HW5

Robert Xi

Nov. 2023

1 Problem 1

1.1 (a)

Here I plotted 10 galaxies' data. From the plot, I see few peaks that occurred around same wavelengths for all different galaxies. Particularly, there are few peaks that are close to each other, which makes me think about that hydrogen atoms are responsible for those lines peaks, and the fact that at around 650 nm there's a peak, which corresponds to the Balmer line for Hydrogen atoms.

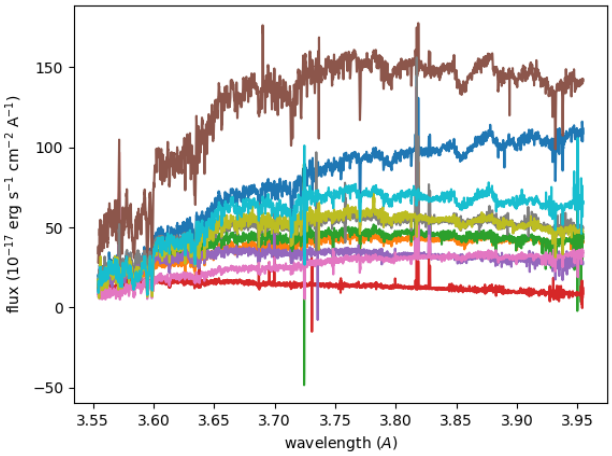

Figure 1: flux against wavelength for 10 galaxies

1.2 (b)

Here I plotted the normalized data for 10 galaxies in figure 2.

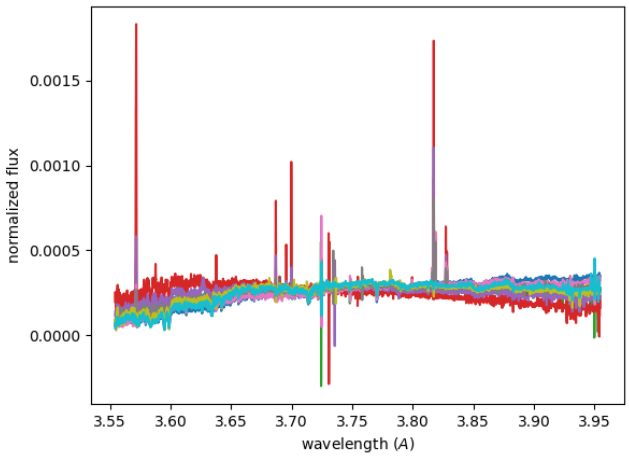

Figure 2: normalized flux against wavelength for 10 galaxies

1.3 (c)

Here I plotted the normalized data with 0 mean for 10 galaxies in figure 3. This data is obtained by subtracting the mean from the data obtained in part b.

1.4 (d)

In this part and next part, we determine the eigenvectors of the corresponding eigenvalue distributions through two ways. In this part, we create the covariance matrix and extract its eigenvectors. In the next part, we extract them directly from the residuals matrix using SVD. Here I demonstrate the first approach. Figure 4 plot the first 5 eigenvectors calculated using covariance matrix. One can see that the data seems to be much cleaner, which is what we wish to see.

1.5 (e)

Continued from last part, here I present the data obtained using SVD method in figure 5. The plot shows that using SVD and covariance matrix methods yield similar results in terms of the wavelength of the peaks. However, they are not exactly the same, each peaks has slightly different amplitudes. I also compared the computational time for each method, and here are the computational time comparison: Computation time using covariance matrix method is 29.8s, whereas the computation time using SVD method is 49.9s, which indicates that using covariance matrix method maybe more efficient. I also include a direct comparison between the calculated eigenvalues using SVD and covariance method in figure 6. One can see that the eigenvalues calculated by both methods are exactly the same.

