1102). So, lets get started!

2 Viewing the Metadata

Each filterbank file has a header, which contains important information about the telescope, back-end, observing frequency, bandwidth etc. This can be achieved by using the command:

```
> readfile B1907+10.fil
```

This will print out all the information contained in the header. Go through the information and list out the following things:

1. Telescope name
2. Observation frequency
3. Bandwidth
4. Bandwidth of a single channel
5. Sampling time
6. Integration time
7. Source co-ordinates (we will come back to this in the end!)

3 Determining the DM of the pulsar

In order to get a useful time of arrival (TOA) from any data, the time-frequency array needs to be de-dispersed to the dispersion measure (DM) of the pulsar. Now, typically, if you are trying to get a timing solution for a new pulsar, you may not have a very accurate DM value to begin with. Let us see what happens if we don't have a very good value of the DM and how we can correct for it. To do this, run the following command:

```
> prepfold -o 1907 -topo -dm 145.0 -n 128 -p 0.283646828 -nosearch B1907+10.fil
```

This will take a lot of time to run(!@#&) and will bring up a plot you may be familiar with (from the talk earlier). What we are interested in here is the frequency sub-band plot in the middle of the figure. You will notice that the pulse appears slanted towards left. This means that the DM we have used to de-disperse the signal is incorrect.

Looking at the sub-band plot, can you say whether the DM is lower or higher than the true DM of the pulsar?

We can actually calculate how wrong the DM is (roughly) from the plot. Dispersion causes lower radio frequencies to travel slower (and hence delayed) with respect to the higher radio frequencies. The delay in ms is given by:

$$\triangle t \simeq 4.15 \times 10^6 \times (f_1^{-2} - f_2^{-2}) \times DM \tag{1}$$

Where both frequencies are in MHz.

Using the information from the header, how much DM offset do we have?

4 Determining the period of the pulsar

In order to reduce the time taken to run the analysis, we will reduce the data size for this exercise by producing a de-dispersed *time series* of the data. This can be done using the command:

```
> prepdata -o 1907_truedm -nobary -dm 149.982 B1907+10.fil
```

This will produce a time series file by the name `1907_truedm.dat`. For determining the effect of wrong period, we will be using this file.

In the last example, we actually used the correct (topocentric) period of the pulsar, so we already know what it is. Let us pretend for the time being that we don't know it yet!
Let us start with a period of 283.5 ms. This can be done through the command:

```
> prepdata -o 1907_truedm -nobary -dm 149.982 B1907+10.fil
```

If you now run the prepfold command:

```
> prepfold -o 1907_truedm_wrongp -topo -p 0.28364 -n 128 -nosearch 1907_truedm.dat
```

This will bring up the time sub-integration plot. We have forced prepfold to fold at the wrong period through the nosearch option.

What is the effect of folding at the wrong period on the pulse profile shape?

Based on the direction of the slant, can you say whether the period we are using is higher or lower than the true period?

In order to calculate the error in the period, we will use the following formalism. Let us call the error in period as δP. Now, we have accumulated a total error of δT (the difference in time between the two bins at the extreme ends of the slant) over an integration time of T. Thus,

$$\delta P = \delta T \left(\frac{P}{T} \right) \tag{2}$$

Where all the quantities are either in ms or s (basically, same units).

Using the above equation, how much is the error in period in our folded profile?

5 Using the profile S/N to estimate the true period

The signal-to-noise ratio (S/N and NOT SNR. SNR is supernova remnant!) of a profile can be defined as the ratio of the mean signal level to the rms of the noise in the *OFF-pulse* region. For a binned profile with N bins, OFF-pulse rms of σ, a mean of $\bar{p}$ is given by:

$$S/N = \frac{1}{\sigma\sqrt{W_{eq}}} \sum_{i=1}^{N} (p_i - \bar{p}) \tag{3}$$

Where W_{eq} is the equivalent width of a square (top-hat shaped) pulse of same area as the pulse from the pulsar. In this section, we are going to get our hands dirty with a lot of calculations! Get ready!

Typically, a program like prepfold would try to do the same thing as we are going to do to figure out the correct pulse period during search analysis. We are going to set-up a grid of trial periods, produce folded profiles, calculate profile S/N and figure out the true period, which is the period at which, we get the maximum profile S/N. Lets start!

1. Set-up a period grid with values separated by 1 μs between 283.64 and 283.65 ms.
2. Produce folded profiles at each trial period using the prepfold command (as in the previous section).
3. Calculate profile S/N for each trial period using Equation 3.
4. Find out the maximum S/N and the corresponding trial period.
5. (Optional) Set-up a finer grid (say 0.5 μs) to obtain a better period estimate.

6 Real-world problems!

If you remember, the co-ordinates for the source and source name itself were not present or garbage in the file header. This happens sometimes due to wrongful entries made by people (which may be you as well!) or sometimes instruments or sometimes, just stupid bugs in the code which fail for very particular cases. This is the reason why astronomers will always have their own version of the observation logs. These situation usually arise without warning, will usually catch you off guard and may render the entire observation utterly useless. We carried out this little exercise in order to be alert and prepared for such situations and impress the fact that it is ALWAYS useful to have at least a customary look at the data before using it for further analysis. Best of Luck!

7 Optional task for if you are interested

You can get another file containing a known pulsar. As you know already, the header does not have co-ordinates, so there is no way for you to know which pulsar it is! The file is called `new_pulsar.fil`. All I will say is that there is a pulsar there with a DM between 150 and 155. Happy pulsar hunting!