

# Lab 2: Astronomical Spectroscopy: Detecting light with a CCD

Diana Kossakowski  
Lab Group: ~foxy~  
Melanie Archipley, Christopher Agostino

October 15, 2015

## Abstract

Using a simple, visible light spectrometer, we were able to measure a variety of spectra from different sources, and explored the noise properties of the CCD detector when changing certain factors, such as integration time or increasing brightness. The intensity follows a Poisson noise distribution, which we discover that a higher brightness reading correlate with more variance. To calibrate the wavelength from pixel, we specifically used the spectra of neon (Ne I) and a fluorescent light (Hg I), where we found the centroids of each peak from the measured spectra and fitted them against the known wavelength peaks using a least squares model, finding that a quadratic (and not linear) fit best calibrates the pixels to wavelengths. We then tested out our calibration on a green laser, red laser, and the solar spectrum to find that it was quite accurate. We also compared our coefficients of the quadratic fit to the coefficients of the Spectrometer and found very little difference.

## 1 Introduction

The purpose of this lab is to explore spectroscopy and its many uses of collecting spectra to deduce what chemicals are being absorbed or emitted by a certain source and how much of that chemical is present. Measured spectra, from stars, planets, and other objects, are taken and then compared to known emission and absorption lines of different chemicals since elements are emitted or absorbed only at certain wavelengths. In this lab, we took different types of spectra from common sources such as neon, fluorescent light (Mercury), the Sun, an incandescent lamp, a green laser, and a red laser.

The spectroscopy works by first diffracting the light coming from the optical fiber as it enters the box. After a series of lenses and mirrors, the charge-coupled device (CCD) detector collects the light (information) by the photoelectric effect, where the photons come and hit the silicon layer and release many free electrons whose voltage is recorded and then converted to a digital number that represents the number of counts of intensity as the unit of ADU (analog-to-digital units). The CCD detector is composed of a 2048 pixel-array, such that we can measure how many photons are collected at each pixel and at which pixel they

are located. Each pixel is exposed to the light for the same amount of time (known as the integration time) and then the CCD detector records how much light each pixel "captured".

Using the spectra measured by the spectrometer, we were able to measure a wavelength calibration using a least squares model with the centroids we calculated and the known peak wavelengths. With an abundant amount of data, we have also found that the light does follow a Poisson noise distribution where higher counts of intensity come with more scatter.

## 2 Equipment, Data, and Methods

For this lab, we used the USB 2000 spectrometer based on diffraction grating with a CCD detector and an optical fiber; and the Ocean Optics Control Software to visualize and to collect the spectra. The CCD detector reads out for 2048 pixels where the number of counts recorded in a pixel is proportional to the number of photoelectrons. Data was read in as pixels and ADU counts with a boxcar width of 0 and integration times of 50ms and 100ms (higher integration times in the order of a few hundred milliseconds were not taken in order to avoid dark current, which tends to amplify unnecessary noise). The analysis of the spectra of finding the centroids and the computation of the least squares fit were made using Python; and the visualizations were made using Python's matplotlib library. The data proved to show no anomalies or prominent systematic errors since the spectra of different sources closely resembled the known spectra and Figure 3 shows how no external factors were negatively affecting the data.

## 3 Analysis

To begin the lab, we collected a 1000 spectra from incandescent light at 90 volts and with an integration time of 100ms; we also collected them at 80 and 85 volts (plotted in Figure 1). We used the spectra of the one at 90 volts to figure out the gain, which is the ratio of photoelectrons to counts of the CCD; we calculated the correlation between the mean and variance to find the gain. We can't see it clearly yet given Figure 1, but there exists more scatter when the ADU reading is higher, and less scatter when the ADU reading is lower.

After exploring the correlation between the mean and the variance, we calibrated the wavelength from the data of the neon and fluorescent light spectra. We used matrix linear algebra to find the least squares fit for both a linear and quadratic fit, finding that the quadratic fit is much better. We used our calibration to see how it worked in transforming the other spectra we took.