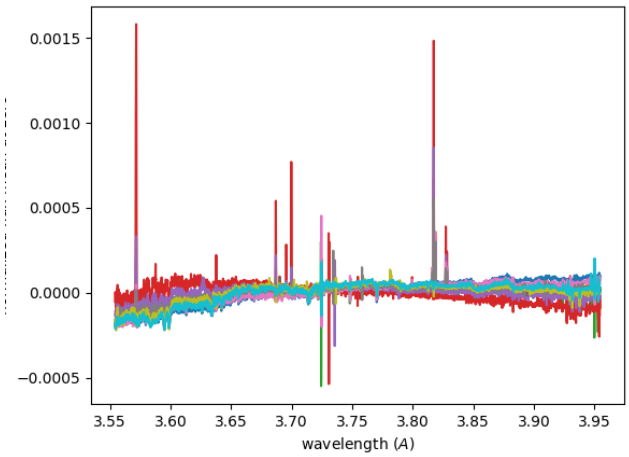

Figure 3: normalized flux with 0 mean against wavelength for 10 galaxies

1.6 (f)

Here's a comparison between the condition numbers for C and R: Condition Number of C is 17410599000.0; Condition Number of R is 6561841.5. One can see that the condition number for C is number larger than that of R and they are both large, which indicates that they are close to being singular. R has smaller value, which may indicate that R's data is more reliable.

1.7 g

Figure 7 is an approximate spectra based on keeping only the first five coefficients. One can see that they are quite similar except a slight deviation the high frequency information.

1.8 h

Below I plotted c_0 vs c_1 and c_0 vs c_2 in figure 8 and 9.

1.9 i

Figure 10 presents the squared fractional residuals between the spectra and the reconstituted for N_c range from 1 to 20. One can see that as N_c increases, the squared fractional residuals drastically decreases. At $N_c = 20$, the value is approaching 0, suggested that the difference between actual data and approximated data is almost the same.

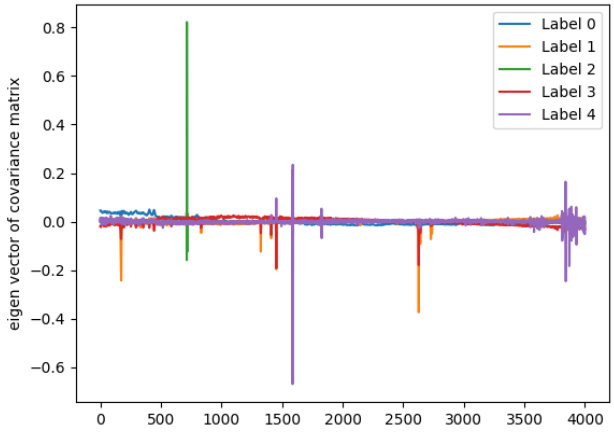

Figure 4: 5 eigenvectors obtained using covariance matrix and extract its eigenvectors

2 Github

username: robertXi6 link: https://github.com/robertXi6/phys-ua210

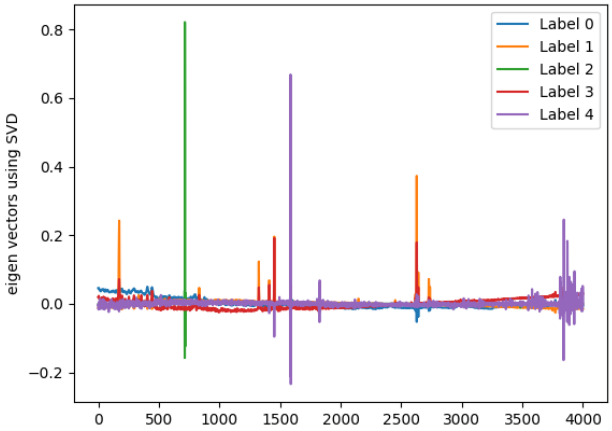

Figure 5: 5 eigenvectors obtained using SVD method

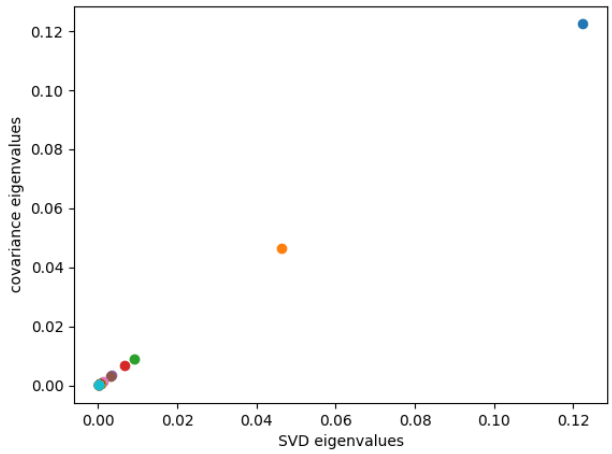

Figure 6: eigenvalues calculated using SVD method against that of covariance method

