Problem Set 6

Ziyao Wang(Github: yktaykketo)

November 1, 2023

1 a

1.1 a

I plot 3 galaxies and find that there are some striking peaks at certain wavelength which is similar to the hydrogen spectrum. It might mean that there are some hydrogen components inside of these galaxies.

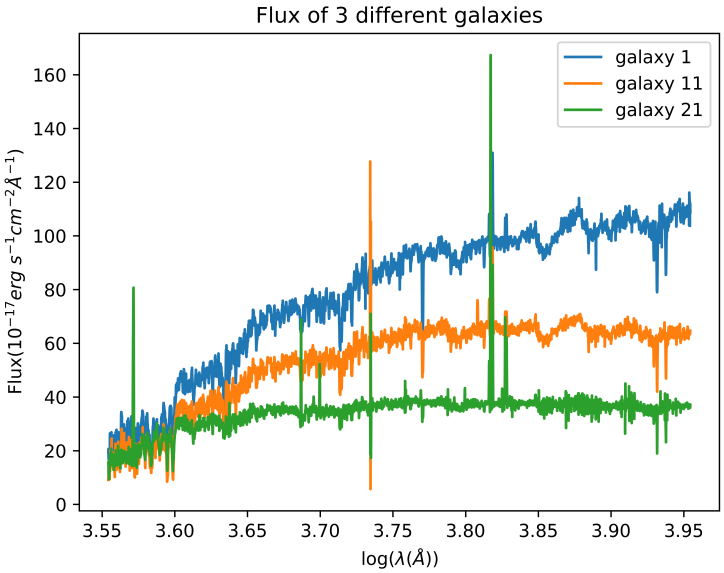

Figure 1: The flux of 3 different galaxies.

1.2 b,c

To make the PCA more meaningful, first, I normalize all the fluxes so their integrals over wavelength are the same. Then, I subtract the mean value of it. Here is one of the galaxies after conducting this process.

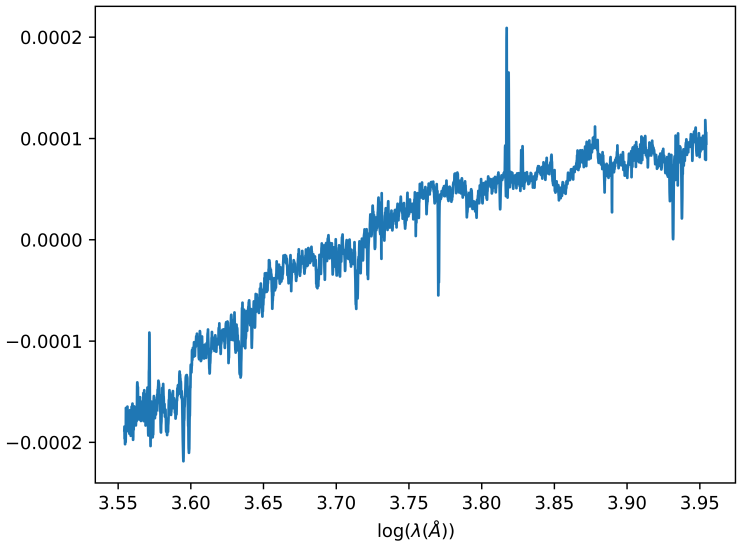

Figure 2: Flux picture after normalization and subtracting the mean value.

1.3 d

The idea of the PCA is to find the eigenvectors of the covariance matrix of the distribution. So my first step is to get the covariance matrix. Then I use the np.linalg.eig function to calculate the eigenvalue and eigenvector. Here I plot the first 5 eigenvectors.