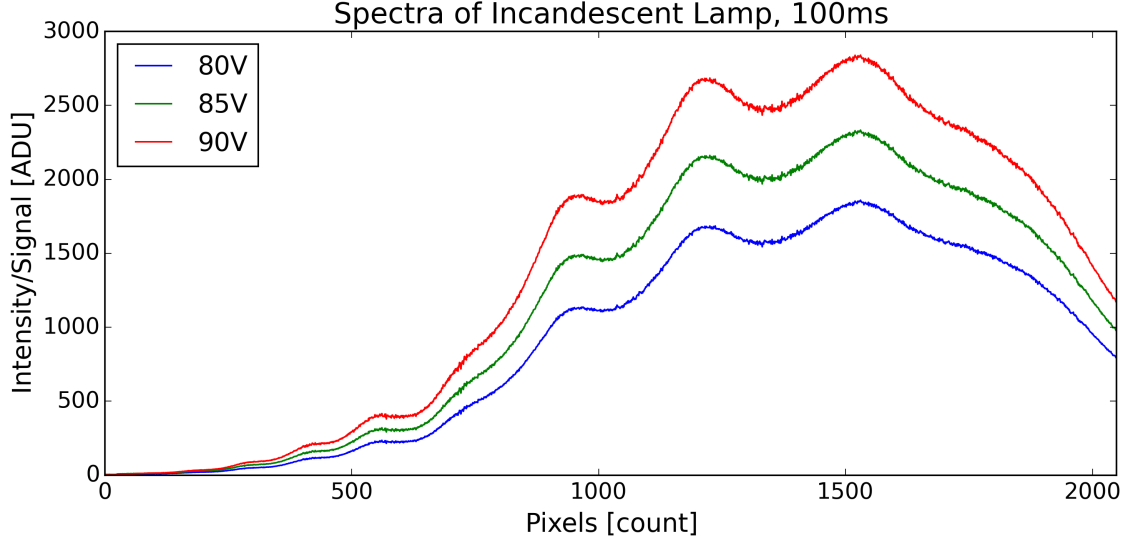


Figure 1: As expected, increasing voltage increase the power output, and thus the CCD detector captures more photons and reads out a higher ADU value

## Saturation Level, Read Noise, and Gain

We found that there are three independent factors that affect the counts of detected photons on the CCD. (1) Increasing the integration time allows for more photons to hit the detector in that given time. This works great for when the signal is a little low and needs a little boost; however, when you allow more signal to be recorded, then more noise will be recorded as well. Therefore choosing 100 milliseconds enables a good balance of the ratio of signal to noise, where it amplifies the signal well enough without allowing the dark current to become apparent. (2) Increasing the voltage of the light naturally increases the counts counted by the CCD since voltage is proportional to power (intensity) out. (3) There exists a "dark current", where electrons are captured by the CCD due to the thermal energy of the CCD and these electrons increase the intensity counts unnecessarily and are independent of the actual intensity given by the source. To fix the dark current, we took spectra trying to cover up all light; it was the "control" spectra of the experiment.

### Saturation Level

The intensity of the data can sometimes reach the saturation level, where no more counts of intensity can be recorded. We were able to find that this level occurs at 4095 ADU from the spectra we took of the red laser (we accidentally allowed too much of the light from the red laser to enter the optical fiber/spectrometer). We concluded that the value of 4095 ADU must be due to the fact that the spectrometer contains a 12-bit ADC, since  $2^{12} - 1$  is 4095.

### Read Noise and Gain

To explain briefly, the CCD detector is made up by combining PN diodes such that the PN junctions allow very little to no current flow. This acts as a capacitor with a dielectric region,

where we as the user measure the voltage by the equation 3 (after subbing in equation 2 into equation 1,

$$V = \frac{Q}{C} \quad (1) \quad Q = Ne \quad (2) \quad V = \frac{Ne}{C} \quad (3)$$

where  $C$  is the capacitance and  $Q$  is the charge that the photon puts on the capacitor due to the Photoelectric Effect, which is proportional to the number,  $N$ , of electrons of charge,  $e$ , that come in. The voltage is measured by an analog digital converter (ADC) and we call these measured units *analog to digital units* (ADU). We know that the ADU increases proportionally to the number of electrons that come in. Therefore, the ADU that we measure is eventually given by,

