The Lomb-Scargle Algorithm

Alex Matheson & Austin Nhung

What is it?

- Lomb-Scargle methods are a way to analyse signals with inconsistent time series
- Generates data on underlying periodicity

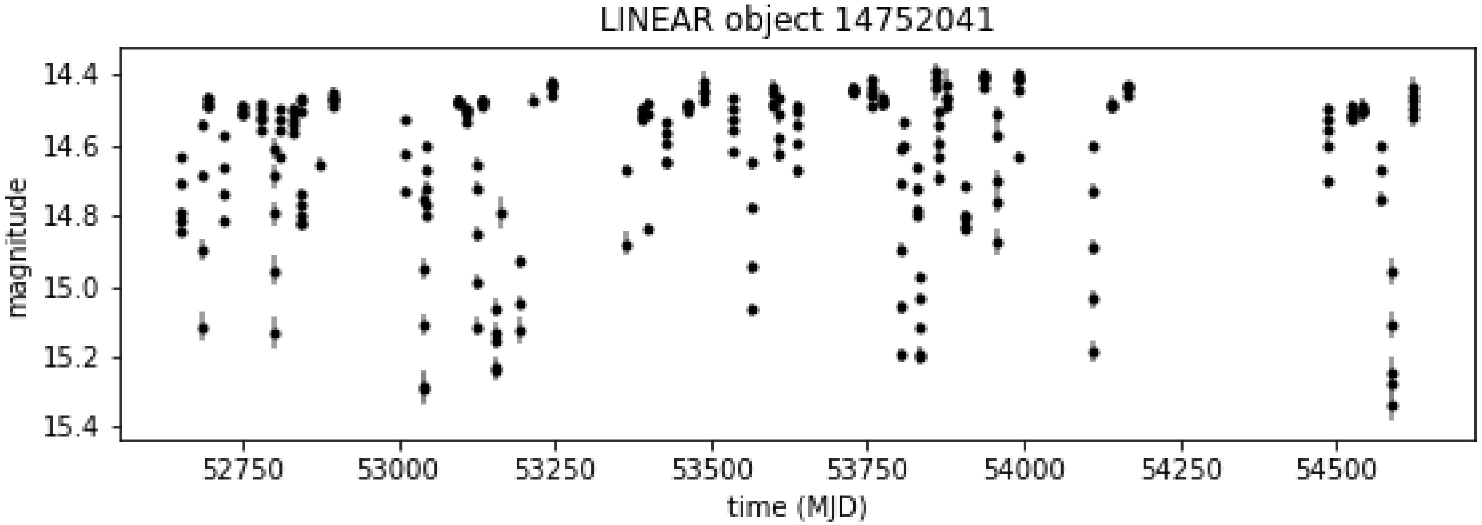


FIG 1. Example data showing light intensity of a star (LINEAR catalogue number 14752041) over time.

Where is it used?

- Many applications where it is not possible to constantly record, or there are gaps in records of past events.
- Sees heavy use in astronomy – weather, day/night cycle, seasonality all can prevent consistent data capture
- Important in climate science – used to deal with gaps in sediment and ice cores

What is a Periodogram?

