Sixth Assignment for Computational Physics

Xinyu Liu

October 15, 2024

Contents

1 My Github Page URL

https://github.com/rising1227/phys-ga2000

2 PCA analysis of a real astrophysical data set

The Associated Code for this assignment is ps6.py.

2.1 A

We've imported the data with astropy package. As a physicist, we know that hydrogen atoms have typical energy level structure with $hc/\lambda = 13.6eV * (1/n^2 - 1/m^2)$. For the related optical frequency in observed data set, we take n = 2 and m = 3,4,5. The transition wavelength should be 656.47nm, 486.27nm and 434.17nm. We plotted several flux data for different galaxy and also plotted some vertical line related to the special transition wavelength. We do observe some kind of structure around those wavelength.

Table 1: Flux wrt wavelength for galaxy 1

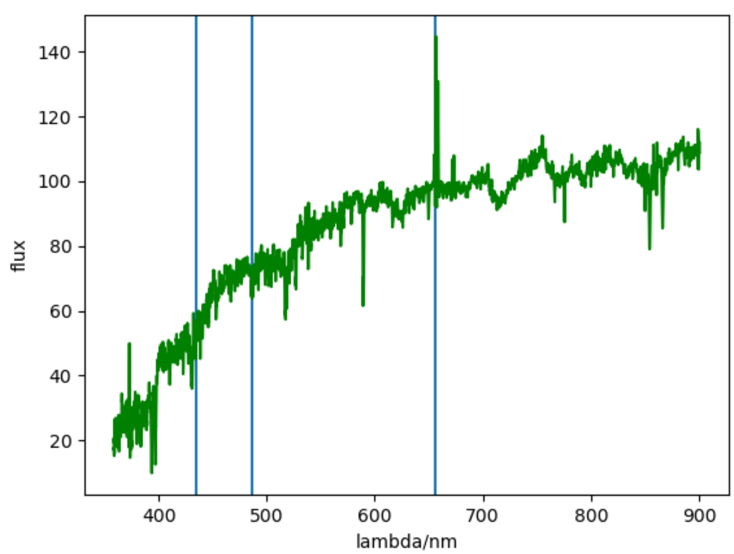


Table 2: Flux wrt wavelength for galaxy 2

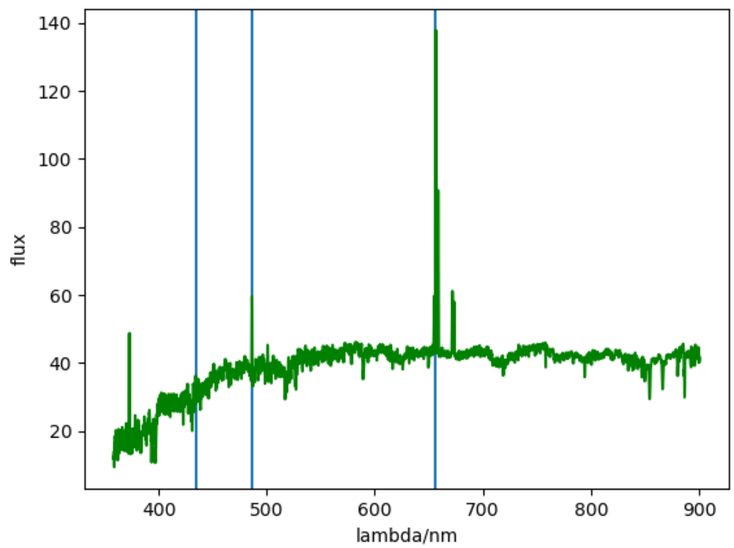


Table 3: Flux wrt wavelength for galaxy 3

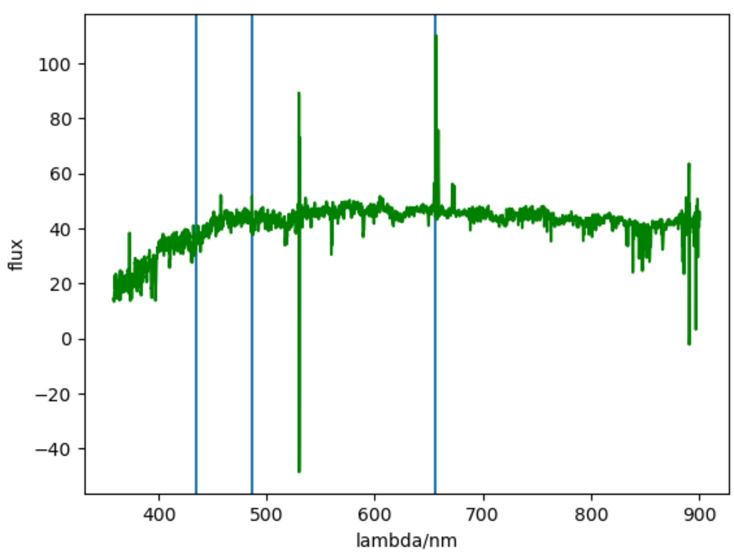


Table 4: Flux wrt wavelength for galaxy 4

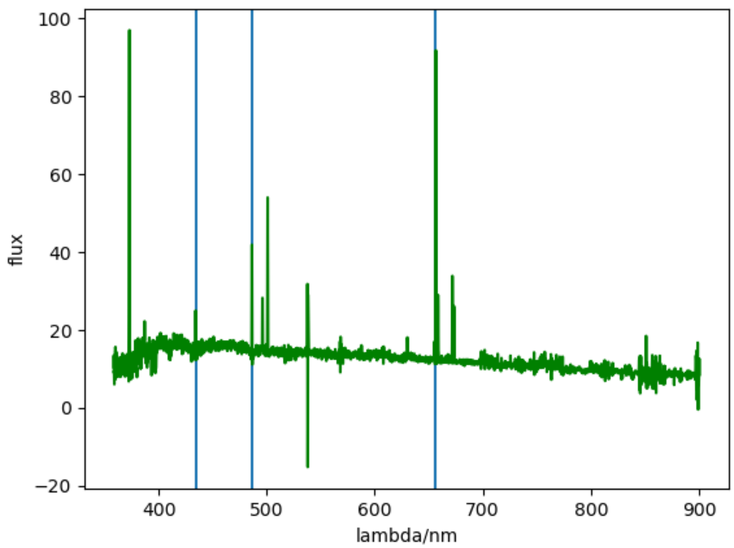
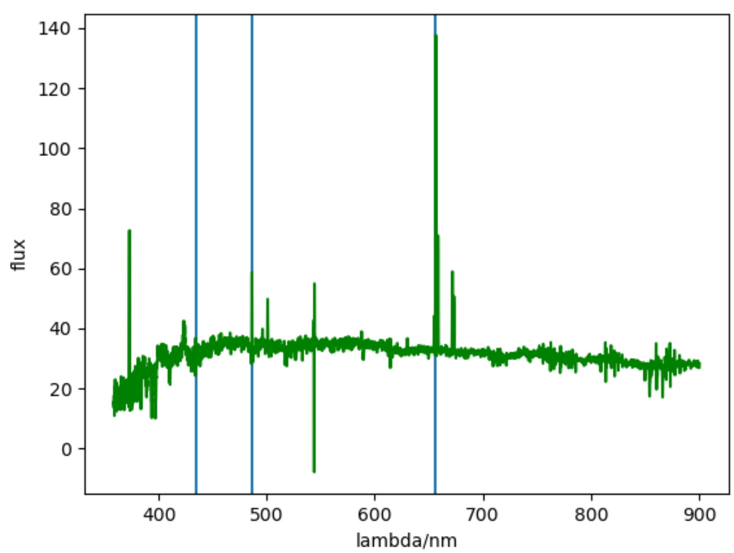


Table 5: Flux wrt wavelength for galaxy 5



2.2 B,C

We've finished the required normalization process in our code. The result is stored in a Array Normalization and Res for future use.

2.3 D

We measure the covariance matrice and carried out the PCA. We plotted five engenstate with largest eigenvalue.

