Draft version March 2, 2021
Typeset using LaTeX **twocolumn** style in AASTeX63

Searching for Planet Signatures in Stellar Radial Velocity Data via the Lomb-Scargle Method

Collin Sinclair[1]

[1] *University of Colorado, Boulder*

Abstract

Inferring planet signatures from data that has been unevenly sampled in time (which is essentially all data) can be tricky. In this work, I analyze radial velocity data from three star systems using the Lomb-Scargle method and determine the periods of periodic signals present in the velocity measurements, which correspond to the orbits of exoplanets around their host stars. I determine that there are two detectable planets orbiting the star in System 1 with periods 5.399 ± 0.005 and 146 ± 4 days, respectively. Additionally, System 2 was found to host one planet signature, with a period of 3.0968 ± 0.0004 days; and System 3 contains two planet signatures with periods 3.53 ± 0.02 and 1.391 ± 0.003 days, respectively.

1. INTRODUCTION

The idea of being alone in the universe is frightening to some and comforting to others. It seems that many consider the possibility of extraterrestrial life to be something of science fiction. Astronomers, however, have a unique vantage point that allows us to peer into the vastness of space with a kind of convinced certainty that we are not alone. This is, in part, due to the now known existence of thousands worlds beyond our own.

Since the detection of the first exoplanet in 1992, scientists have confirmed the existence of more than 4,000 others. Many of these are thought to be hot-Jupiters close to their host star. More and more, we see evidence of large rocky planets - super Earths. Today, it is believed that almost every stellar system is planetary.

There are several approaches to exoplanet detection: Transit photometry (searching for shadows), direct imaging (taking pictures), gravitational microlensing (light in a gravity lens), and so on. One method is accountable for almost a fifth of all exoplanet discoveries: the radial velocity, or Doppler, method (watching for wobble).

Orbiting planets cause their host stars to "wobble" in space (both bodies are actually orbiting their combined center of mass). This changes the color of light astronomers observe from the star; as the star moves towards Earth in its wobble, its light appears bluer - that is, it has been blueshifted. When the star moves away, its light becomes redshifted. By examining how red- or blueshifted starlight is, astronomers are able to determine the *radial velocity* of the star in question. If a star's radial velocity has a periodic component, that is, it increases and decreases regularly, we consider the possibility that the star has a companion - an exoplanet.

In order to further investigate the nature of periodic changes in stellar radial velocity data, we use a tool called the Lomb-Scargle periodogram (Lomb 1976; Scargle 1982). This is a method by which astronomers are able to determine, characterize, and error-analyze the periodicity of a dataset - specifically those which are unevenly sampled in time. In this paper, I use the Lomb-Scargle periodogram to analyze radial velocity data from three stellar systems in hopes of discovering planet signatures.

2. ANALYSIS

In order to efficiently and consistently analyze the data, I have written a program that reads in and plots the data; computes the Lomb-Scargle periodogram and determines the false alarm probability of the periodogram maximum; phases the data and fits the Lomb-Scargle model; states the period of the found planet signature with uncertainty; removes the most prominent periodic signal; and gives the user the option to repeat the process on the "new"' dataset (which now excludes information about the period that was just found).

This approach allows the user detect potential orbits, confirm their existence, and check for additional planetary companions. Due to the sensitivity of phasing the data over several hundreds of periods, a periodic signal will only vanish completely when the period has been very accurately determined. Thus, when a periodic signal has been effectively removed, its parameters have been correctly established.

I explain the process outlined above at greater length in the subsections that follow.[1]

[1] System 1 corresponds to UnknownPlanetStar1, System 2 to UnknownPlanetStar2, and System 3 to Zach Star 1 Data.

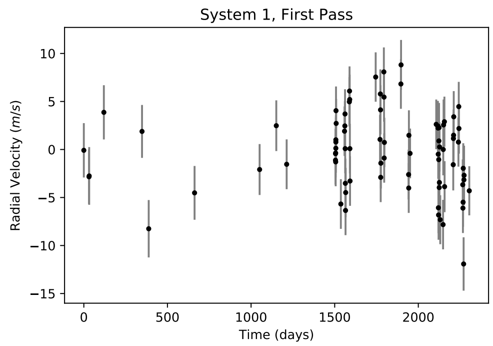

Figure 1. A plot of the stellar radial velocity data over time. Most of the observations occur in the last third of the observing epoch. The uncertainties constitute a significant fraction of the total spread of the data.

