

Computational Physics - Problem Set 6

Frederik Holst Knudsen

October 14, 2024

Github URL: <https://github.com/frederikholst/phys-ga2000>

1 Part A

The data is imported and loaded using the astropy package. The peaks we see from several of the galaxies in Figure 1 are probably from the Balmer series in the Hydrogen atom, where higher orbitals larger than $n=6$ transition to $n=2$, where n is the principal quantum number of the Hydrogen atom.

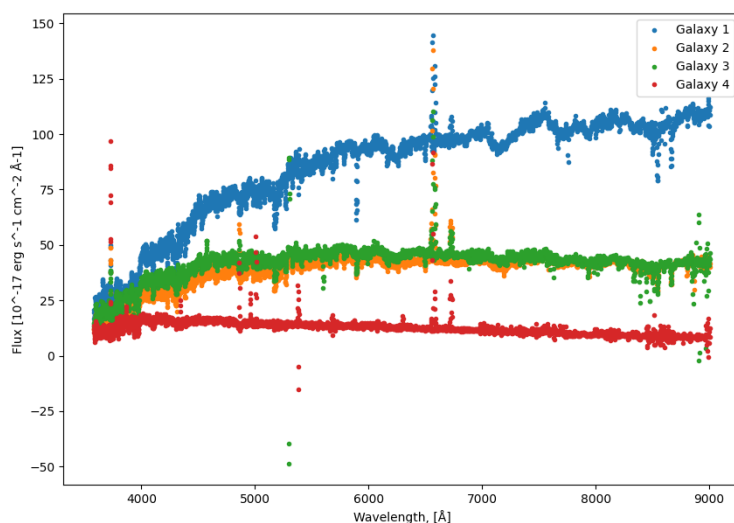


Figure 1: The spectra of the first five galaxies are seen plotted against wavelength.

2 Part B

We now normalize the flux by dividing each spectrum with their integral, see Figure 2.

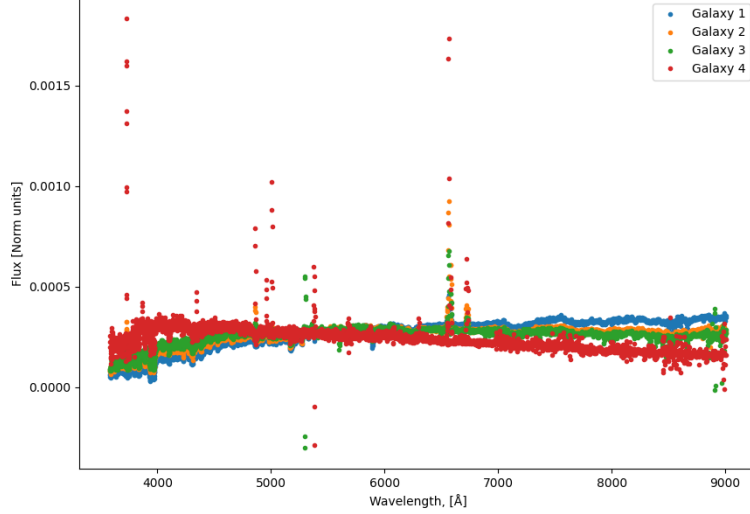


Figure 2: The spectra of the first five galaxies as seen in Figure 1, but now normalized.

3 Part C

We will subtract the mean from each spectra and keep the mean stored in an array for later. See Figure 3.

4 Part D

The PCA is performed by computing the covariance matrix: $C = R \cdot R^T$, where R is the residual matrix that we have found previously. The eigenvalues and eigenvectors are found using `np.linalg` methods and are plotted in Figure 4.

5 Part E

Using the SVD technique, we find the eigenvectors of $C = R \cdot R^T$, by using that:

$$R \cdot R^T = V \cdot W \cdot U^T \cdot U \cdot W \cdot V^T$$

Here The eigenvectors of V corresponds to that of C . In Figure 5, the residuals between the SVD on R and the regular eigenvector method on C are plotted for the first five eigenvectors. Due to sign ambiguity of the eigenvectors, a small dot product test is implemented to make sure the correct sign is chosen for the computation of the residuals.

We compute the computation time for a sample size of 500 to be 4.587 seconds for the SVD method and 0.203 seconds for the `np.linalg(C)` method.