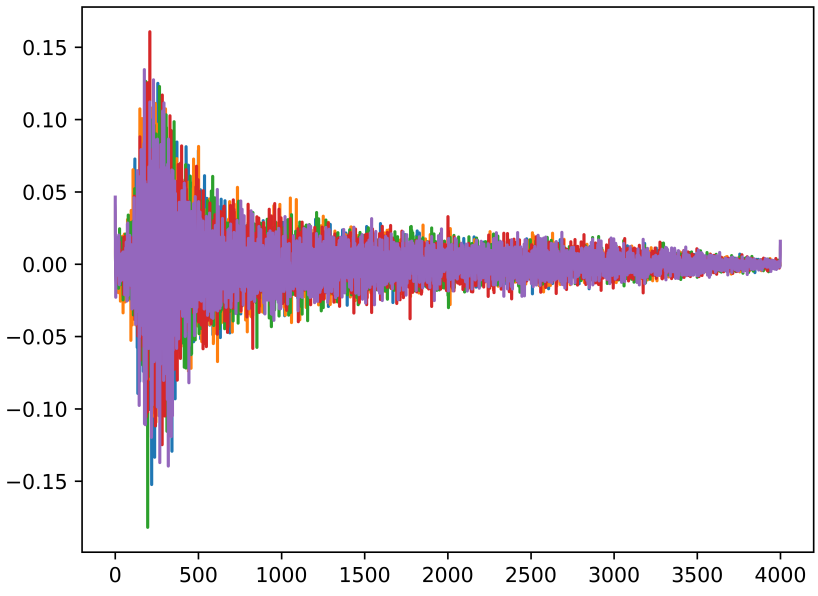

Figure 3: First five eigenvectors using the covariance eigenvalue methods.

1.4 e

In this part, I calculate the eigenvectors using an SVD decomposition of R. These vectors are the same as the previous one except that there is a minus sign difference in some eigenvectors. The SVD method takes more time than the previous one.

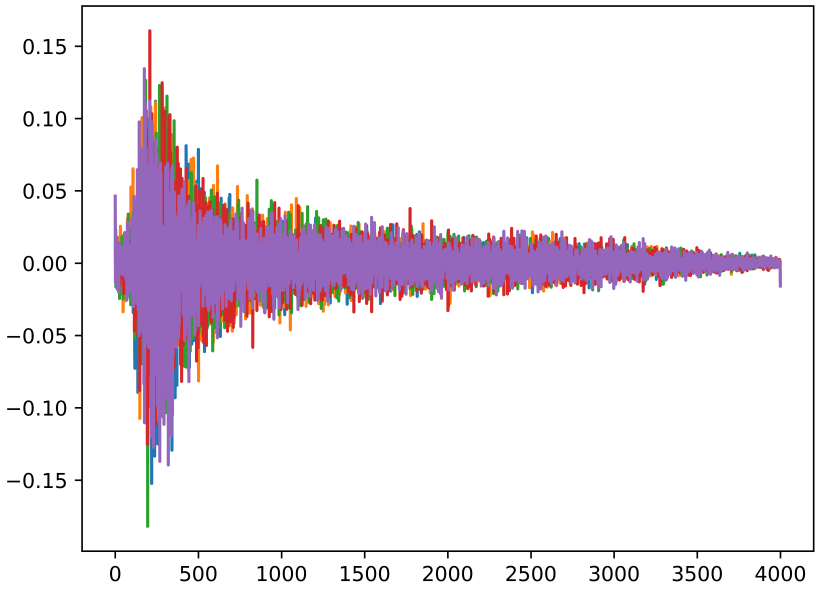

Figure 4: First 5 eigenvectors using the SVD decomposition of R.

1.5 f

The conditional number of C in the covariance method is the square of that of R in the SVD method. So using the SVD method is a more stable way.

1.6 g

I keep the first Nc = 5 coefficients and create the approximate spectra. We can see that using the first 5 coefficients reconstructs the original data very well. If we use more coefficients, the data that we construct will become closer to the original data.

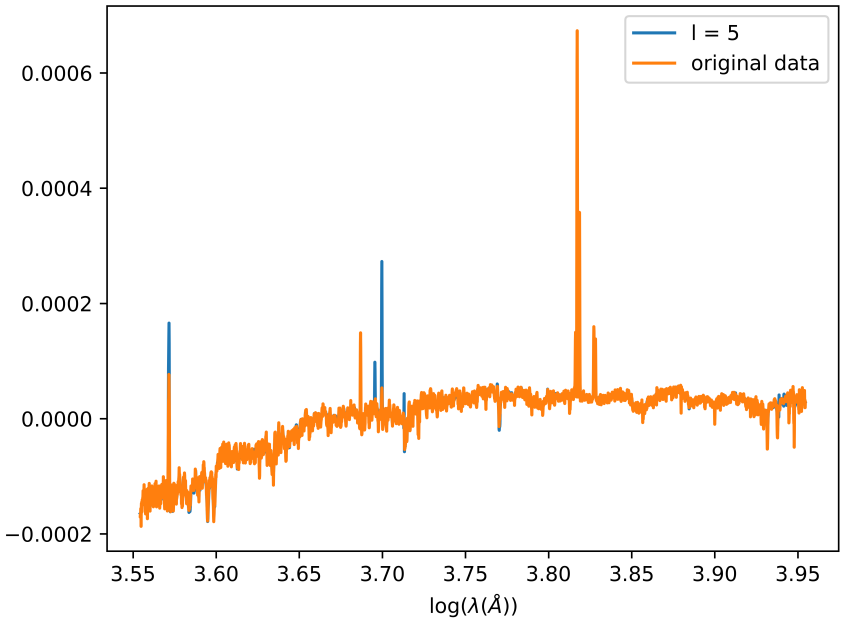

Figure 5: Reconstruct the data using the first 5 coefficients and compare it to the original one. We can see that the error is small.

