Spectral Line Spectroscopy

MA342 - Numerical Analysis

Dylan Allen, Kellen Miller

February 28, 2017

Abstract

In this paper we analyzed spectral data from the galaxy ngc1275 to determine the overall strength under several peaks of interest. The data set contained variables, wavelength and relative intensity, from the black hole "Galactic Spaghetti Monster," and were the measurements of interest throughout our analysis. We fit a cubic baseline function to the data after removing our peaks, because the peaks would not have been present if there was no substance in the galaxy. Once the baseline function was determined we used numerical differentiation to find the bounds of integration for each peak. We then used the numeric integration techniques, Simpson's method and Trapezoidal rule, to solve for the total area, (strength in $\frac{W}{m^2}$), of each peak. The strengths in ascending order with regards to wavelength are as follows: peak at 3800 Å: $3.8771 \cdot 10^{12}$, 5100 Å: $1.3062 \cdot 10^{12}$, 6400 Å: $6.7341 \cdot 10^{11}$, 6700 Å: Peak 1 $2.9372 \cdot 10^{12}$, 6700 Å: Peak 2 $2.2694 \cdot 10^{12}$. We then found the margin of error by continuously integrating numerous amounts of points corresponding to each peak. To finish, we determined an approximate redshift assuming a universal linearity among all redshifts. We assume the redshift from this galaxy to be ≈ 50Å.

Contents

List of Figures

List of Tables

1 Overview

This paper contains an analysis of spectral data from the galaxy ngc1275, whose black hole is known as the "Galactic Spaghetti Monster". We were most interested in the strength of the 5 key peaks from the Excel spreadsheet data, provided by Dr. Sullivan. In Section 2, we prepared the data for further investigation through data conversion and data cleaning. In Section 3, we determined our bounds for integration by numerical differentiation and exploratory data analysis. Once we determined the bounds we created a baseline that best fit the spectral data in Section 4. We conclude the paper in Section 5 and 6, where we integrated over the primary peaks of interest, determined the margin of error corresponding to each peak, and found the approximated red-shift for each peak.

2 Converting to Frequency

There are 5 key peaks that we are interested in examining. These peaks are located at 3800 Å, 5100 Å, 6400 Å, and a double peak at 6700 Å. To analyze these peaks we began by converting wavelength to frequency using Equation 1:

$$f = \frac{c}{\lambda} \tag{1}$$

In the equation above, f is frequency, c is the speed of light ($299,792,458\frac{m}{s}$), and λ is wavelength. Because λ was measured in Angstroms; we had to convert it to meters using the conversion: 1 Å$= 10^{-10}$m. We then solved for frequency using Equation 1. We acquired a range of frequencies spanning from $4.22243 \cdot 10^{14}$ Hz to $8.21349 \cdot 10^{14}$ Hz. It is important to note that wavelength and frequency are inversely proportional, therefore the peak with the greatest wavelength will have the smallest correspondent frequency and vice-versa.

3 Determination of Bounds of Integration

Two methods were utilized in calculating the bounds by which we integrated. The first and primary method was numerical differentiation—using Matlab code to find derivatives from data sets—and the second was an analysis of the plot of intensity vs. frequency.

3.1 Numerical Differentiation Plot Analysis

The initial approach to determine the bounds of integration was to analyze a plot of our peaks' derivatives, and note where the derivative was equal to 0. We used numerical differentiation on the data where the x values corresponded to wavelength and y corresponded to relative intensity. The relationship between x and $y'(x)$ was then plotted for each peak, these relationships can be seen in Figure 1

A few characteristics to note from the plots in Figure 1. First, the red points correspond to our chosen bounds of integration. Once the wavelength bounds were determined, we recorded the frequencies and these would be