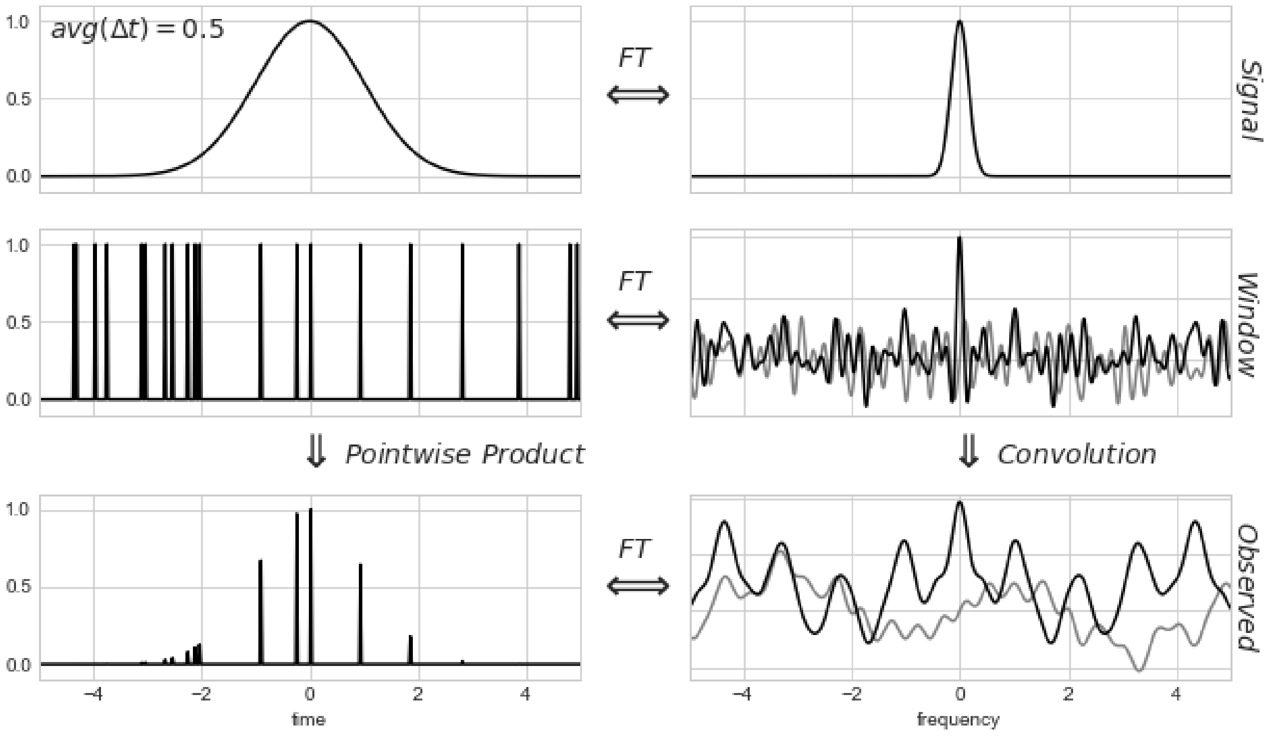
- An <u>estimate</u> of the power spectral density, characterizing the signal in terms of the power of each frequency
- The classical periodogram is the square of the Fourier transform, and can be computed with the DFT

$$S(\omega) = \frac{1}{N}\left\|\sum_{j=1}^{N} x(t_j)e^{-i\omega t_j}\right\|^2$$

$$= \frac{1}{N}\left(\left(\sum_{j=1}^{N} x_j \cos(\omega t_j)\right)^2 + \left(\sum_{j=1}^{N} x_j \sin(\omega t_j)\right)^2\right)$$

Why not FFT?

- For the FFT, an uneven delta-comb sampling window is asymmetric, and has a noisy Fourier inverse. This gives a noisy result.
- Spectral leakage from high frequencies to low frequencies is also a common problem



Source: Vanderplas 2017

Mathematics

- Relies on Fourier series. Instead of Fourier transforming the data straight away, it determines the frequencies of interest, then fits coefficients to the frequencies using a least-squares method.

$$\phi \approx \mathbf{A}x$$

$$\mathbf{A} = \begin{bmatrix} \sin(\omega_0 t_0 + \phi_0) & \sin(\omega_1 t_0 + \phi_1) & \dots & \sin(\omega_n t_0 + \phi_n) \\ \sin(\omega_0 t_1 + \phi_0) & \sin(\omega_1 t_1 + \phi_1) & \dots & \sin(\omega_n t_1 + \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\omega_0 t_n + \phi_0) & \sin(\omega_1 t_n + \phi_1) & \dots & \sin(\omega_n t_n + \phi_n) \end{bmatrix} \qquad x = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Choosing Frequencies