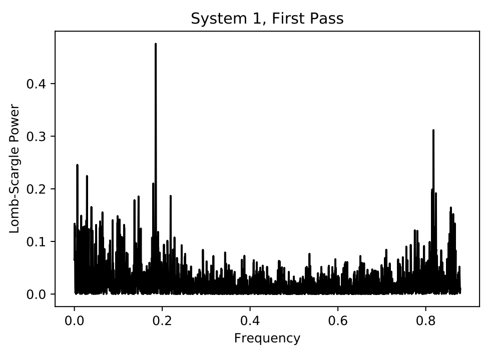

Figure 2. A Lomb-Scargle periodogram analysis of the stellar radial velocity data. There is a clear peak near 0.2 cycles per day, with another outstanding peak near 0.8.

2.1. *System 1*

We begin (as always) with a plot of the data, shown in Figure 1. Despite the non-uniformity of the observations in time, we can compute a Lomb-Scargle periodogram to determine if there is any periodic component hidden within. This process consists of checking a set of potential frequencies against the observed data and establishing the likelihood that any one of these frequencies is truly present. There has been much discussion about the selection of frequencies to test among the scientific community, specifically regarding the upper bound, or Nyquist limit.

The Nyquist limit provides a mathematical justification for the existence of an upper limit to potential frequencies in sampled data which exists in principle, and explains how to find such a limit. It is defined for time series whose entries are evenly-spaced, but has been adapted to work with data sets whose entries are not evenly spaced in time.

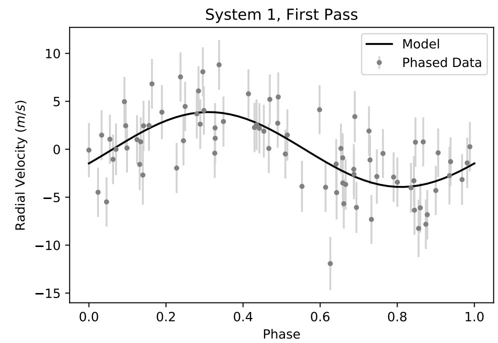

Figure 3. A plot of the phased data with a model overlaid. The data and Lomb-Scargle model appear to agree quite well.

The routine used to compute the periodograms[2] in this paper includes a heuristic approach that automatically determines the frequency grid to test based on the input data. It also considers measurement uncertainty in order to provide the most accurate report possible. Figure 2 shows the periodogram for System 1.

To determine the significance of the peaks in the periodogram, we calculate the *false alarm probability* of the highest peak via a "bootstrap" simulation: under the assumption that there is no periodic signal in the data, we will observe a peak with a Lomb-Scargle power of 0.48 or above approximately 0.00% of the time[3]. This gives a very strong indication that a periodic signal is present in the data. The period corresponding to the highest peak is $T = 5.40$ days, seemingly short to us Earth-dwellers, but quite reasonable for an exoplanet very close to its host star.

To further confirm the existence of a companion, we *phase* the data, that is, "fold" the data back over itself in time such that radial velocity measurements corresponding to the same point in a planet's orbit are shown at the same x-coordinate. If the period exists and is correct, then we should see a sinusoid. Figure 3 shows this in action. Continuing to validate the detected periodic signal, we subtract the modelled velocities from the measured velocities and perform the above tests again. If the periodic signal was correctly identified, then subtracting the model from the data should remove that period and any of its aliases entirely.

2.1.1. *Second Pass*

After having removed the initial periodic signal with period $T = 5.40$ days, we re-analyze the updated radial velocity data to check for additional planet signatures

[2] `astropy.timeseries.LombScargle`

[3] The 0.00% arises because the bootstrap simulation never recorded a peak at 0.48 or higher.

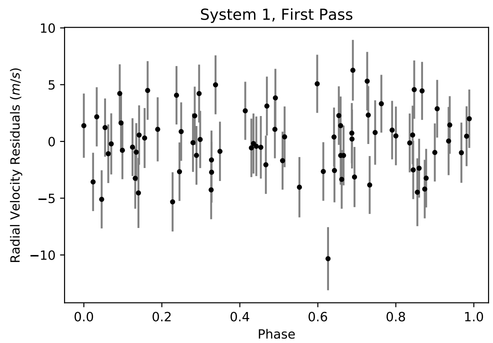

Figure 4. A plot of the residuals resulting from the subtraction of the model from the phased data. This is a great visual check that the periodic signal was correctly identified - there no longer appears to be any sort of sinusoidal pattern.