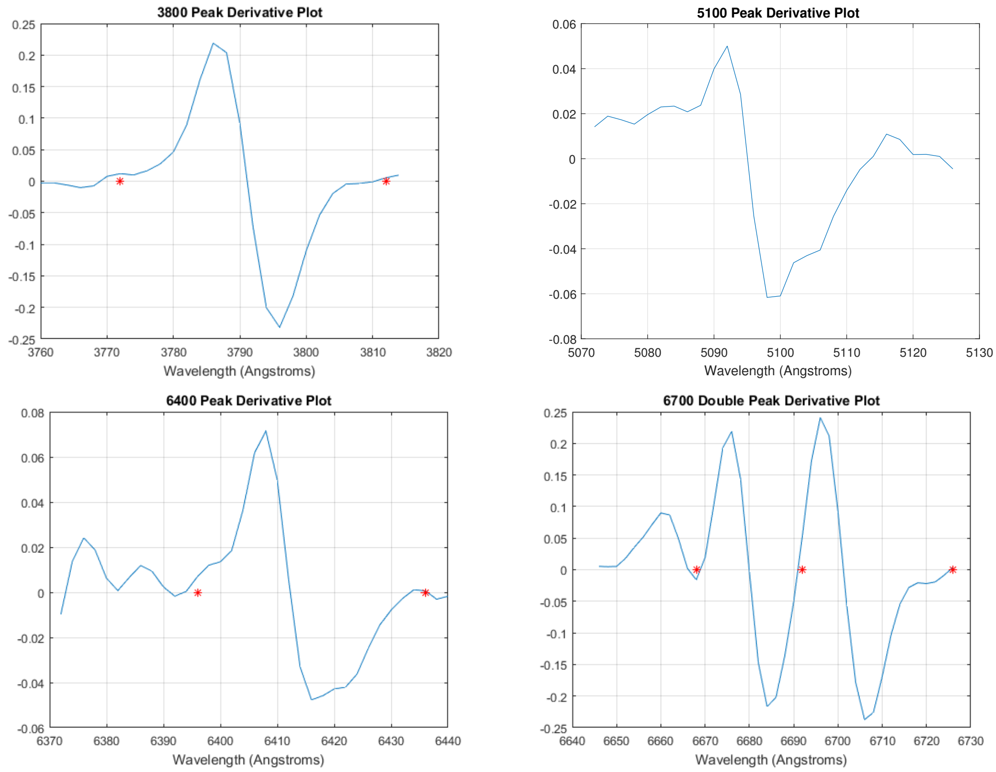

Figure 1: Derivative plots of the peaks with wavelengths at approximately 3800, 5100, 6400, and 6700

our bounds for integration. Next, the plot corresponding to the peak at 6700 Åhas an unusual characteristic; unlike the previous three peaks, the peak at 6700 Åhas two peaks. For that reason we had to analyze this peak as two separate peaks. After plotting the peak corresponding to a wavelength of 5100 Å, we noticed that there was too much noise, and the bounds needed to be approximated differently.

3.2 Intensity vs. Frequency Plot Analysis

Our second method of determining our bounds of integration was a thorough analysis of the Intensity vs. Frequency plots. With regards to the peak at 5100 Å, we had to trim off values that were near consideration as peak values, but didn't quite make the cut based purely on our judgment. This trimming left us with figure 2. Considering the jaggedness of the derivative plot, redetermining our bounds of integration by analyzing the intensity vs. frequency plot definitely helped acquire a less noisy peak.

Regarding the double peak at $\approx$ 6700 Å, we knew we could not just split them in half and use that midpoint, (where the derivative = 0), as a bound of integration. This is because spectral lines are based on the presence of certain compounds. The double peak represents the presence of 2 compounds. If one of the compounds in the double peak was not present, then there would only be one peak—in its entirety. Splitting the double-peak in half would result in limiting each of the compounds peaks, thus leading us away from the answer we were looking for. The approach we took to solving this problem was to treat each of these as individual peaks, and create a tangent line for each of the peaks' sides that were cut off when splitting them up. In Figure 3, the plot for Peak 4 has what looks like an obtrusive line sticking out the right side of it. We decided this was acceptable

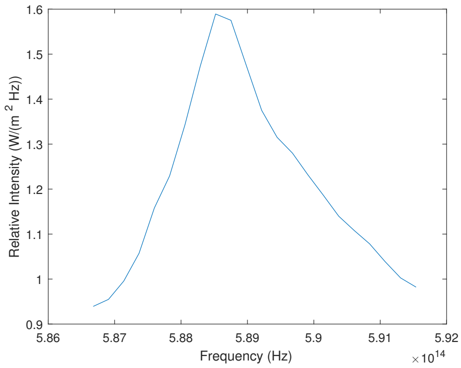

Figure 2: Plot of our new peak at 5100 Å, with noise reduced

because there was a considerable amount of noise between the left and right sides of this line. We put a line through this noise, between two points, to try and generalize this noise into something useful. The tangent line described earlier was actually implemented on the left side of this plot. On Peak 5 there is a noticeable hangover one the left. We decided to keep this based on our judgment of the determination of of integration bounds by differentiation.

