Problem Set 6

Evan Arch (https://github.com/evan-nyu/phys-ga2000)

October 16, 2024

Problem 1

Note: throughout this problem set I am only working with the first 10 galaxies. Working with more causes my terminal to crash. The fits data file contains the spectrum of numerous galaxies. Normalized plots of the flux versus wavelength for the first 4 galaxies in the data set are shown in figure 1.

To preform principle component analysis, we need to find the covariant matrix C given by $\mathbf{C} = \mathbf{R^T} \cdot \mathbf{R}$ where $\mathbf{R}$ is the matrix containing the fluxes of each galaxy. We can find the eigenvectors of $\mathbf{C}$ using numpy's built in methods. The five eigenvectors with the largest corresponding eigenvalue are plotted in figure 2. Finding the eigenvectors with this method took 18.3 seconds.

We can also find the eigenvectors using SVD on $\mathbf{R}$. In this case, the eigenvectors will be contained within $\mathbf{V}$. A plot of the eigenvectors found using this method is shown in figure 3. These eigenvectors are equivalent to those found using numpy's build in method up to a scalar. This method took 0.14 seconds.

The condition number of $\mathbf{R}$, 35.4, is much lower (and thus better), than the condition number of $\mathbf{C}$, 66732165000000.0. Thus, it seems that SVD is the better technique to use in most situations.

An approximation of the any spectru can be found by decomposing the spectrum into five constants multiplied by the 5 largest eigenvectors. This is shown for the first galaxy in figure 4. Plots of the constants, c_1 vs c_0 and c_2 vs c_0 are shown in figure 5 and figure 6 respectively.

The squared residuals for approximations using various numbers of eigenvectors are shown in figure 7.

The RMS residual for $N = 20$ eigenvectors was found to be $2.2116232 \times 10^{-10}$ which indicates that this is a very good approximation of the original function.

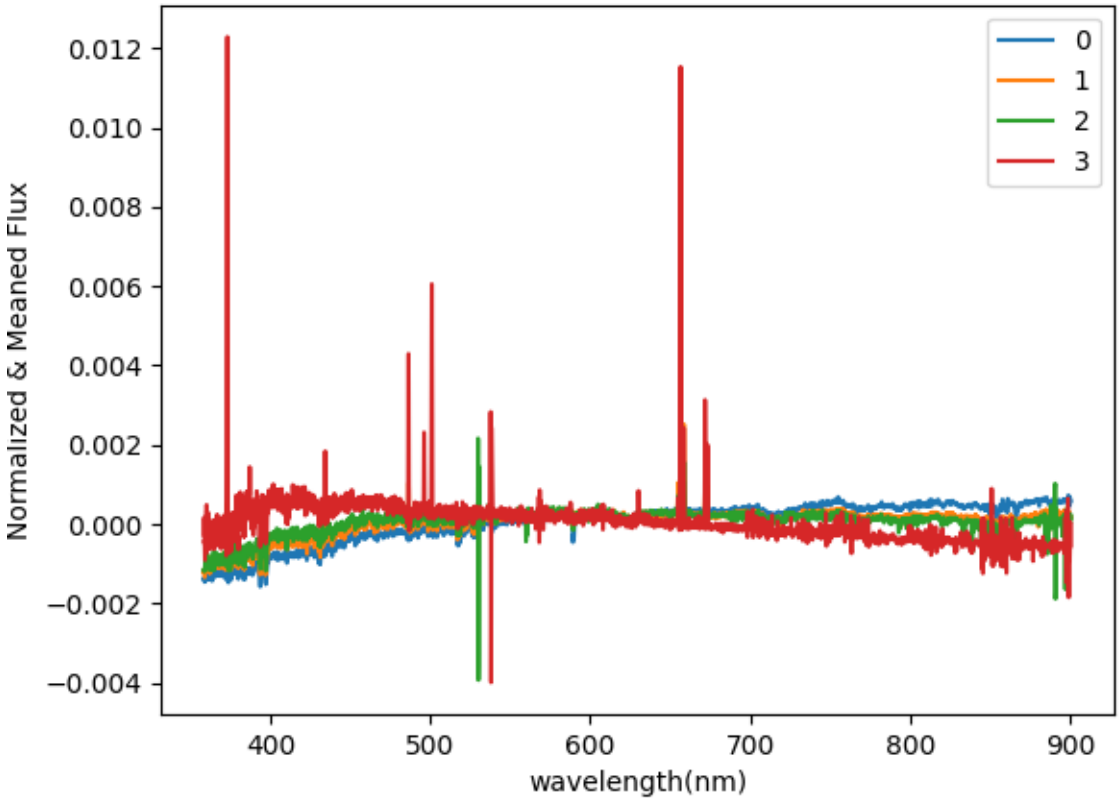

Figure 1: Normalized and meaned spectra of the first 5 galaxies in the data set. Note the spike at 650 nm corresponding to the lowest energy Balmer transition in hydrogen.