- Different frequencies may be selected for the A matrix. A good selection of frequencies can improve results.
- The minimum frequency is often set as the inverse of the test duration
- The maximum frequency is often based on the Nyquist frequency.
  - The Nyquist frequency doesn’t apply for non-uniform data. We can use an approximation for a benchmark. (We can actually find patterns below Nyquist)
  - The frequency is then oversampled for safety
  - Non-uniform Nyquist model:

$$t_n = t_0 + k * p, k \epsilon \mathbb{N} \qquad f_{Ny} = \frac{1}{2p}$$

Phase

- Each frequency should be fit to both amplitude and phase. We will fit the phase first, then fit the amplitudes.
- The phase should shift the proposal wave to have a maximum at the maximum of the given data
- The phase τ is calculated for each frequency f by:

$$\tau = \frac{1}{4\pi f} \tan^{-1}\left(\frac{\sum_n \sin(4\pi f t_n)}{\sum_n \cos(4\pi f t_n)}\right)$$

Power

- The least-squares method result is the following equation:

$$P(f) = \frac{1}{2}\left(\frac{\left(\sum_n g_n \cos(2\pi f[t_n - \tau])\right)^2}{\sum_n \cos^2(2\pi f[t_n - \tau])} + \frac{\left(\sum_n g_n \sin(2\pi f[t_n - \tau])\right)^2}{\sum_n \sin^2(2\pi f[t_n - \tau])} \right)$$

Error

- The previous equations rely on some approximations, namely:

$$\sum_{n=1}^{N} \cos^2(2\pi f t_n) = \sum_{n=1}^{N} \sin^2(2\pi f t_n) = \frac{N}{2}$$

- This error is χ^2 distributed, meaning we can calculate the likelihood that a peak is due to periodic behavior.

