Searching for and Finding Pulsars

2015 NRAO REU Student Workshop

Goals

In this activity you will work with real data, going through the steps to find and identify an otherwise unknown pulsar. You will examine the radio frequency interference present during the observation, as well as the time and Fourier domain representations of the data. Using programs from the `PRESTO` software suite, you will identify candidate pulsars, fold them, and ultimately find the real pulsar.

Introduction

Searching for pulsars is a computationally demanding and time consuming task. You will be going through a simplified version of this process to illustrate the key steps, which are:

1) Identify and remove radio frequency interference (RFI).
2) Remove (partially) the effects of the interstellar medium (ISM) through the process of de-dispersion and create a time-series at a number of trial dispersion measures (DM).
3) Fourier transform the time-series.
4) Search for statistically significant signals in the Fourier-transformed data. This may involve searching for pulsars that are accelerating due to their presence in a binary system.
5) Filter the candidate pulsars.
6) Fold the most promising candidates.
7) Inspect the results and decide which candidates are worthy of follow-up.

You will be using some pre-existing software to gain a better understanding of each step. The data you will analyze comes from the Green Bank North Celestial Cap pulsar survey and is known to contain a pulsar. Once you have successfully identified the pulsar you will extract a pulse time of arrival from the observation.

Dispersion and De-dispersion

Radio waves, like all electromagnetic radiation, travel at the speed of light, *c*, in a vacuum. But space is not a vacuum. Rather, it is pervaded by the interstellar medium (when referring to the space between stars in galaxies), which is in part an ionized plasma. As the radio waves emitted by pulsars travel through this ionized plasma, they move with a speed less than *c*, thus arriving at the Earth later than they would had they traveled through a true vacuum. Furthermore, the amount that they slow down depends on the radio frequency, so that signals at a relatively high frequency travel faster than signals at a lower frequency. This is known as dispersion, and the difference in travel time between radiation emitted at two different radio frequencies is known as the dispersion delay.

The formula for calculating the dispersion delay between two frequencies is

$$\Delta t = \frac{e^2}{2\pi m_e c} \times \left(\frac{1}{f_1^2} - \frac{1}{f_2^2}\right) \times DM$$

where e is the electron charge, m_e is the electron mass, f_1 and f_2 are the lower and higher radio frequencies (expressed in MHz), and DM is the dispersion measure. The prefactor with the all the physical constants turns out to be about 4.15×10^6 ms.

The dispersion measure is a combination of the number density of electrons encountered by the radio signal and the distance traveled. Expressed mathematically it is defined as

$$DM = \int_0^d n_e dl$$

where n_e is the electron number density, *d* is the distance to the source of the radio waves, and the integral is taken over the path between the source and the observer, *l* (you may recognize this as a line integral, and the DM as a column density). The units for DM are parsecs (an astronomical unit of distance) per cubic centimeter, or pc/cm^3.

Why does all this matter? Suppose a pulsar emits a broadband signal, meaning that the signal is emitted at a range of radio frequencies. The high frequency components of this pulse will arrive at the Earth *before* the low frequency components. What will this do to the pulsar signal?

Your task: Calculate the dispersive delay in a pulsar signal.

Suppose we try to observe pulsar with a spin period of 3 ms and a dispersion measure of 20 pc/cm^3. What is the dispersive delay *relative to the 400 MHz* for signals at the following frequencies: 400 MHz, 375 MHz, 350 MHz, and 325 MHz? How does the dispersive delay between the highest and lower frequencies compare to the period of the pulsar? What do you think will happen if we try to naively detect a pulse from this pulsar *without taking into account the dispersive delay*?

As the above example illustrates, we need to account for the dispersive delay if we want to detect pulsars, especially millisecond pulsars. When searching for pulsars, we use a technique known as *incoherent dedispersion*. This involves breaking the radio signal down into discrete radio frequency channels of finite width, and then subtracting the appropriate dispersive delay from each channel, thus aligning all of the pulses. This does *not* complete remove the effects of dispersion, however, because individual radio frequency channels have a finite width, and so still suffer a small dispersive delay between the top and bottom of the channel. This is often referred to as dispersive smearing. We cannot remove this through incoherent dedispersion.

Your task: Calculate the dispersive smearing for a pulsar observation.

Suppose we are observing a pulsar at a center frequency of 350 MHz using a bandwidth of 100 MHz broken down into 4096 individual frequency channels, with a sampling time of 81.92 microseconds. The period of the pulsar is 3 ms and the DM is 20 pc/cm3. What is the dispersive delay in the center frequency channel?

We will break this down into several parts.

1) What is the width of an individual frequency channel?

2) The formula for dispersive delay in a frequency channel of width Δf >> f is

$$t_{DM} = 8.3 \times 10^6 ms \times DM \times \frac{\Delta f}{f^3}$$

(As an extra exercise, you can try to derive this from the formula for dispersive delay. Hint: try using a Taylor expansion).
What is the dispersive smearing for this observing set up at the center frequency of 350 MHz?

3) How does this dispersive smearing compare to the pulsar period? How does it compare to the sampling time used for this observation?