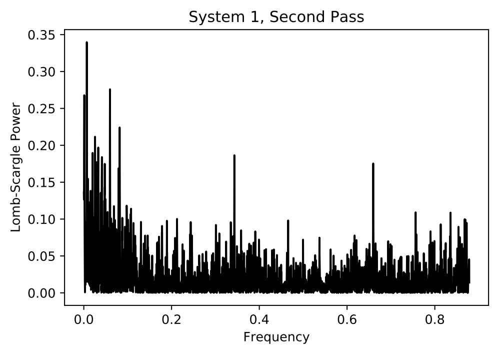

Figure 5. A Lomb-Scargle periodogram generated from stellar radial velocity data after the initial periodic signal was removed.

in much the same way as described above. Let's take a look at the Lomb-Scargle periodogram in Figure 5 on this "second pass." Notice that the peak found in Figure 2 (near frequency $f = 0.2$) has disappeared. This is an excellent sign: it shows that there was indeed a periodic signal present in the data, and that there is likely a planet in orbit around the observed star with period $T = 5.40$ Earth days.

Computing the false alarm probability for the periodogram generated in the second pass, we find again that there is an extremely low chance that we observe a peak of Lomb-Scargle power 0.34 under the assumption that there is no periodic signal in the data (nominally, 0.00% - the bootstrap simulation never saw a peak at the height of the maximum in Figure 5).

Fitting the Lomb-Scargle model to our observations, it appears that there is a periodic signal with period $T = 146.80$ days present. The model again agrees with the phased data, as shown in Figure 6.

Removing this signal from the data, we are left with the residuals shown in Figure 7, which visually lack a periodic component. To confirm this, we generate yet another periodogram based on the remaining information after the model from the second pass has been removed.

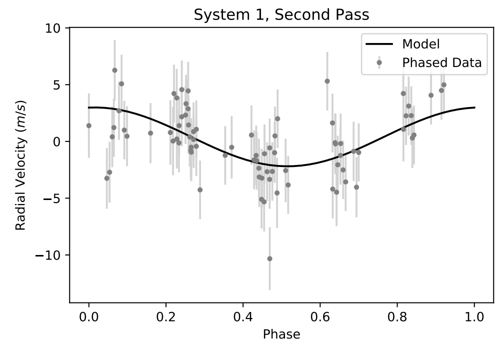

Figure 6. A plot of the phased data on the second pass of analysis with the Lomb-Scargle model overlaid.

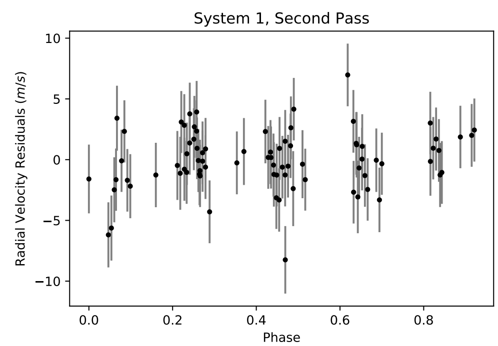

Figure 7. A plot of the residuals after the periodic signal in the second pass has been removed. Notice the lack of oscillatory patterns - a good sign!

The uncertainties in these periods are calculated from a derivation by Kovacs (1981):

$$\delta T = T \times \frac{\delta\omega}{\omega} = \frac{T}{\omega} \times \left(\frac{3\pi\sigma_N}{2N_0^{1/2}TA} \right)$$

where σ_N^2 is the variance of the residuals, N_0 is the number of data points, T is the length of the dataset, and A is the amplitude of the signal.

Applying this method of determining uncertainties to the periodic signals found in System 1, it comes out that the signal with period $T = 5.399$ days has an uncertainty of 0.005 days; likewise, the signal with period $T = 146$ days has an uncertainty of 4 days.