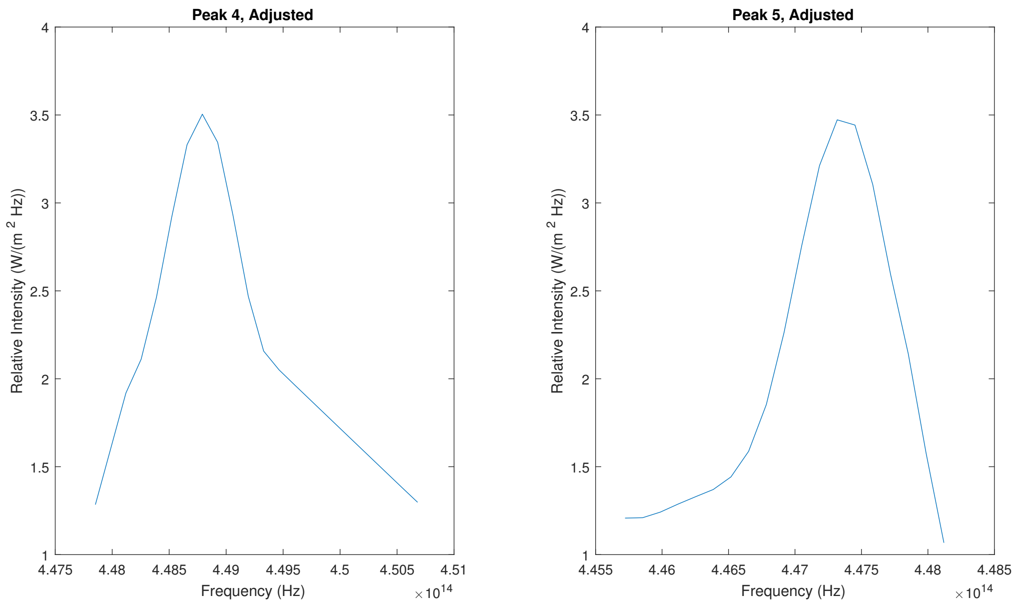

Figure 3: The plots of our final choices for the split between the double peak

4 Determining Baseline function

Our baseline function was determined by utilizing Excel's curve-fitting feature. We took the data that was supplied to us, solved for frequency, and plotted intensity vs. frequency. If we were to assume the absence of all compounds and elements in our spectral lines, there would be no peaks. So, we went through our data and deleted what we assumed to be peaks, see Figure 4. When applying a trendline to this data, the coefficients to our variable extended far past machine precision, and we assumed we would run into many problems with that. To combat this, we scaled down our frequency by dividing it by 10^{14}. This allowed us to acquire Equation 2, clearly much more manageable. Reintroducing our peaks and applying the trendline, we acquired Figure 5. Our baseline hugs the bottom of all of our peaks, so we now have an integrable area.