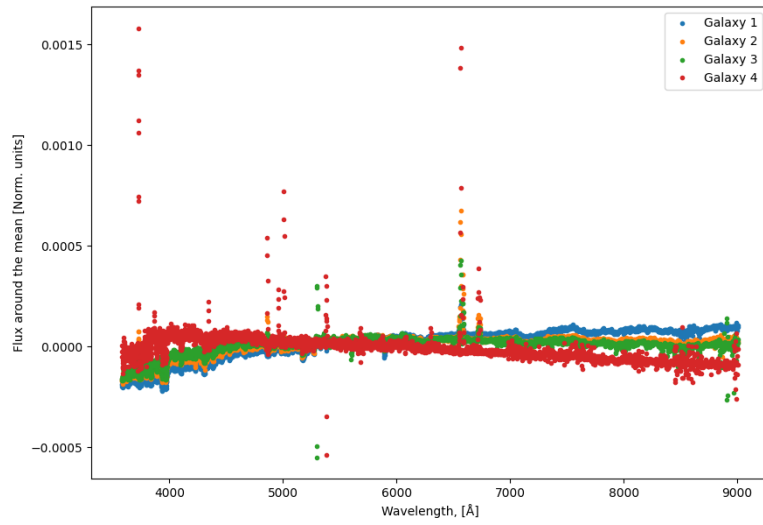


Figure 3: The normalized spectra of the first five galaxies with subtracted mean.

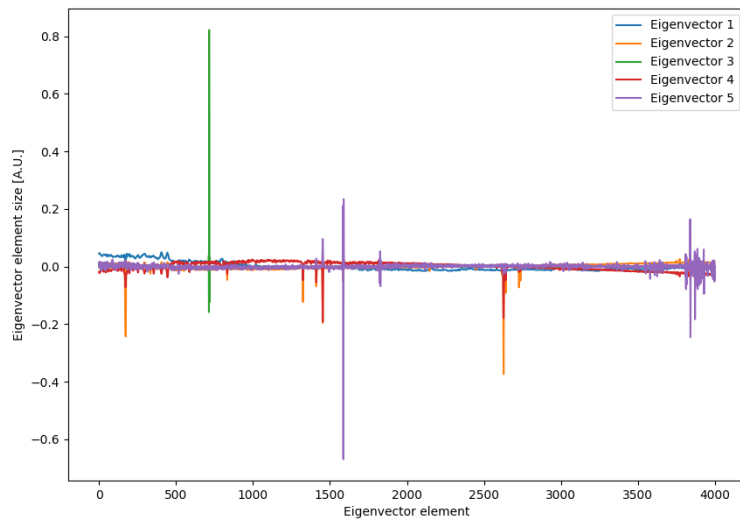


Figure 4: The first five eigenvectors plotted with the value of each element in the y-axis.

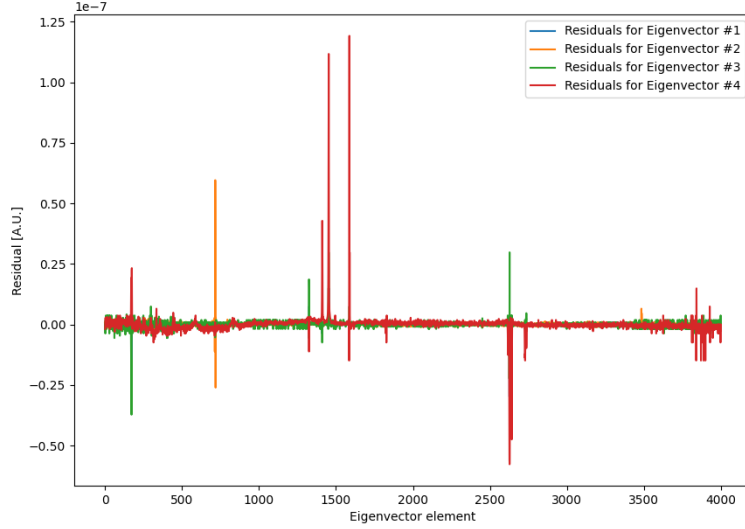


Figure 5: The residuals are plotted between the two different methods of finding the eigenvectors.

6 PART F

A way to assess the two types of methods besides the computation time, is to look at the condition number. For the Covariance matrix the condition number is 524599.2 and for the residual matrix, R the condition number is only 724.32874, meaning that the residual matrix is way less sensitive to input parameters and errors than that of the covariance matrix. So, using SVD directly on R leads to more reliable eigenvector and eigenvalue computations, but at the cost of longer computation time.