$$ADU = g \frac{Ne}{C} + ADU_0 \quad (4)$$

where  $g$  is the "gain" and  $ADU_0$  is the bias. Equation 4 leaves out the effect of dark current given that only short exposure times (less than a few hundred milliseconds) are used. Next, we want to find the variance (fluctuation) in ADU,  $\sigma_{ADU}^2$ , which we can do using error propagation. Equation 5 and 6 are the functions, and equation 7 and 8 are the error propagations. We can set the covariance to 0 when dealing with  $ADU$  since  $N$  and  $ADU_0$  are independently interacting.

$$f = f(u, v, w) \quad (5) \quad \sigma_f^2 = \left( \frac{\partial f}{\partial u} \right)^2 \sigma_u^2 + \left( \frac{\partial f}{\partial v} \right)^2 \sigma_v^2 + \dots + \text{covariance} \quad (7)$$

$$ADU = ADU(N, ADU_0) \quad (6) \quad \sigma_{ADU}^2 = \left( \frac{\partial ADU}{\partial N} \right)^2 \sigma_N^2 + \left( \frac{\partial ADU}{\partial ADU_0} \right)^2 \sigma_{ADU_0}^2 \quad (8)$$

where  $\sigma_{ADU_0}$  is called the "read noise" which describes how good the voltmeter is; and  $\sigma_N^2$  is the variance of the photons coming in, which follows a Poisson distribution, where we know from Lab 1 that the variance equals the mean,  $\sigma^2 = \mu$ . In this scenario, we can say that,  $\mu = N$ , since  $N$  gives us the best approximation for  $\mu$ , and thus,  $\sigma_N^2 = N$ . With that in mind, we can then write  $N$  in terms of equation (4) and then plug that back into equation (8) to get the final noise model,

$$\sigma_{ADU}^2 = \left( \frac{eq}{C} \right) (ADU - ADU_0) + \sigma_{ADU_0}^2 \quad (9)$$

where  $\sigma_{ADU}^2$  describes the fluctuation in signal and we redefine "gain" as  $\left( \frac{eq}{C} \right)$ . "Gain" tells us how many electrons we need to get to obtain one "count" (ADU), as shown in Figure 2, we need 10 electrons to increase one count of intensity. We can see that there is more fluctuation as brightness (voltage) increases, which follows the Poisson distribution that more information will bring more scatter (more noise). It is also important to note that we have only made this one measurement of the gain. If we were to keep taking more and

more independently measured spectra and then measure the gain, we would see that the gain would become more precise as it converges to its mean (reference back to Lab 1).