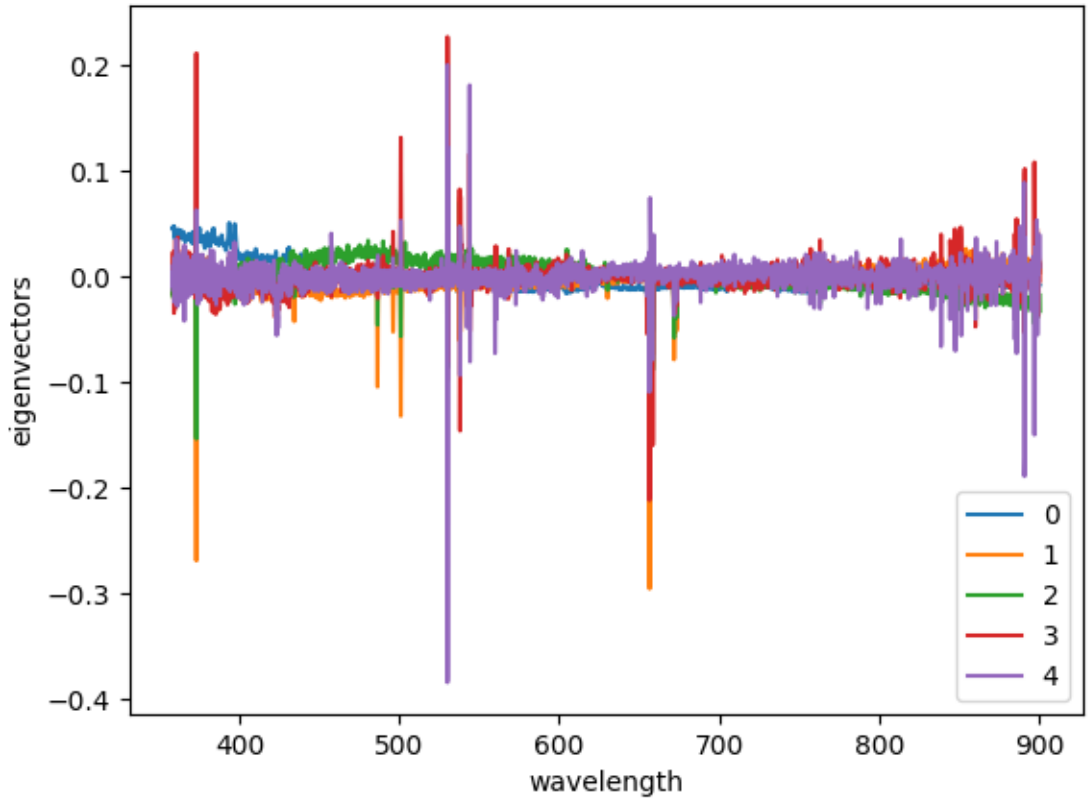

Figure 2: Plots of the 5 largest eigenvectors found using numpy built in linalg.eig method plotted as a function of wavelength.