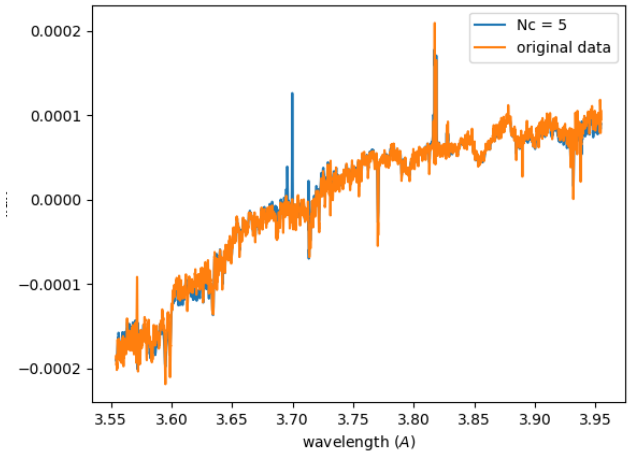

Figure 7: original data vs approximate data, only first 5 coefficient is used

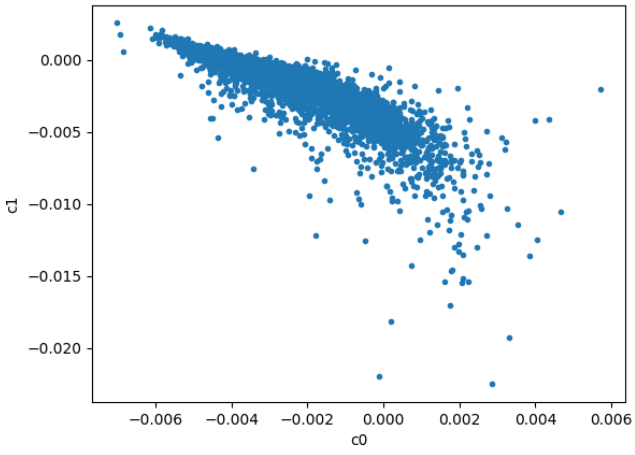

Figure 8: c0 vs c1

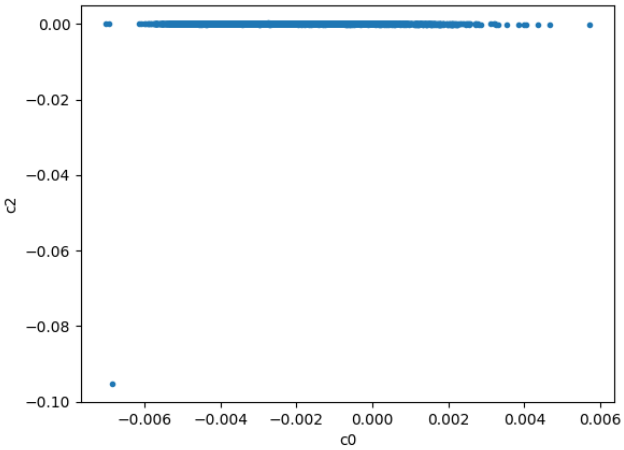

Figure 9: c0 vs c2

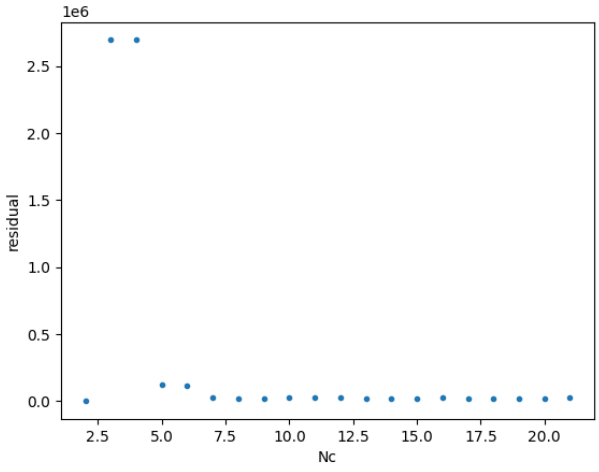

Figure 10: N_c versus squared fractional residual