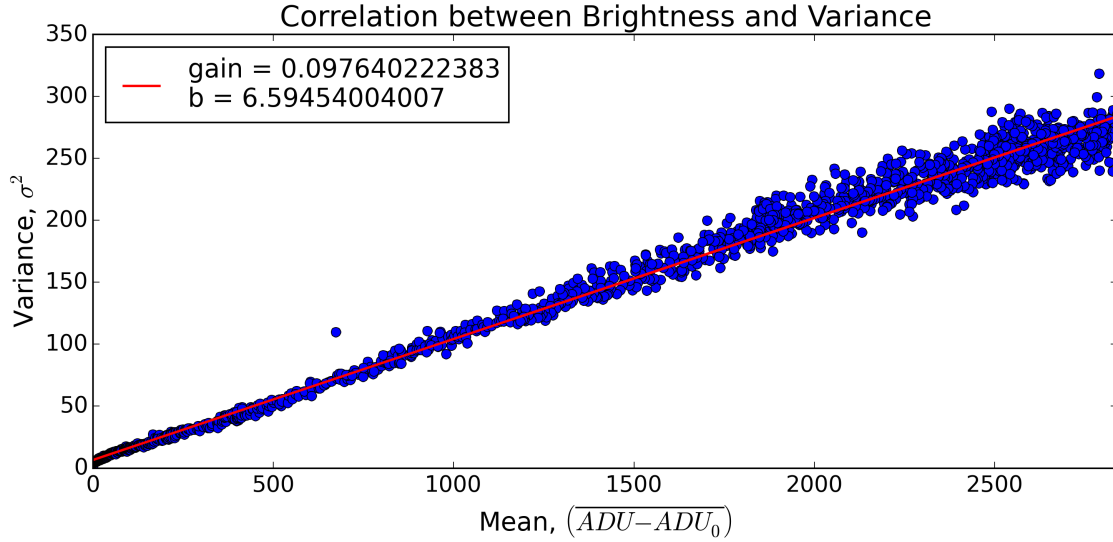


Figure 2: The means and variances from each pixel from 1000 (dark subtracted) spectra are plotted against each other and show a Poisson noise, where there exists a correlation of how increasing brightness increases the variance as well. *Note: there is one outlier (around pixel 600) that may suggest this is a malfunctioning pixel. Removing it from the fit would make a negligible difference.*

The purpose of Figure 3 is to ensure that no external factors are affecting the variance of the signal. If we were to find that this plot dips down because some light had been blocked, then the variance would inappropriately be higher and not accurately represent the actual variance. From what we have, we can conclude that no discrepancies are present.

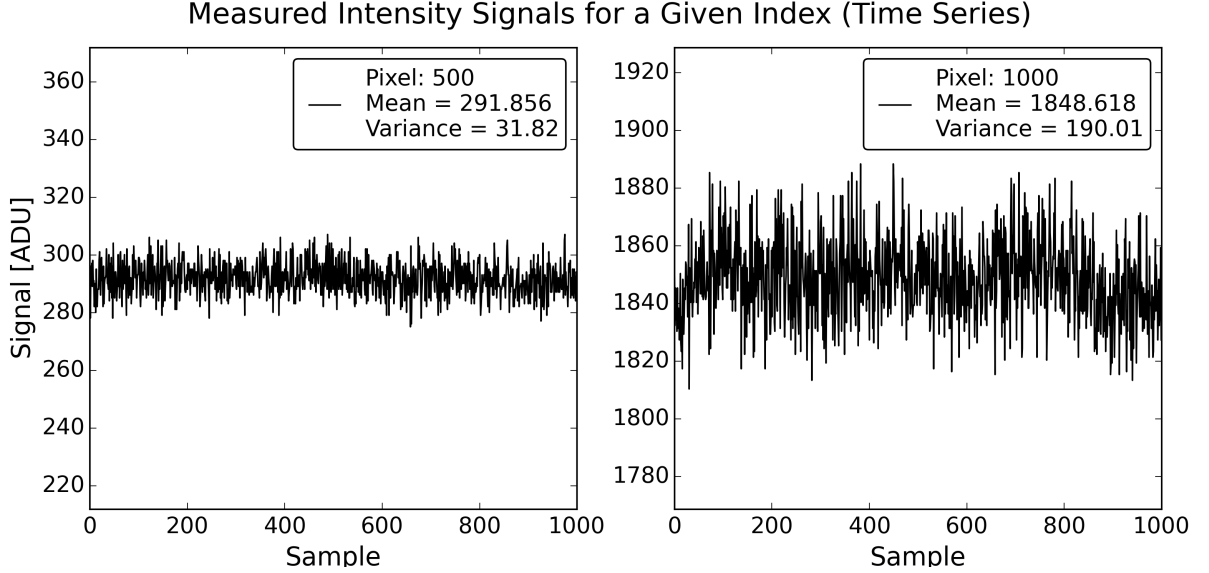


Figure 3: 1000 spectra (dark subtracted) from an incandescent bulb at 90V were measured every 100ms. (a) and (b) look at pixel 500 and 1000, respectfully. The x-axis represents the  $n^{th}$  measured spectra and the y-axis represents the measured signal for the  $n^{th}$  spectra. The y-axis range is the same for both plots to explicitly show the difference in variance. A full correlation between brightness and variance for all 2048 pixels is shown in Figure 2.

## Wavelength Calibration

To obtain the wavelength calibration, we took the spectra of both neon light (Ne I) and fluorescent light (Hg I) since they have distinct peaks at higher wavelengths and lower wavelengths, respectfully. Using both spectra enables us to achieve an accurate calibration.