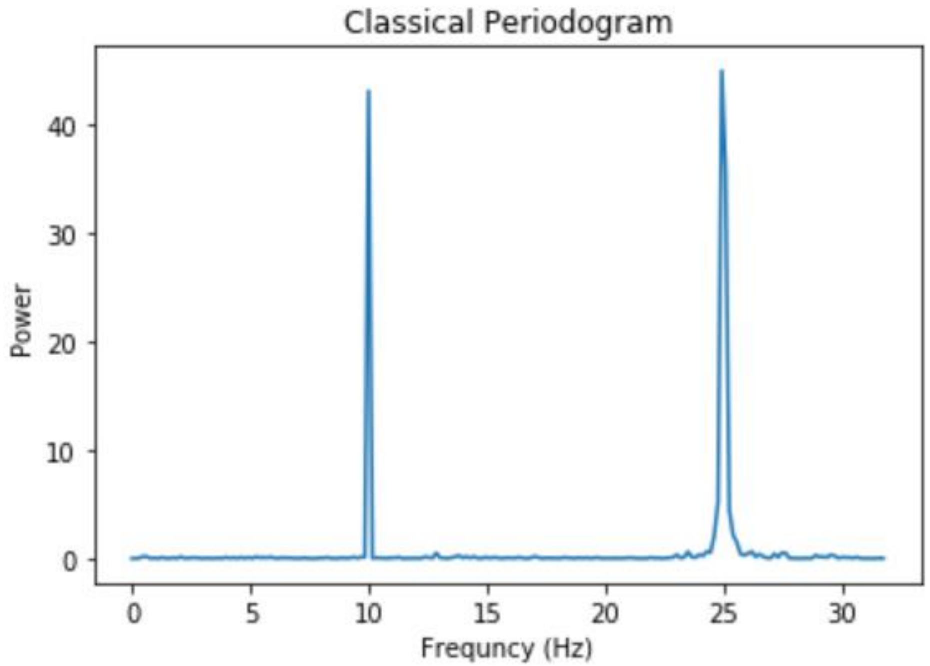
Lomb-Scargle vs. FFT

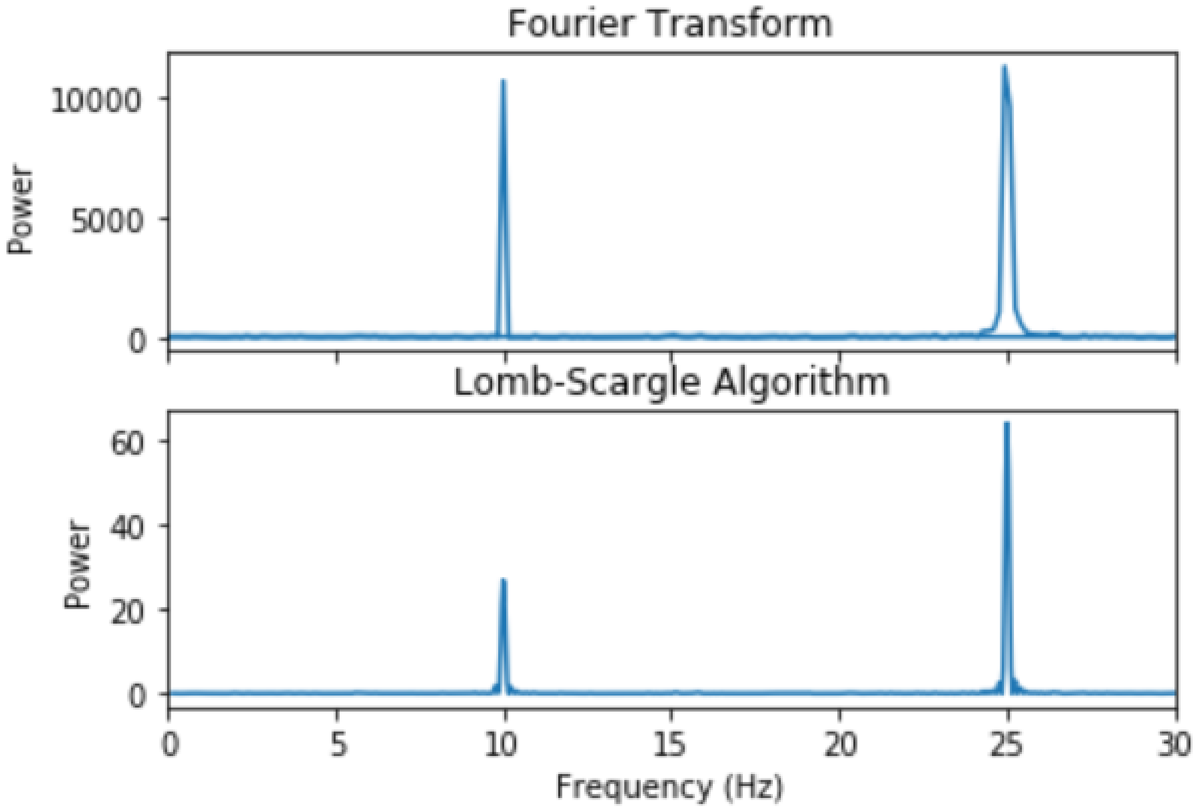
- The Lomb-Scargle algorithm is much slower than DFT, but faster than FFT, operating at order O(n) instead of $O(n^2)$ and O(nlog2(n)) respectively
- When you can use both Lomb-Scargle and FFT, FFT is actually faster. This is because Lomb-Scargle relies on oversampling.