As you can see, dispersive smearing requires that we use large numbers of frequency channels if we want to retain sensitivity to fast pulsars. Unfortunately, we can only measure the DM of a pulsar *after* we discover it. So how do we find new pulsars, for which the DM is not known ahead of time? The only way to do so is to de-disperse our data at many trial DMs and search the data for new pulsars, hoping we get lucky. A trial DM is unlikely to match exactly the true DM of a new pulsar, so there almost always be *additional* smearing from using a non-ideal trial DM. For this reason, the step size between trial DMs must be small, so that the DM error is not too large. This requires using many trial DMs, each of which must be searched independently, and that requires lots of computing power.

In PRESTO, di-dispersion is handled by one of two programs: `prepsubband` and `prepdata`. `prepsubband` is more useful when creating many de-dispersed time series. The following command was used to generate 10 de-dispersed time series for you to work with:

```
prepsubband -o guppi_55366_GBNCC16246_0087_0001 -numout 1458152 -lodm 17.75
-dmstep 0.1 -numdms 10 -mask guppi_55366_GBNCC16246_0087_0001_rfifind.mask
guppi_55366_GBNCC16246_0087_0001.fits
```

As before, the `-o` flag specifies the output file name and the last argument is the name of our raw data. The `-numout` flag specifies the number of output data points, which must be even for when we take the Fourier transform (more on this soon). The `-lodm` flag specifies the lowest DM that we will create a time series for, `-dmstep` is the step size between DMs (both of these are in units of pc/cm^3) and `-numdms` specifies the number of trial DMs to use. The `-mask` flag specifies the RFI mask that will be applied to the data.

Exploring a de-dispersed time series

A de-dispersed time series is created by subtracting the appropriate dispersive delay from each frequency channel, after which the frequency channels are added together, resulting in a record of the power detected by the telescope as a function of time. You will explore one of the 10 time series created using the `prepsubband` command, as described above.
Your Task: Explore a de-dispersed time series

Use the following command to get a graphical representation of a time series file:

```
exploredat guppi_55366_GBNCC16246_0087_0001_DM18.25.dat
```

You can use the following commands to zoom in/out, and move left and right in the file:

Button or Key	**Effect**
Left Mouse or I or A	Zoom in by a factor of 2
Right Mouse or O or X	Zoom out by a factor of 2
<	Shift left by a full screen width
>	Shift right by a full screen width
,	Shift left by 1/8 of the screen width
.	Shift right by 1/8 of the screen width
?	Show this help screen
Q	Quit

Explore the data by zooming in and out, moving right (forward in time) or left (back in time), and noting the axis labels. Write down anything you notice about the data, such as the general trend, abrupt changes, etc. You already know that there *is* a pulsar in this data---do you see any pulses? Does that surprise you? What about this data confuses you? What makes sense?

The Fourier Transform

Searching for a pulsar with unknown period in the time domain is very difficult. Unless the pulsar is bright enough that individual pulses can be clearly identified (and most pulsars *aren't* that bright), we would need to fold the data at a wide range of trial periods and look for a pulsar-like signal. That isn't feasible in a blind search. Luckily, we have a tool that is designed exactly for identifying periodic signals in complex data: the Fourier transform.

Entire college courses are taught on the Fourier transform. Obviously, that is beyond the scope of this activity, so we will just touch on some of the basics. The Fourier transform of a function *f*(*x*) is defined as

$$F(k) = \int_{-\infty}^{+\infty} f(x)\, e^{-2\pi i x k} dx$$

where *i* is the imaginary number. This is the definition of the Fourier transform for a continuous function, but in practice we work with discretely sampled data. The discrete Fourier transform (DFT) is then defined as

$$X_k = \sum_{n=0}^{N-1} x_n\, e^{-2\pi i k n/N}$$

This may seem a little bit confusing, so it may make more sense if we take advantage of a pretty cool identity:

$$e^{ix} = cos(x) + i\, sin(x)$$

This let's us write the DFT as

$$X_k = \sum_{n=0}^{N-1} x_n\ [cos(2\pi kn/N) - i\, sin(2\pi kn/N)]$$

Let's describe this in words. It turns out that any function can be described as a sum of many (potentially infinite) sine and cosine functions, each with different frequencies and amplitudes. The DFT of a function tells us the contribution of these sine and cosine functions at each possible frequency.

As an example, suppose our data is noise free and contains a single cosine function with a frequency of exactly 2 Hz. The DFT of this data would be zero at every frequency *except* 2 Hz, where we would see a spike in power.

Pulsar signals are not usually well described by a single sine or cosine. Instead, they are better described by a repeating Gaussian function (this is the classic bell curve). But a repeating Gaussian can be described by (in principle) an infinite sum of sine and cosine functions with different frequencies. In practice, however, not all of these sines and cosines contribute significantly to the final Gaussian. If a Gaussian has a width δ, then it turns out that the number of sinusoids we need to adequately describe the function is proportional to $1/\delta$. Furthermore, if the Gaussian repeats with a frequency of f, then these additional sinusoidal components will have frequencies of $2*f$, $3*f$, $4*f$, … . These higher frequency components are known as *harmonics*.