## Centroids

Using the raw data given as pixels [count] versus intensity [ADU] by the Spectrometer, we plotted the results of 100 averaged, dark subtracted spectra for each element in Figure 4. To find the peaks, we collected all the data points that were the local maxima, only considered the ones greater than a certain intensity ( $> 100ADU$ ), and then kept the ones that were the maxima within 10 pixels (we assumed each peak is roughly 10 pixels). This method worked great for the spectrum of Neon due to the fact that Neon has very prominently distinct and well-defined peaks, whereas at higher pixel counts for Mercury ( $> 1450$ ), the spectrum peaks are not as clear and may come within 10 pixels of each other, so we decided to not include them since Neon already covers the higher pixel range. Next, we calculated the center of "mass", known as the centroid, of each peak assuming that each peak is roughly 10 pixels and using the center of mass formula,

$$x_{peak} = \left(\frac{1}{2}\right) (x_i + x_{width}) \quad (10)$$

$$x_{centroid} = \left(\frac{\sum_i^{width} I_i x_i}{\sum_i^{width} I_i}\right) \quad (11)$$

The beautiful spectra of neon and the fluorescent light, along with the centroids plotted as vertical lines, are found in Figure 4. We calculated the centroid error using these formulae,

$$\sigma_{\langle x \rangle}^2 = \frac{s^2}{F}, \text{ where } s^2 = \left( \frac{\sum_i (x - \langle x \rangle)^2 I_i}{\sum_i I_i} \right), \quad F = \sum_i I_i \quad (12)$$

where  $s^2$  represents the rms width of the peak given by the variance. We found that these errors were on the magnitude of  $10^{-4}$ , and therefore, the error bar for a calculated centroid would not show up in Figure 5.

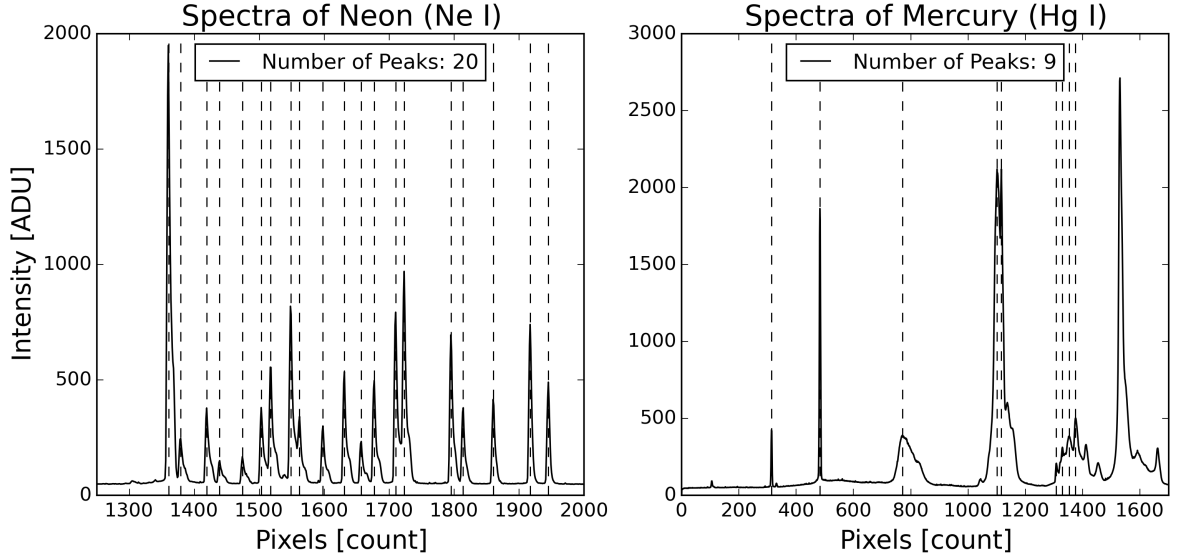


Figure 4: Both spectra were averaged over 100 spectra at 100ms time intervals. The Neon spectrum (a) dominates in the higher pixel range and the Mercury spectrum (b) covers the lower pixel range. Combining the peak information from both spectra allows for a more accurate calibration.