Example: Mathematical

Simple example signal

$$\frac{1}{2}\sin(2\pi 10t) + \frac{4}{5}\sin(2\pi 25t)$$



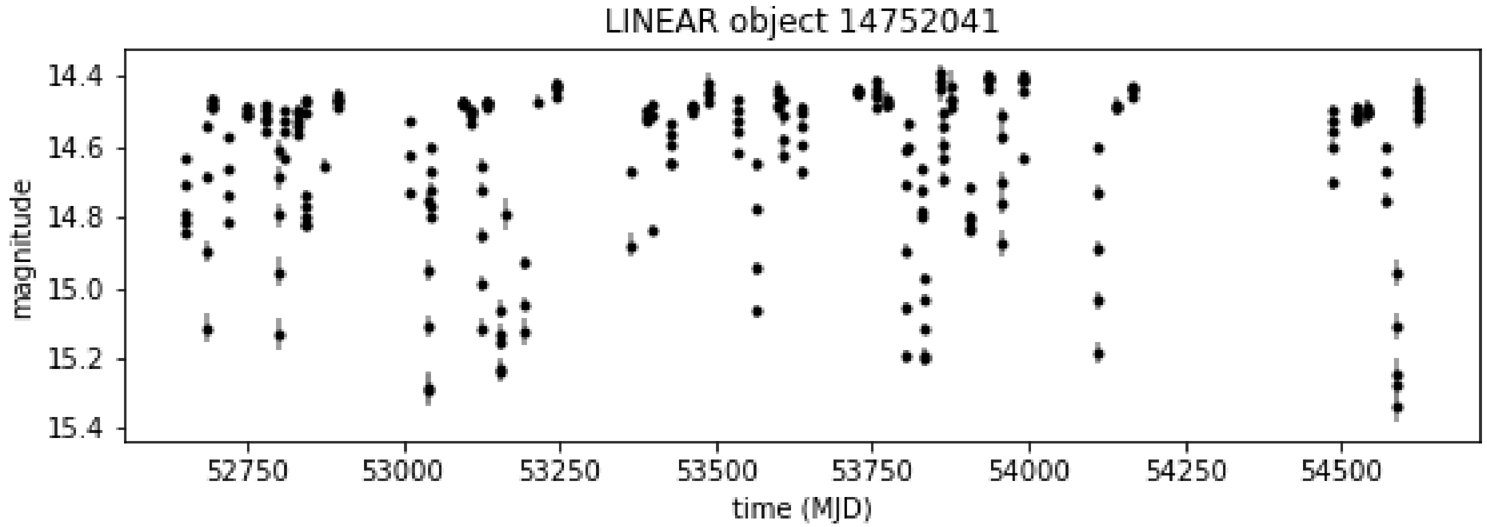


Astropy

- Astropy has developed a simple class for handling Lomb-Scargle peridograms

Pros	Cons
Object-oriented	Automatically sets the A matrix
Simple commands	Manual selection of A values far less simple
Includes 6 different algorithm implementations	Can’t handle error in time measurements
Can handle error in signal amplitude	
Built-in tests for units, significance	