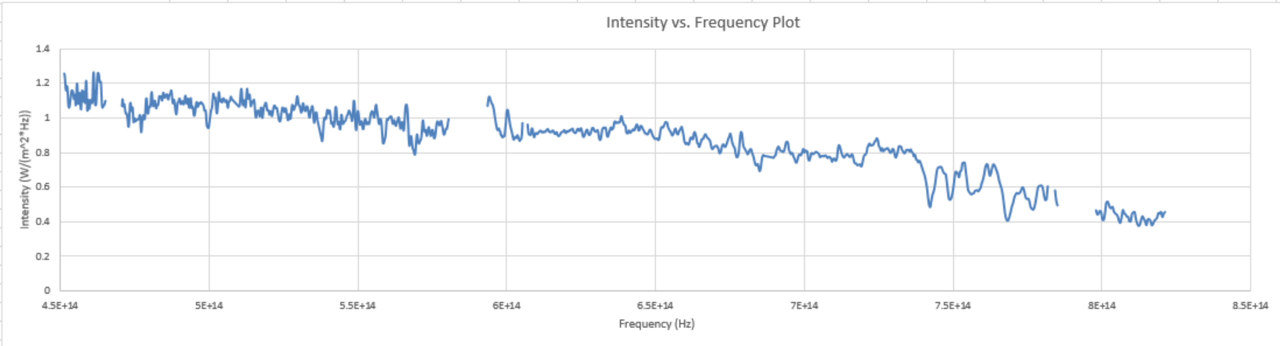

Figure 4: Plot of Intensity vs. Frequency with the absence of Peaks

$$b(x) = -0.0238x^3 + 0.409x^2 - 2.4404x + 6.0068 \tag{2}$$

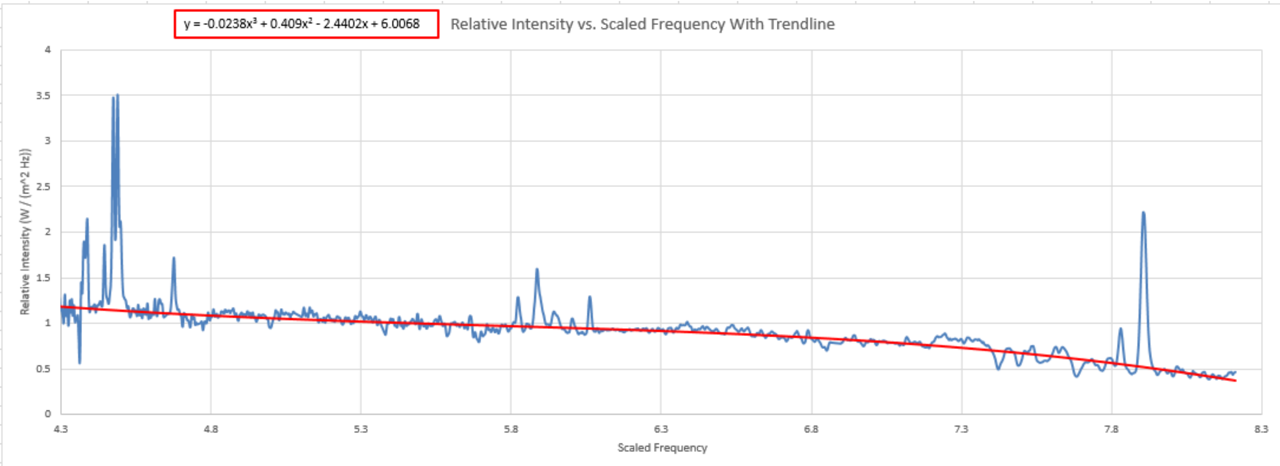

Figure 5: Plot of Intensity vs. Scaled Frequency with baseline equation

5 Numerical Integration

5.1 Process of Integration

Seeing how we scaled down our frequency in order to form an equation to handle our integration, we had to re-scale it to its appropriate magnitude. We did this by running through our Simpson's method, and then multiplying our answer by 10^{14}. To acquire the area under the peaks, we used the trapezoidal rule with unscaled data. To calculate the area between the peaks, we subtracted the integral of our Simpson's Rule integration from the trapezoidal area calculations of our peaks. This gave us the data in Table 1.

Table 1: Approximate peak strength (10^{12})

	Strength ($\frac{W}{m^2}$)
Peak 1	$3.8771 \cdot 10^{12}$
Peak 2	$1.3062 \cdot 10^{12}$
Peak 3	$6.7341 \cdot 10^{11}$
Peak 4	$2.9372 \cdot 10^{12}$
Peak 5	$2.2694 \cdot 10^{12}$

5.2 Margin of Error

In order to determine the margin of error we conjectured that the range of integration determined in Sections 3.1 and 3.2 were correct. Once this was determined we kept the step size constant for each interval and integrated from five points less than the lower bound and 5 points more than the upper bound. The output of this computation was a vector of values corresponding to the strength of each peak using different bounds of integration. The max and min of each vector of strengths was then found to get the total range of strengths given the different bounds. We continued this step for each of the five peaks. The table below is a summary of the distribution of strengths for the three peaks as well as the double peak (4 and 5).

Table 2: Descriptive statistics on a distribution of peak strength (10^{12})

	Predicted	Min	Max
Peak 1	3.8771	3.4024	3.8985
Peak 2	1.3062	1.1811	1.3189
Peak 3	0.67341	0.60603	0.67788
Peak 4	2.9372	2.5475	4.1692
Peak 5	2.2694	1.5894	3.3589

A few things should be mentioned from the table above. First, the strength values are dependent on the baseline function; therefore the table above is representation of the strength where the baseline equation (Equation 2) is used. Next, we should notice that peaks 4 and 5 have a larger range than peaks 1-3. This is because the two peaks are so close together, and when taking the strengths using the different bounds it was also computing the area under the coupled peak. For example, when taking the integral of the first peak we began at 5 points less than the smallest frequency. When we finished taking the final integral of the range the upper bound of the integral was actually part of peak 5 and increased the overall strength. Therefore, given our previous assumptions, we can suspect that the true strength of each peak lies within the range of values presented in Table 2.