## Least-Squares Fitting

Now that we have our centroids spanning the visible spectrum, we can compare them to the known wavelengths of the peaks of Ne I and Hg I, and then we can fit the data. When looking for the best fit to the data, we are trying to minimize the distance from each point and the fit,

$$d_{linear} = y_i - (mx_i + c) \quad (13) \quad d_{quad} = y_i - (a_2x_i^2 + a_1x_i + a_0) \quad (14)$$

$$\chi^2 = \sum_i (y_i - (mx_i + c))^2 \quad (15) \quad \frac{\partial \chi^2}{\partial m} = 0; \quad \frac{\partial \chi^2}{\partial x} = 0; \quad (16)$$

where the index  $i$  spans through all points and  $d_{linear}$  and  $d_{quad}$  correspond to a linear and quadratic fit, respectfully. The  $x_i$  variable is known perfectly as it is the measured

quantity, which is represented by the centroids we have computed in Figure 4. The  $y_i$  holds all the error and is represented by the actual wavelengths of Ne I and Hg I. To find the best coefficients for the fit, we have to solve Equation 16, which becomes very easy and doable if we express it as a system of linear equations, or a matrix as below,

$$\begin{pmatrix} N & \sum x & \sum x^2 & \dots & \sum x^m \\ \sum x & \sum x^2 & \sum x^3 & \dots & \sum x^{m+1} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x^m & \sum x^{m+1} & \sum x^{m+2} & \dots & \sum x^{2m} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum y \\ \sum y * x \\ \vdots \\ \sum y * x^{m-1} \end{pmatrix} \quad (17)$$

The next graph (Figure 5) shows how the data of pixels versus wavelength is fit. We can see how the neon spectrum covers the higher pixels and wavelengths and the mercury spectrum covers the lower pixels and wavelengths. The coefficients for the quadratic fit are shown in Table 1.

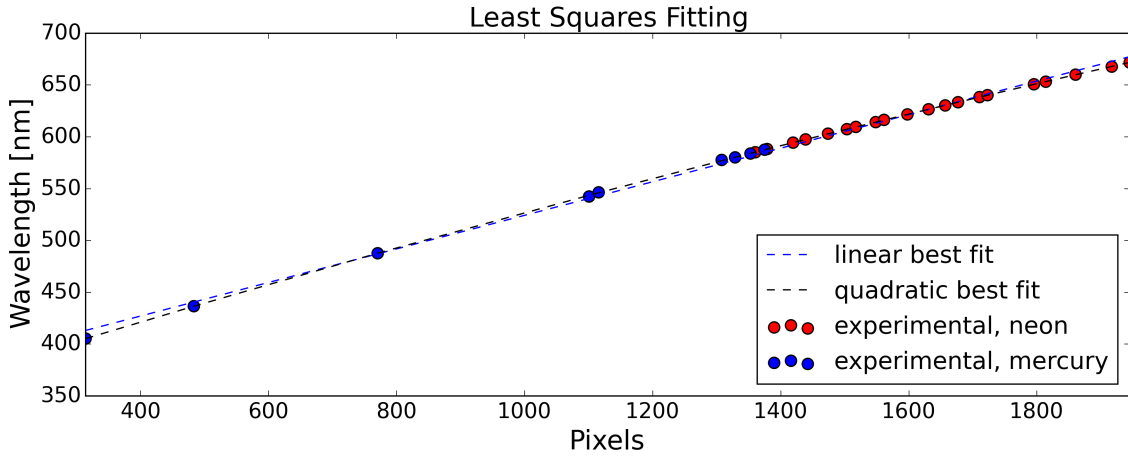


Figure 5: Here are both fits plotted, and as we can see, they look practically identical.