Astropy Example

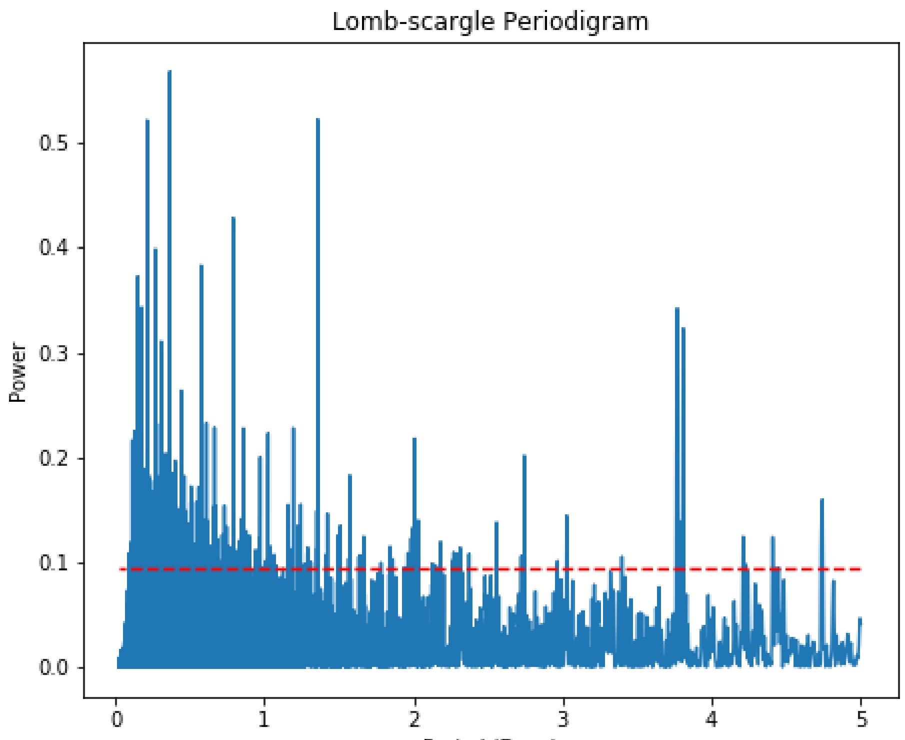


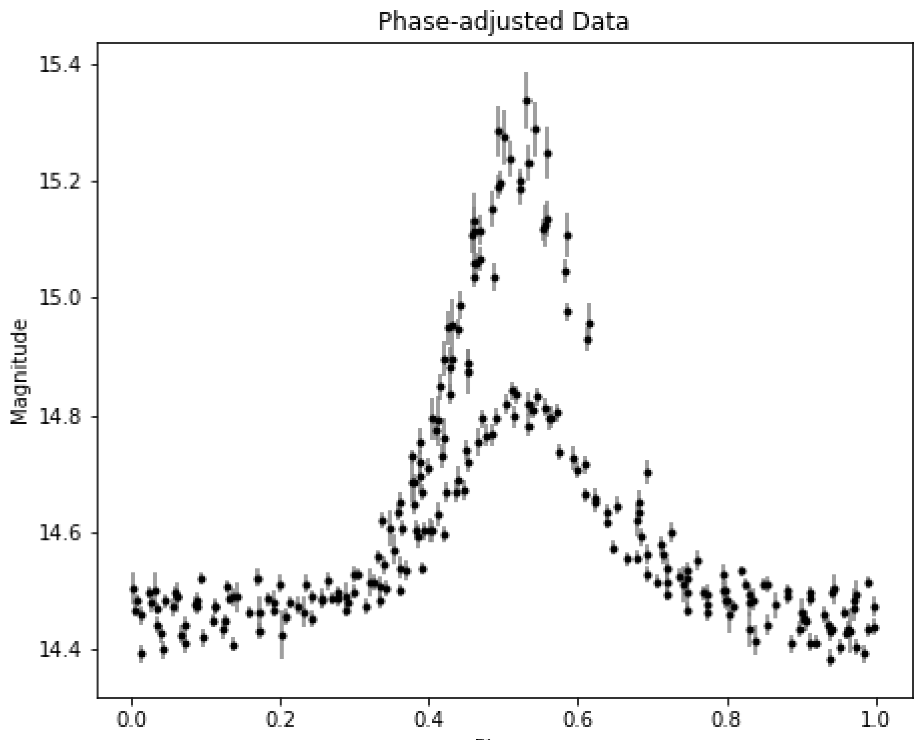
- Recall our initial example. Astropy can analyse this data set using its class notation

```
#Get the lomb-scargle parameters !These will be in frequency units!
ls = LombScargle(data.t, data.mag, data.magerr)
freq, power = ls.autopower(nyquist_factor=500,
                           minimum_frequency=0.2)
```

Astropy Example

- Once we plot the periodogram, we can see that there are many significant peaks. We can obtain a periodic signal using the most significant peak (p=1.9e-40)





Conclusion

- The Lomb-Scargle periodogram can find periodic behaviour like the FFT, but is useful in different situations:
  - It is mainly used when data points are not evenly spaced
  - It may be used to find periodic data above the Nyquist frequency
- Unlike the FFT, the Lomb-Scargle method introduces error that may be tracked
- Successful use of the method requires selecting a good frequency grid
- Python libraries such as scipy and astropy provide quick and easy to use tools to apply the method to data