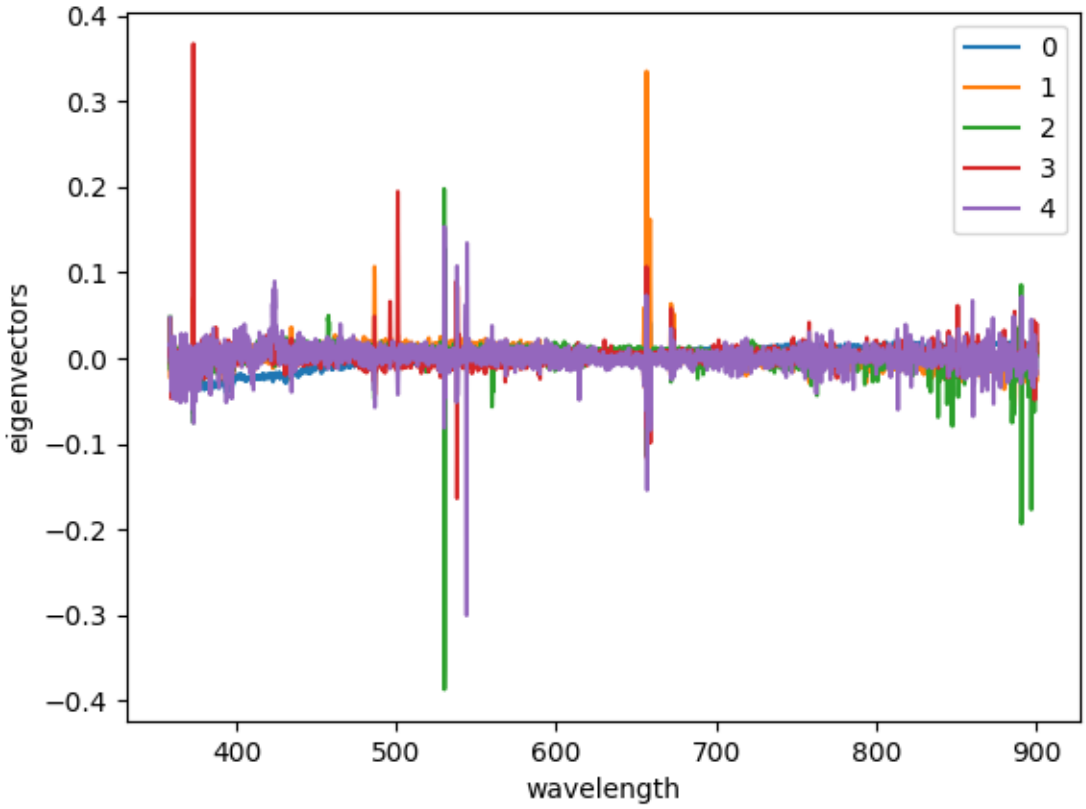

Figure 3: Eigenvectors found using SVD versus wavelength.