This is a lot to of information to digest! So take a few minutes to discuss with your neighbor the preceding section, and ask some questions if you have them.

Exploring a DFT of the Time Series

You will now take advantage of a PRESTO program to create DFTs of the de-dispersed time series. Use the command

```
realfft guppi_55366_GBNCC16246_0087_0001_DM18.25.dat
```

Your task: Explore the DFT of the time series

Use the command

```
explorefft guppi_55366_GBNCC16246_0087_0001_DM18.25.fft
```

Next, press the “N” key, followed by “r”. This will normalize the values by their raw power. Now use the following commands to explore the file:

Left Mouse or I	Zoom in by a factor of 2
Right Mouse or O	Zoom out by a factor of 2
<	Shift left by a full screen width
>	Shift right by a full screen width
,	Shift left by 1/8 of the screen width
.	Shift right by 1/8 of the screen width
N	Re-normalize the powers by one of several methods
G	Go to a specified frequency
?	Show this help screen
Q	Quit

As with the time series, write down some general features of the DFT, looking at things like axis labels, trends, and frequencies that have a lot of power. What do you notice about these frequencies? What could might be the source of some of these signals? In our examples we discussed *noiseless* data, but our data has noise in it. How does this affect the DFT? What is the behavior of the DFT at very low frequencies (approaching 0 Hz)? Write your observations in the space below.

Searching a DFT for Pulsars

In principle, to find a pulsar, we need to look for the Fourier frequencies with significant power, fold the original data at those frequencies, and then determine if the signal is a pulsar or RFI. In practice, there are some subtleties to consider. For example, pulsars have significant Fourier power at their harmonic frequencies. If we sum the potential harmonics, we can find weaker pulsars, but this requires some care. We also want to find *binary pulsars* whose apparent periods are changing due to Doppler acceleration during the course of an observation. This spreads power out among different Fourier frequencies. We need to account for this to find binary pulsars. Luckily, tools exist that take these and other issues into consideration to produce a list of candidate pulsars. One such tool is called `accelsearch`. We will use it now.

Your task: Use accelsearch to find candidate pulsars in the DFTs

Issue the command

```
accelsearch -zmax 0 -sigma 5.0 guppi_55366_GBNCC16246_0087_0001_DM18.25.fft
```

The option `-zmax` says to search only for signals with a Fourier acceleration (z) of 0. In other words, we will not be searching for binary pulsars right now. Note that we are also cheating a bit here by ignoring everything with a significance (sigma value) less than 5.0, because we already know there is a strong pulsar in this data. This just helps reduce the number of candidates you need to consider.

`accelsearch` creates several output files. One group of files ends in ACCEL_0. You can see these files by typing `ls`. These are text files that contain a list of the candidate pulsars found by accelsearch. Use the following command to explore each file:

```
less -S guppi_55366_GBNCC16246_0087_0001_DM18.25_ACCEL_0
```

You can then use the left, right, up, and down arrow keys to look at the list of candidate. Focus on the top few rows, which are a summary of what was found. What do you make of these candidates? Do any look particularly interesting to you?

Sifting and Folding

We have only searched 1 DM, but typically we search 10s of thousands! That produces a lot of candidates, some of which may be duplicated at different DMs, of marginal significance, or known RFI. So we typically use some intelligent sifting algorithm to reduce the number of candidates that are actually interesting enough to fold, though that number can still be in the hundreds for a single observation. This is a complication you should be aware of, but that we will not get into now.

Folding is a technique you have probably heard of before. When we fold data, we divide it into segments that are each one rotation period long (for the purported rotation period of the candidate). We then add these segments. If there is a true periodic signal in the data, then whenever the signal is “on”, the correlated signal should increase linearly, but whenever the signal is “off”, the uncorrelated noise level should increase as the square root of the number of additions. As such, we should see the ratio of signal and noise (not surprisingly called the signal-to-noise ratio, or SNR, or S/N) increase as $N/N^{0.5} = N^{0.5}$, and our signal should get stronger. In this way we can see a pulsar whose individual pulses are actually lower than the noise level!

In PRESTO, folding is accomplished using the command `prepfold`. The most important thing we need to know when folding a candidate is the period. There are many ways to specify this in `prepfold`, but the simplest is using the `-p` option.

Your task: Fold some of the candidates from your search

Examine the list of candidates for various trial DMs. Choose some that you think might be pulsars, and examine the output files to find the period of the candidate. Then issue the following command:

```
prepfold -p <period> guppi_55366_GBNCC16246_0087_0001_DM18.25.dat
```

where you replace `<period>` with the candidate period **in seconds**.

`prepfold` should work pretty quickly. Examine the output plot. Do you think this is candidate is an actual pulsar? Ask one of the activity leaders for their opinion.

Now we will fold the full raw data file, retaining frequency information (remember that we summed over radio frequency when making the time series files). Issue the following command:

```
prepfold -p <period> -dm 18.25 guppi_55366_GBNCC16246_0087_0001.fits
```

This command will take a little longer to run. Once it is done we will discuss the result and answer any other questions you have about the activity. Nice work!