Polynomial Coefficient	Value
$a_0$	344.9778
$a_1$	.1966
$a_2$	$-1.4716 \times 10^{-5}$

Table 1: Quadratic fit for the data in Figure 4

When we plot the linear and quadratic fits in Figure 5, we notice how both fits look practically identical and we are under the impression that the linear fit does a good enough job; however, this is not the case as Figure 6 shows. The residuals explicitly show how the quadratic fit is better; if a pattern emerges from the residuals, as the linear fit residuals do, then it is not an adequate fit.



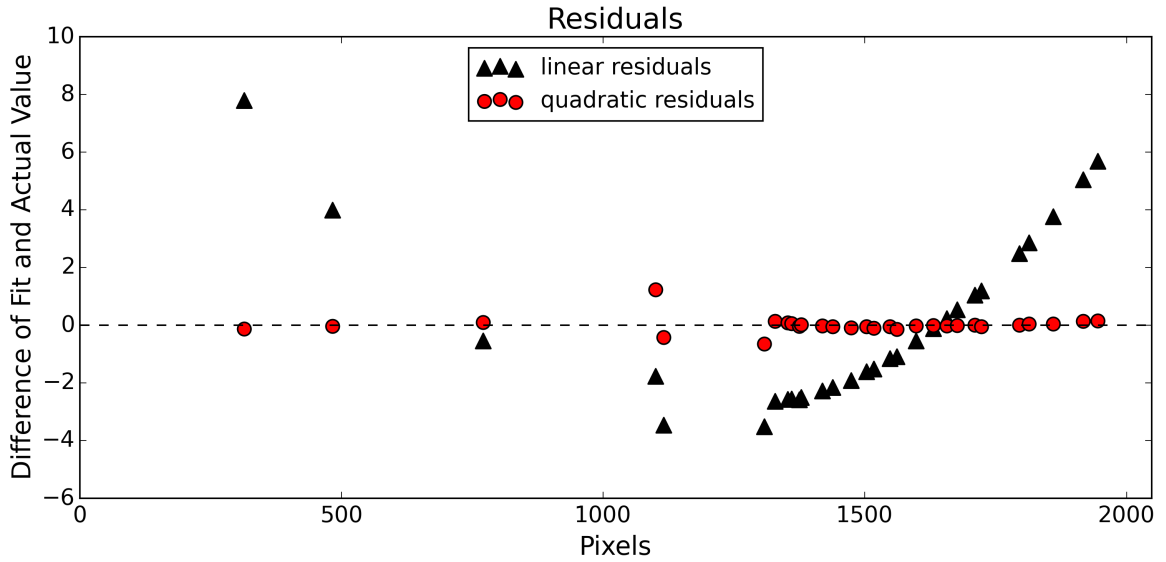


Figure 6: The linear residuals (triangles) show a clear quadratic pattern, whereas the quadratic residuals (circles) appear to be completely scattered around zero indicating a good fit.

## Results

I decided to see how well our calibration works by applying it to the spectra of a green laser and of a red laser! These spectra have also been averaged over 100 individual spectra at 100ms time intervals. As a side note, the green laser is not as high intensity because of systematic errors (it was difficult to hold the laser still for  $100ms * 100$  seconds). Figure 8 shows how the percent error from the true wavelengths of high-end laser pointers were only a fraction of a percent off. The error for the red laser was a bit more than the error for the green laser, which may be accounted by the fact that the red laser has a higher intensity, and therefore more scatter.