2.1.2. *Third Pass*

Figure 8 shows that the maximum in Figure 5 has now disappeared, validating our identification of another periodic signal in the observation data. At this point, we have confirmed the existence of two periodic signals, and

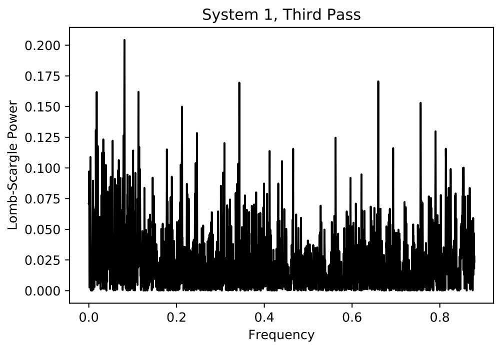

Figure 8. A Lomb-Scargle periodogram generated from stellar radial velocity data after the periodic signals from passes one and two have been removed.

with them, planetary companions of the star in question. You may notice that the periodogram in Figure 8 lacks any outstanding peaks, and that the overall range of Lomb-Scargle powers is significantly lower than in previous analyses for this system. The false alarm probability agrees with your intuition, now producing a nonzero value of 3.20%. This indicates that we may now be in the regime of superfluous analysis. Notice also that the error bars in Figure 6 span a significant range of the total amplitude of the data. Because the peak from Figure 5 has completely disappeared after the removal of that signal, we can be confident in the fact that it existed. Further analysis beyond the third pass, however, shows that the periodogram in Figure 8 (any any further passes) does not contain information that allows us to infer the existence of any further planetary orbits.

2.2. *Systems 2 and 3*

After having performed a very similar analysis on the second and third star systems in question, we find only one planet signature with a period of $T = 3.0968 \pm 0.0004$ days for the second system and two planet signatures with periods $T_1 = 3.53 \pm 0.02$ and $T_2 = 1.391 \pm 0.003$ days, respectively.

3. RESULTS

In total, five reliable planet signatures were found within three stellar radial velocity datasets: System 1 contains two planet signatures with periods 5.399 ± 0.005 and 146 ± 4 days, respectively; System 2 contains one planet signature with period 3.0968 ± 0.0004 days; and System 3 contains two planet signatures with periods 3.53 ± 0.02 and 1.391 ± 0.003 days, respectively.

4. DISCUSSION

The method outlined above opens the door to conversations about several topics regarding this kind of statistical analysis. To begin, let us talk about the normalization of the periodograms. This subject has seen much discussion in the academic community and, similarly to the Nyquist limit, lacks an exact answer. In their paper, Horne & Baliunas (1986) suggest normalizing the Lomb-Scargle periodogram by dividing each Lomb-Scargle power by the variance on the tested data (in our case, that would be the radial velocity) and continue to prescribe a formula for determining the false alarm probability given that their normalization style was used. Baluev (2008) revisits this discussion, providing four methods by which to normalize periodogram peaks.

In my analysis, I have elected to use the standard method built into the Python routine that I used to generate and analyze the data, which is normalized by the residuals of the data around the constant reference model:

$$P_{standard} = \frac{\chi^2_{ref} - \chi^2(f)}{\chi^2_{ref}}$$

The resulting power P is a dimensionless quantity that lies in the range $0 \leq P \leq 1$ (VanderPlas 2018). As mentioned in Section 2, the false alarm probabilities were calculated via a bootstrap method that does not depend on the way the periodogram was normalized.

It is also notable that in System 1 the most prominent periodic signal in the data corresponded to a planet with a very short period (by Earthly standards), whereas the second signature to be found had a period nearly 30 times greater than the first. This makes sense considering Kepler's third law, which states that the closer a planet is to its host star, the short its period will be. By Newton's law of universal gravitation, planets closer to their host star will produce a larger effect in the movement of the star than planets of the same mass further away. The reverse is true for System 3, where the planet with the larger period influenced the movement of the star to a greater extent. This occurs when the planet further away from the star has a great enough mass to offset its distance, compared to a less-massive planet that orbits closer to its host (think Sun, Earth and Jupiter!). Both scenarios are feasible.

A note on uncertainties: the errors reported with the periods of the planet signatures represent the error in the period *assuming* it exists. That is to say, the presence of such small errors does not testify to the certainty of the existence of the signal; only to the value of the period should the period be present in the first place. The confidence in the existence of a periodic signal at all comes from the false alarm probabilities.