7 PART G

We create approximate spectra based on keeping only the first five coefficients of the PCA. First the weight coefficients for the principal components are found by rotating the spectra into the eigenspectrum basis. This is implemented by projecting the eigenvectors of C onto R . See Figure 6, for comparison of the first spectrum with the PCA reconstruction.

8 PART H

See Figure 7 for plots of c_0 vs. c_1 and c_2 respectively.

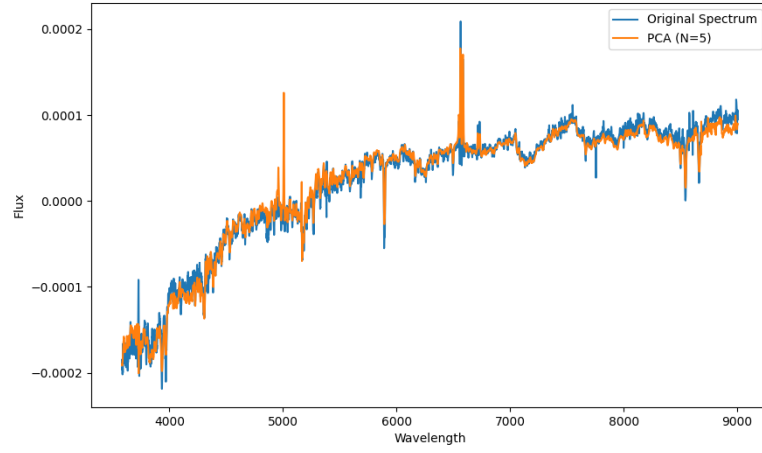


Figure 6: The two spectra are plotted and we see very strong agreement considering only five principal components are used.

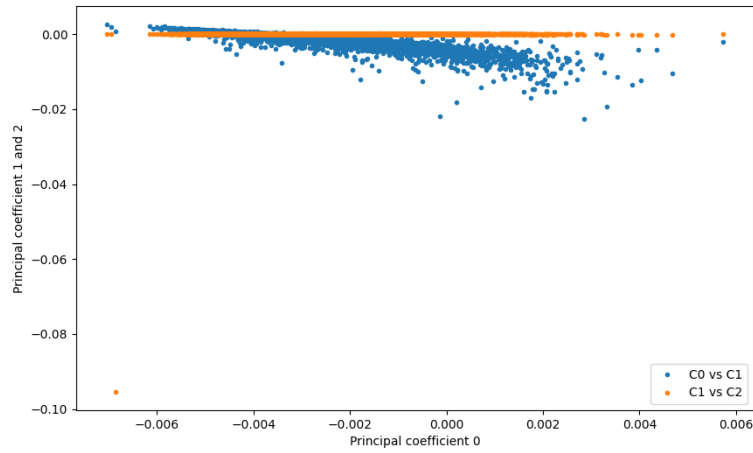


Figure 7: The relation between the first few PCA coefficients are plotted.

9 PART I

See Figure 8 for how the squared residuals of the PCA reconstruction decreases as we increase the number of PCA coefficients we include in the reconstruction.

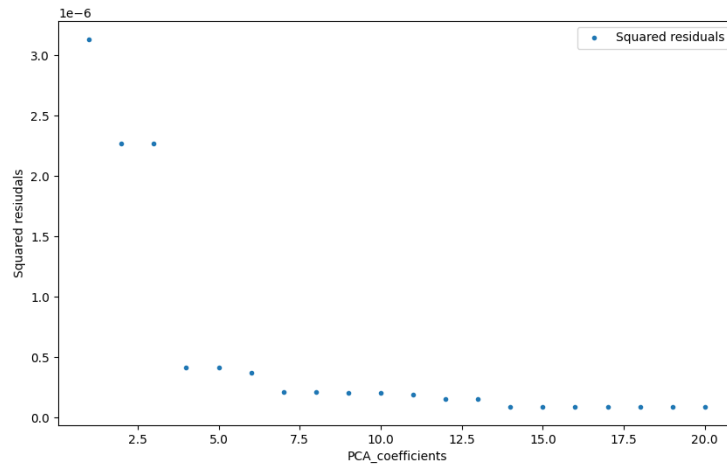


Figure 8: We see that the squared residuals fall off very quickly. Notice the scale at $1e-6$. At $N_c=20$ we have the squared residuals at 9.36×10^{-8} which is well within the machine error.