1.7 h

The points in the c0 vs c2 graph are almost horizontal along c0. We can see that the spread of the c1 vs c0 are wider. This suggests that the previous principal component captures more information than the later ones.

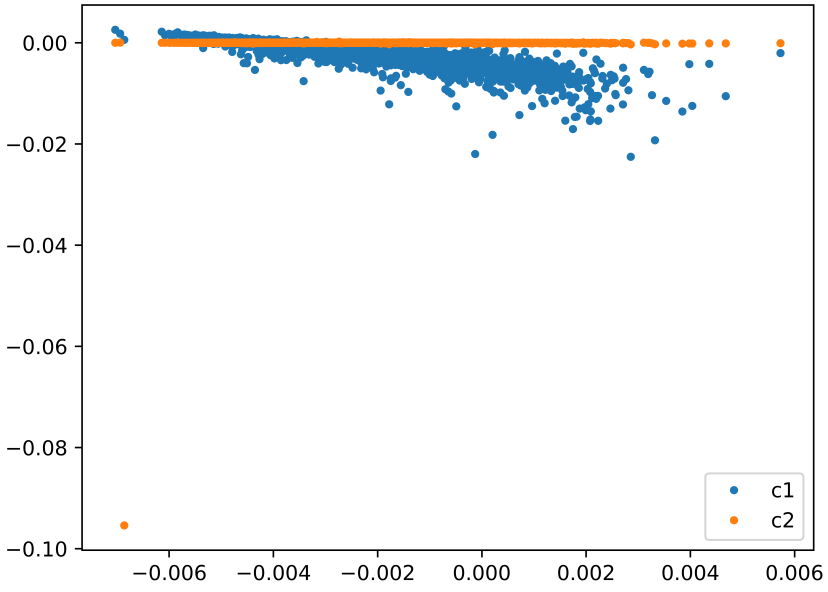

Figure 6: c0 vs c1 and c0 vs c2.

1.8 i

I plot the squared fractional residuals between the spectra and the reconstituted, as a function of N_c. We can see that the residuals decline. As the number of coefficients gets closer to 20, the residuals are about the amount of 10^{-7} magnitude so it is a good approximation. This means that this PCA method compacts the data in a good way and keep a lot of information about the original data. The more coefficients we use, the more precise reconstructed data we get.

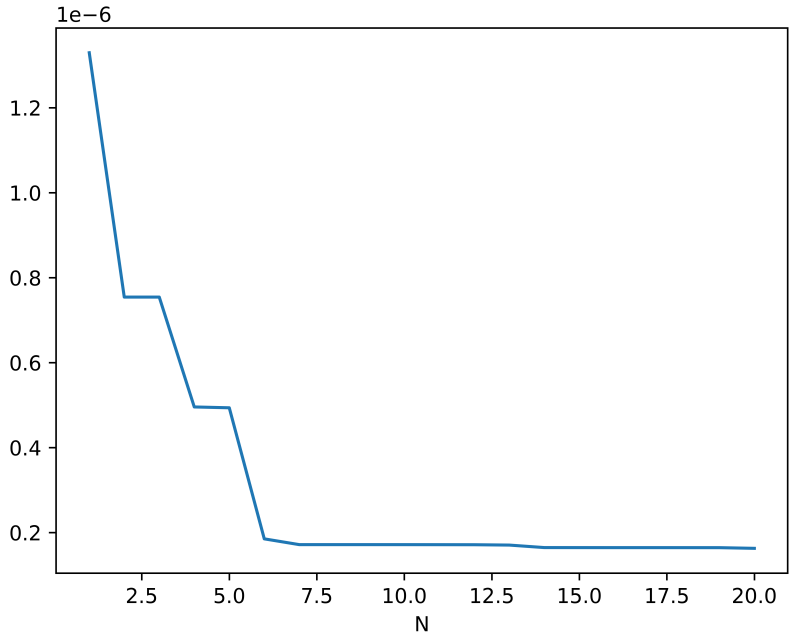

Figure 7: Squared fractional residuals between the spectra and the reconstituted, as a function of Nc.