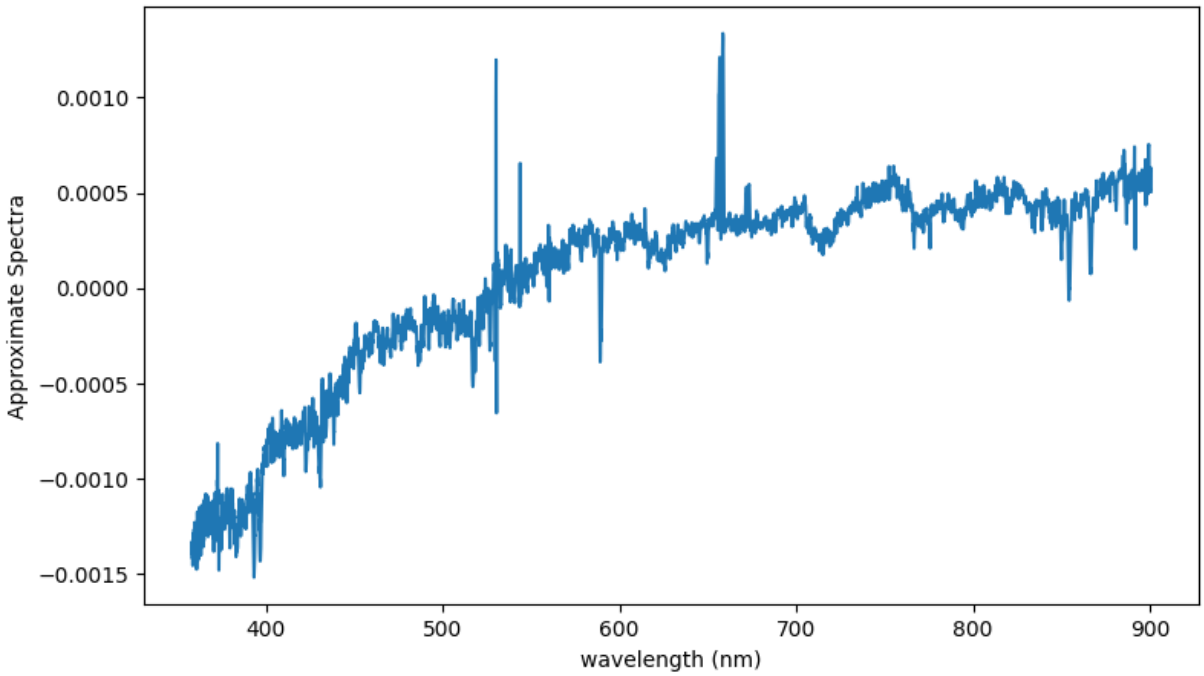

Figure 4: Spectrum of galaxy 1 from 5 largest eigenvectors

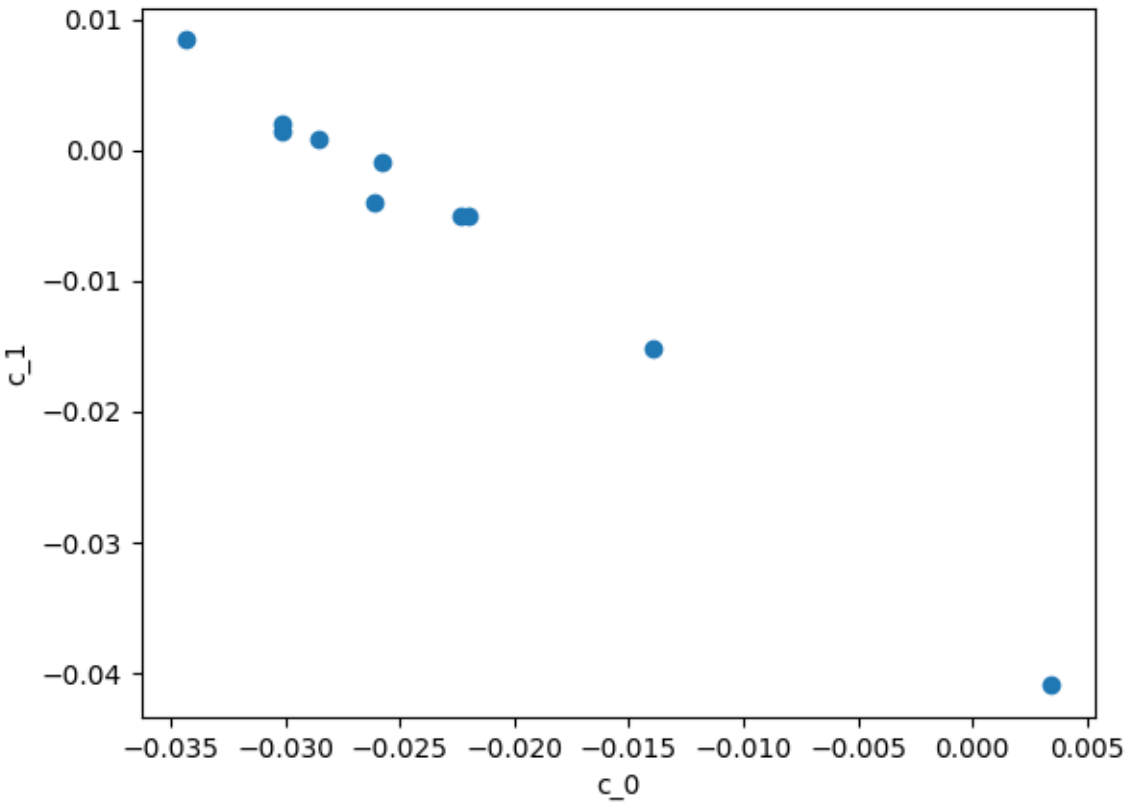

Figure 5: c_1 vs c_0