5. CONCLUSION

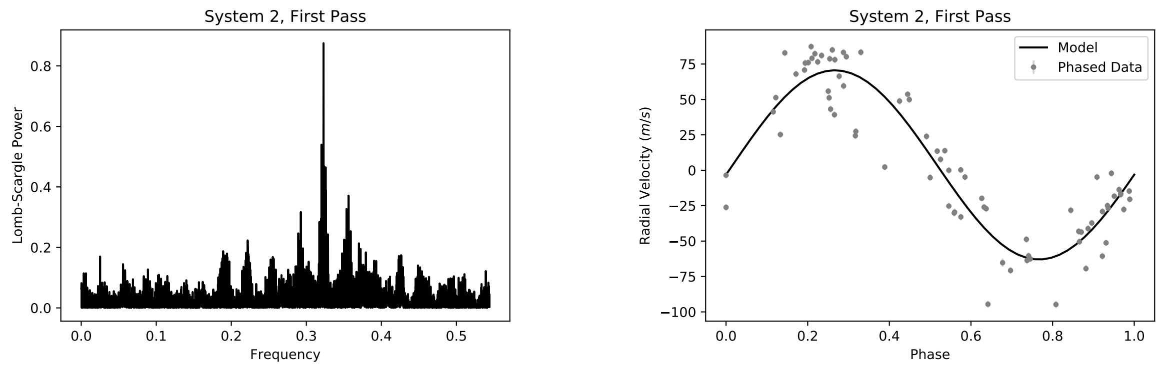

Figure 9. The Lomb-Scargle periodogram and corresponding Lomb-Scargle model for system two.

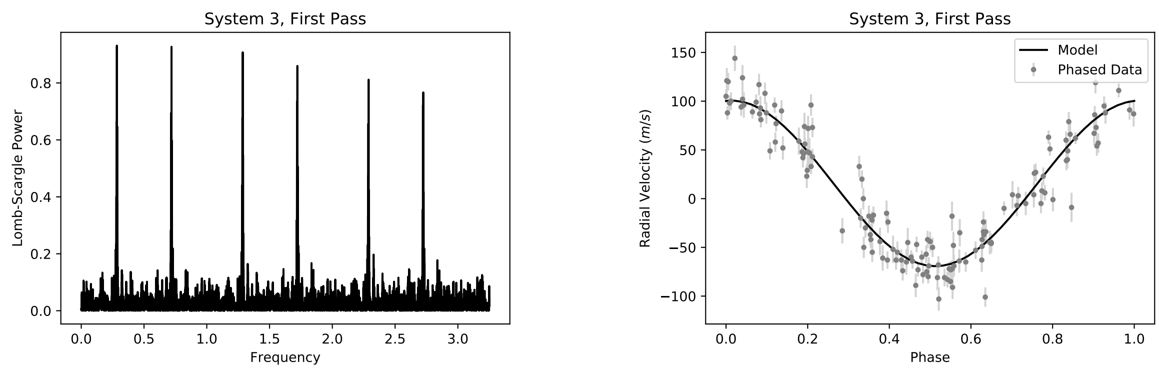

Figure 10. The Lomb-Scargle periodogram and corresponding Lomb-Scargle model for system two. Notice the aliases (several tall spikes) that occur. These vanish after subtracting the periodic signal that corresponds to their primary frequency near $f = 0.25$ - a good indication that the period has been identified.

Searching for exoplanets is perhaps one of the best ways to spend one's time. This project has been incredibly gratifying and enlightening, and I hope to do more like it soon. Speaking of the future, there exists an entirely different way to test datasets for periodic signals than what I have done in this paper: Markov Chain Monte Carlo simulations. This method is one that I hope to explore in the near future, and then compare and contrast with the results from the periodogram method.

On a personal note, I would like to express my gratitude to Dr. Doug Duncan and Mireille Fehler for a fantastic semester. This class is easily one of the best I have had the pleasure of taking, and I see myself using what it taught me well into my career.

REFERENCES

Baluev, R. V. 2008, MNRAS, 385, 1279, doi: 10.1111/j.1365-2966.2008.12689.x

Horne, J. H., & Baliunas, S. L. 1986, ApJ, 302, 757, doi: 10.1086/164037

Kovacs, G. 1981, Ap&SS, 78, 175, doi: 10.1007/BF00654032

Lomb, N. R. 1976, Ap&SS, 39, 447, doi: 10.1007/BF00648343

Scargle, J. D. 1982, ApJ, 263, 835, doi: 10.1086/160554

VanderPlas, J. T. 2018, The Astrophysical Journal Supplement Series, 236, 16, doi: 10.3847/1538-4365/aab766