6 Redshift Analysis

Looking at Figure 6 below, it appears as if the redshift occurs in a nearly linear pattern. The top spectral line image is of the sun, whereas the bottom colors are from galaxy BAS11, which is about 4 times farther away than ngc1275. There are other things one needs to worry about when considering redshift, such as gravitational interactions and universal expansion, but approximating using the image and ruler software will yield a result somewhat near the ballpark, just 4 times larger. To do this, we overlayed a visible light spectrum that had the wavelengths of the visible spectrum labeled in 50 nm steps. We then measured the number of pixels between 50 nm of the spectrum, then measured the pixels between two black spectroscopy lines. We then used the ratio between these numbers to determine the distance along the spectrum the substances shifted, but divided by 4 considering BAS11 is about 4 times farther away than ngc1275. Our ratio equation is Equation 3. Our final answer for the shift ended up being ≈ 50 Å. If we were to determine the substances in this galaxy, we would need to take this redshift into consideration, and realign our plot and baseline function, respectively. This would still only yield an approximate answer, because we did not perform any higher-level mathematics that consider universal expansion, gravitational interaction, or different rates of shift among different wavelengths.

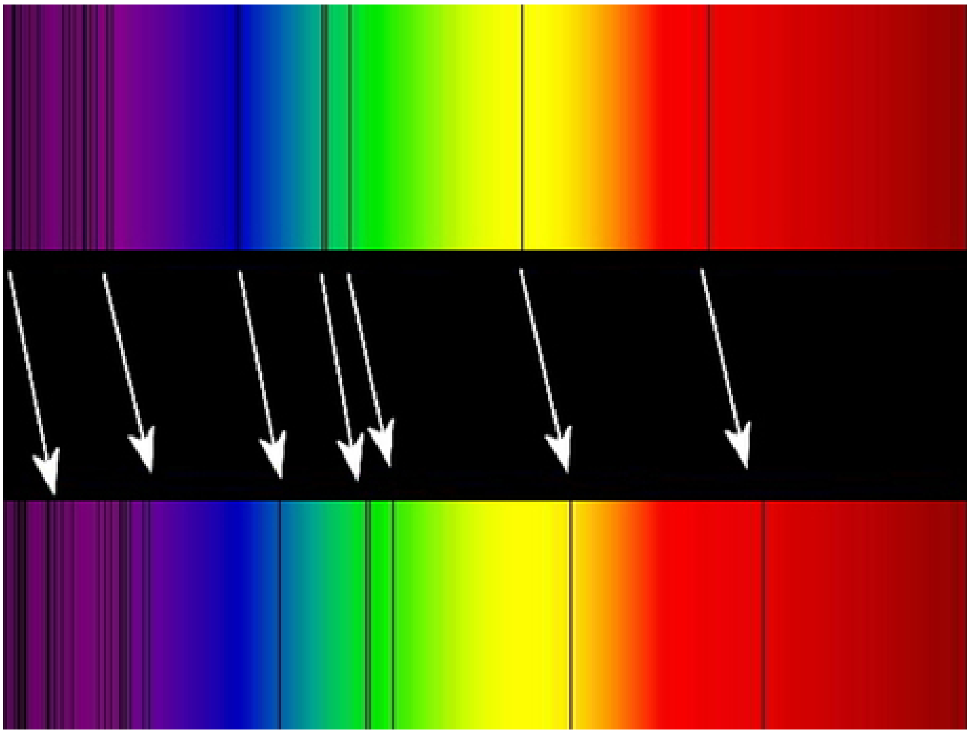

Figure 6: http://www.redshift-live.com/en/magazine/articles/Astronomy/18416-Redshift-1.html

$$\frac{61}{50} = \frac{26}{x} \tag{3}$$

Note: x is the shift for the BAS11 galaxy. We had to divide x by 4 to get approximate shift for ngc1275

7 Conclusion

In the previous sections we analyzed spectral data from the galaxy ngc1275 to determine the overall strength under several peaks of interest; three single peaks and one double peak. The data set contained variables, wavelength and relative intensity, obtained from the black hole "Galactic Spaghetti Monster. They were the measurements of interest throughout our analysis. In order to determine the strength, we first had to convert wavelength to frequency, which was later used as the bounds for integration. We took two steps before the final strength under each peak were determined: First, a cubic baseline function was derived as a lower bound of integration. Second, we found the bounds of integration using numerical differentiation techniques. From this we determined that the five major peaks ranged from $6.7341 \cdot 10^{11}$ to $3.8771 \cdot 10^{12}$. We also performed a red shift analysis, assuming a linearity among the entire spectrum.

References

[1] Mishra, Deepak. "Integration of Chromatographic Peaks." LinkedIn SlideShare. LinkedIn, 16 Sept. 2015. Web. 27 Feb. 2017.