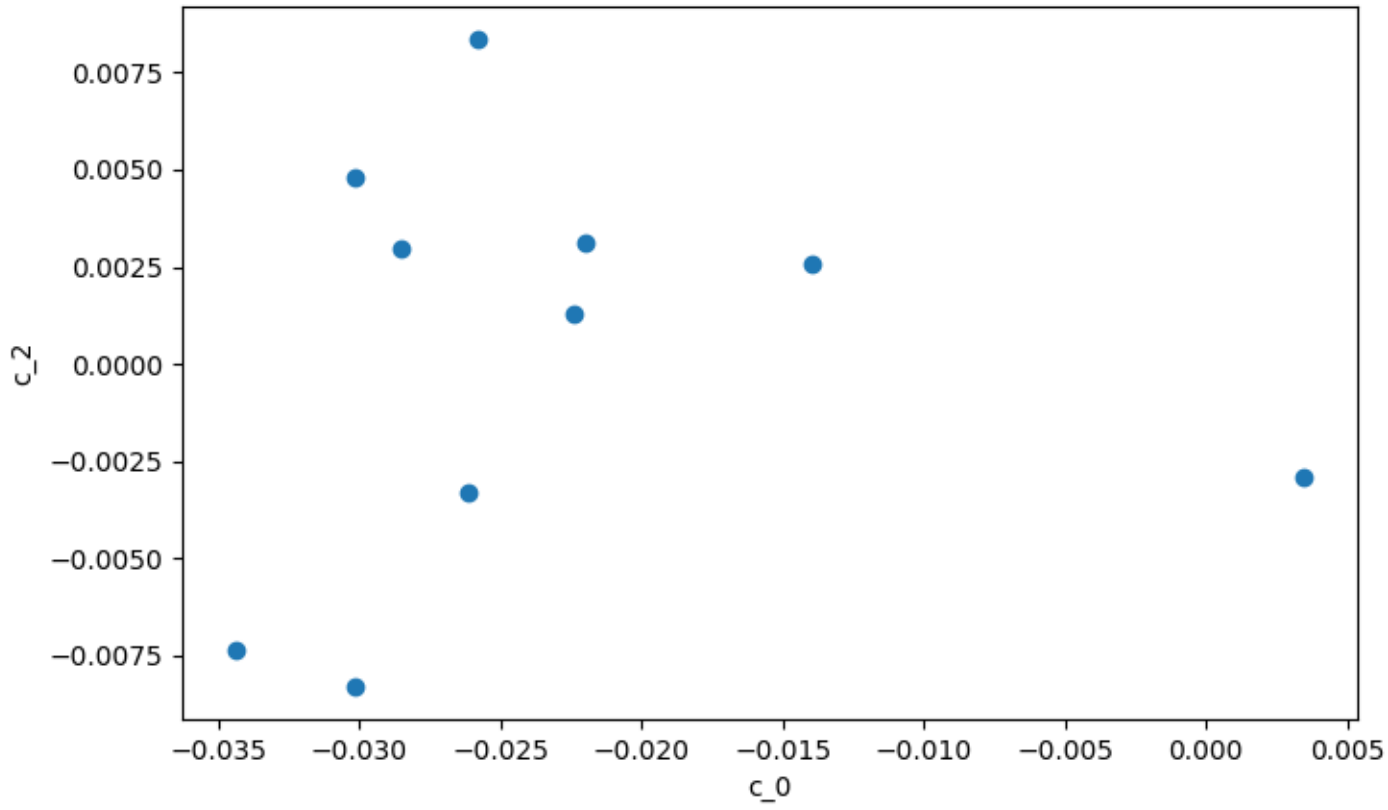

Figure 6: c_2 vs c_0

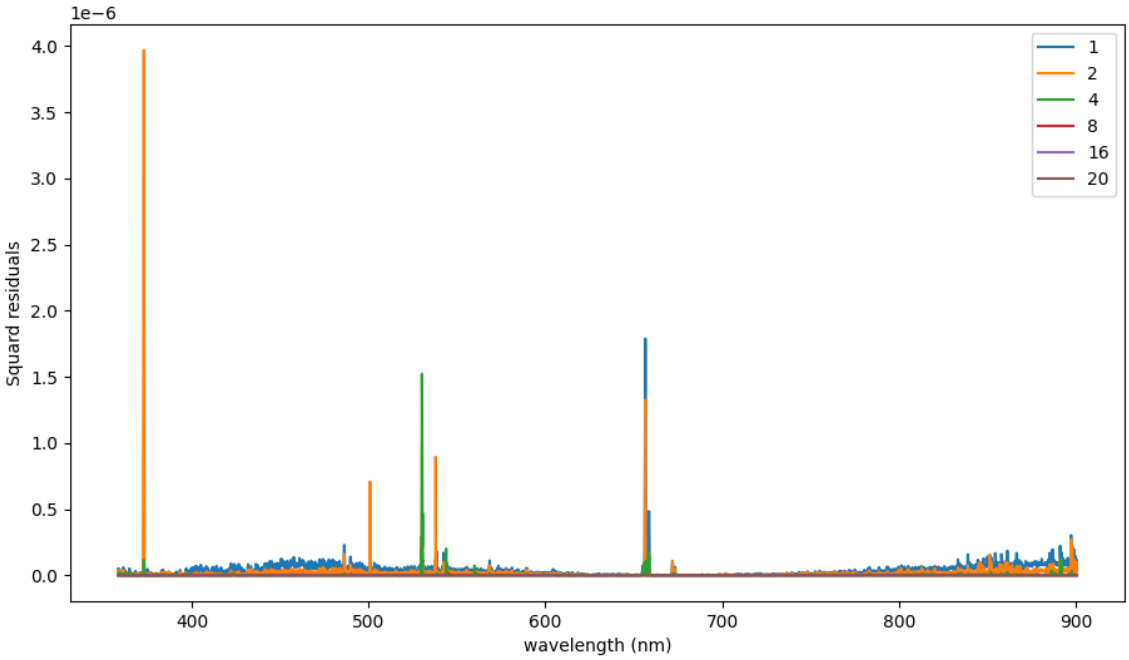

Figure 7: Squared residuals as a function of wavelength for $N = 1, 2, 4, 8, 16, 20$ eigenvectors.