Problem Set 6

Jason Li

November 2, 2023

Contents

Problem 1

(a) The graph of the first 5 wavelengths are displayed below. In the figure,

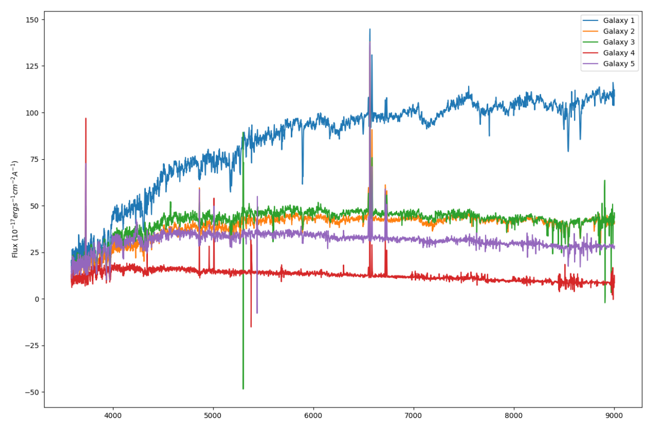

Figure 1: Wavelength of first 5 galaxies

we can observe several spikes in the spectra. These probably correspond to the emissions of photons by electrons during their excited states.

(b) See code.

(c) See code.

(d) The plot of the first 5 eigenvectors are displayed below.

(e) Using the integrated scipy SVD decomposition, we can use SVD decomposition to obtain **R**. We compare the eigenvectors obtained from the

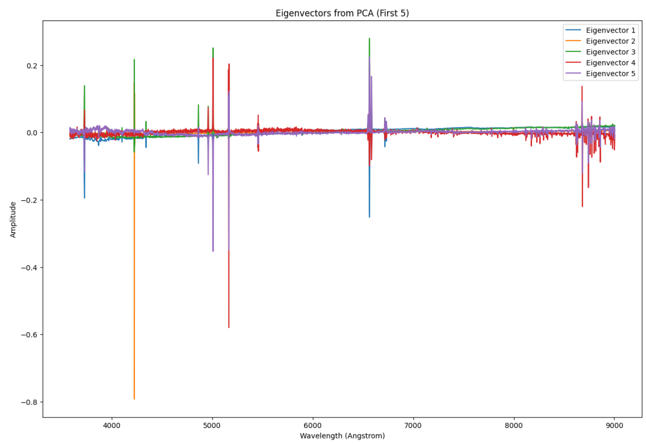

Figure 2: Eigenvectors of first 5 galaxies

covariance matrix, $\mathbf{E}$, with those obtained from the singular value decomposition, $\mathbf{E}_{\mathrm{SVD}}$, by computing the norm of their difference. The comparison accounts for the potential sign ambiguity in eigenvectors, as they can be negated and still represent the same subspace. This is corrected by ensuring the dot product between corresponding pairs of eigenvectors is positive before the comparison:

$$\text{For each } i, \text{ if } \langle \mathbf{E}_{:,i}, \mathbf{E}_{\mathrm{SVD},:,i} \rangle < 0, \text{ then } \mathbf{E}_{\mathrm{SVD},:,i} = -\mathbf{E}_{\mathrm{SVD},:,i}.$$

After aligning the eigenvectors, we calculate the Euclidean norm (or l_2 norm) of their difference for each corresponding pair:

$$\|\mathbf{E}_{:,i} - \mathbf{E}_{\mathrm{SVD},:,i}\|_2 = \sqrt{\sum_j (\mathbf{E}_{j,i} - \mathbf{E}_{\mathrm{SVD},j,i})^2},$$

where $\mathbf{E}_{:,i}$ and $\mathbf{E}_{\mathrm{SVD},:,i}$ represent the i-th columns (eigenvectors) of $\mathbf{E}$ and $\mathbf{E}_{\mathrm{SVD}}$, respectively, and j indexes over the elements of the eigenvectors. A smaller norm indicates a higher similarity between the eigenvectors from the two different methods.

The norms of the differences between the two are displayed in the code, and at first glance, the error seems to be quite close to 0, indicating a high similarity between the two eigenvectors. There seem to be some entries with noticeable differences however.

If the data is large, since the computation of the covariance matrix involves matrix multiplication, it is less efficient than if one was directly to compute the SVD.

(f) As previously mentioned, if the data is large, since the computation of the covariance matrix involves matrix multiplication, it is less efficient than if one was directly to compute the SVD. Additionally, a high condition number indicates that computations involving the matrix is numerically unstable. As seen in the code, the condition number of the covariance matrix and the residuals respectively are as follows:

Condition number of C: 41343037000.0
Condition number of R: 5149336.0

Therefore, using the SVD decomposition is more numerically stable, as the condition number of R, while large, is significantly smaller than that of C.

(g) See code.

(h) The plots for c_0 vs c_1 and c_0 vs c_2 are displayed below:

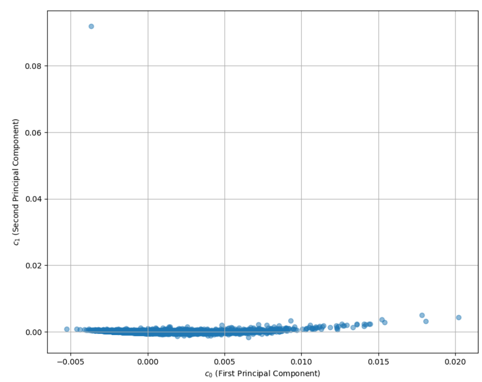

Figure 3: Plot of c_0 vs c_1

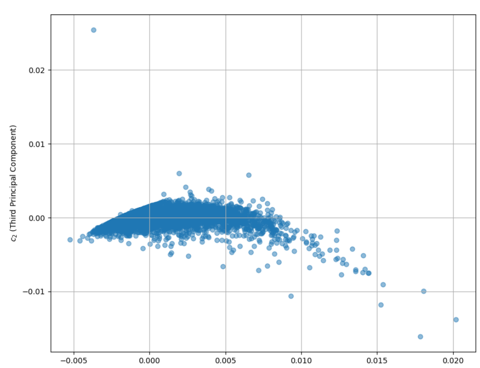

Figure 4: Plot of c_0 vs c_2

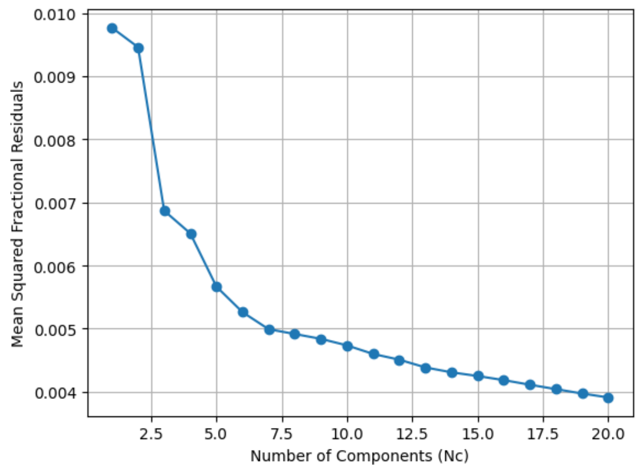

Figure 5: Fractional error vs N_c

(i) The plots for Fractional error vs N_c are displayed above. As you can observe, the squared residuals are declining. The fractional error for $N_c = 20$ is 0.029336463660001755