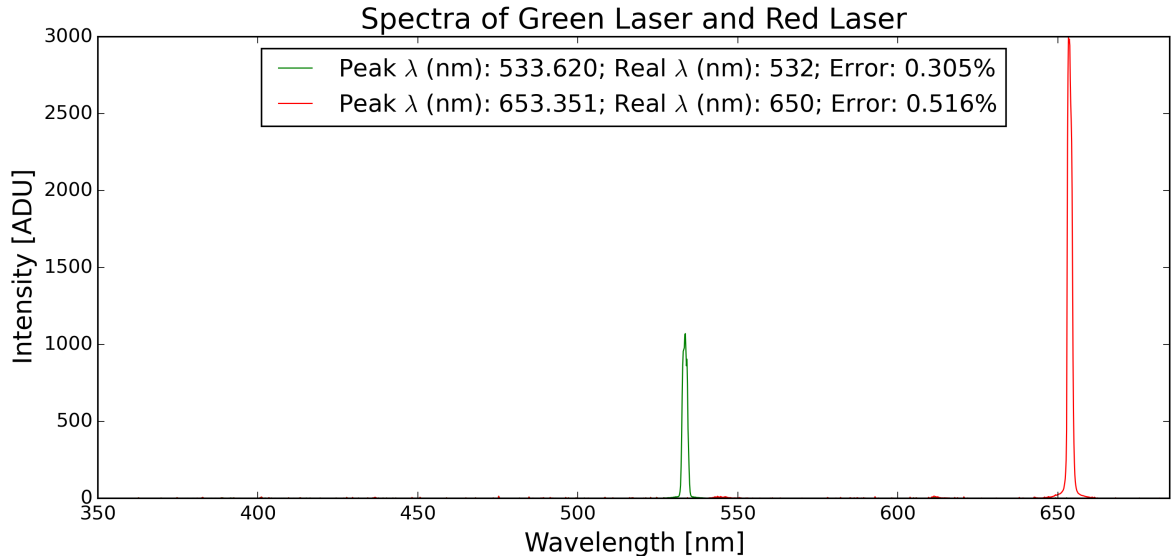


Figure 7: Spectra of a green laser and a red laser were averaged over 100 individual spectra at 100ms time intervals.

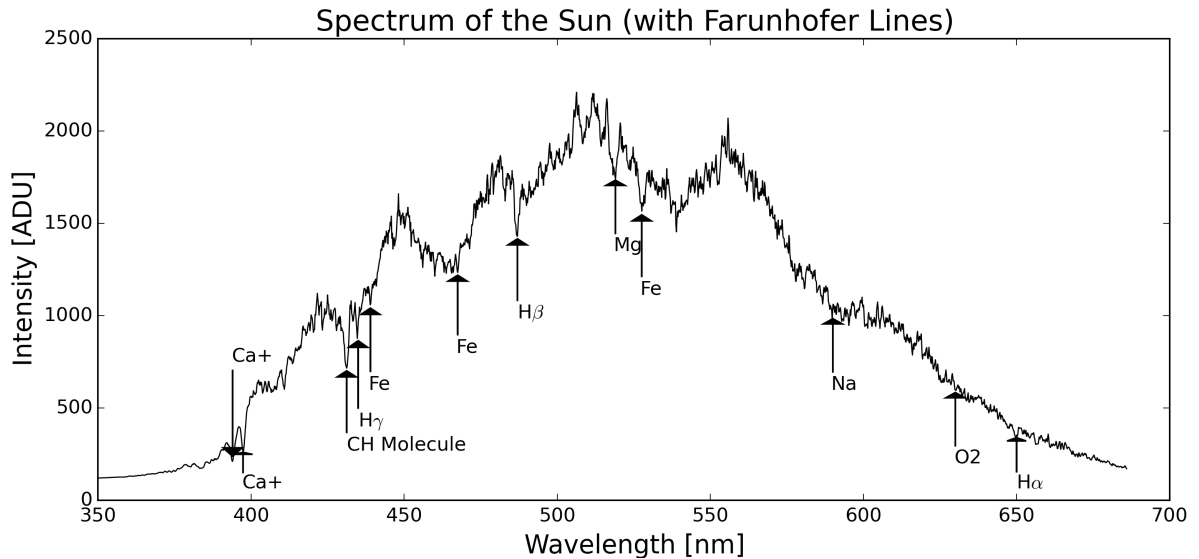


Figure 8: The solar spectrum and the absorption lines (Fraunhofer Lines). The spectrum was taken when the Sun was at high noon on a cloudless day through a glass window.

## Conclusion

From this lab, I learned the usefulness of spectroscopy for spectra observation, how a spectrometer works by diffracting light for the CCD detector to count the number of photoelectrons per pixel, and how error propagation plays a role in collecting data. The ultimate goal was to calibrate the data from the spectra, which our group did very successfully as the calibrated spectra show very good precision to the real wavelengths. Overall, I am quite pleased with the results we obtained.