Table 6: 1st PCA eigenstate

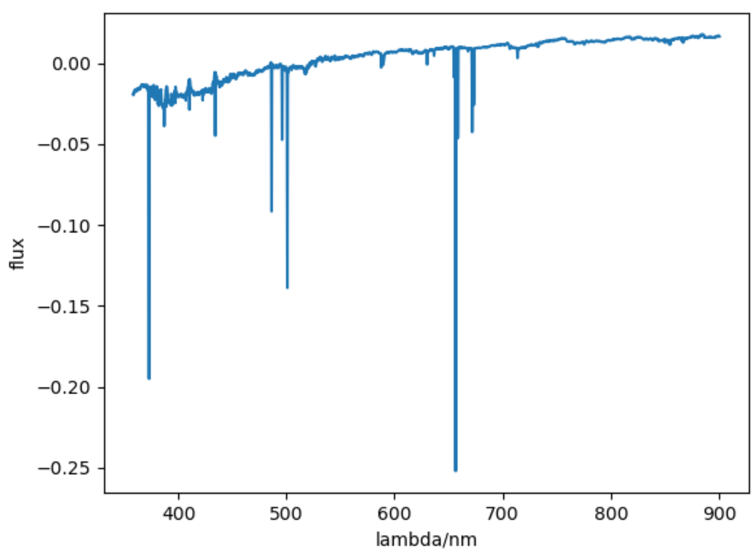


Table 7: 2nd PCA eigenstate

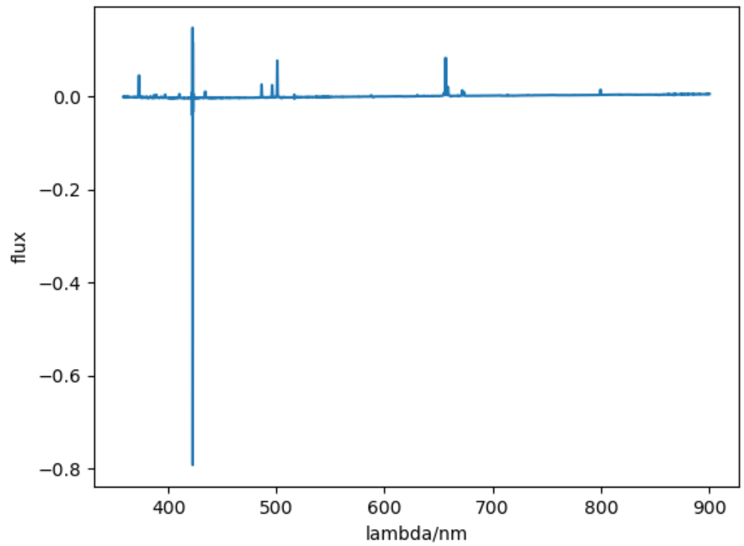


Table 8: 3rd PCA eigenstate

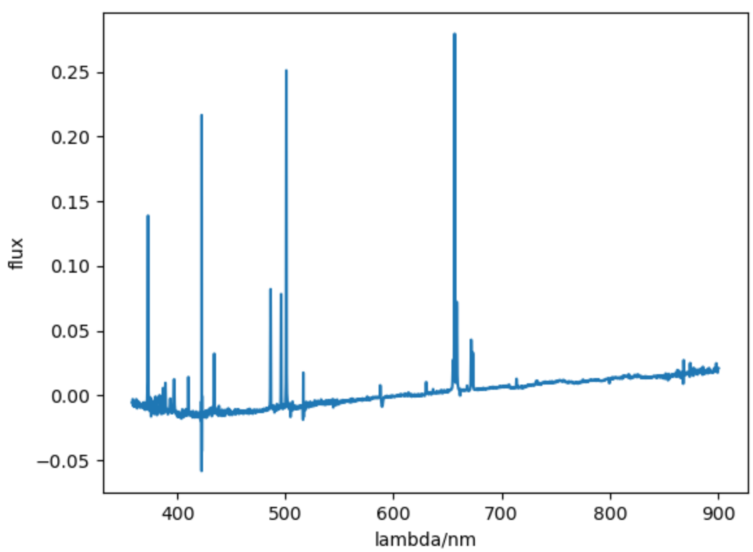


Table 9: 4th PCA eigenstate

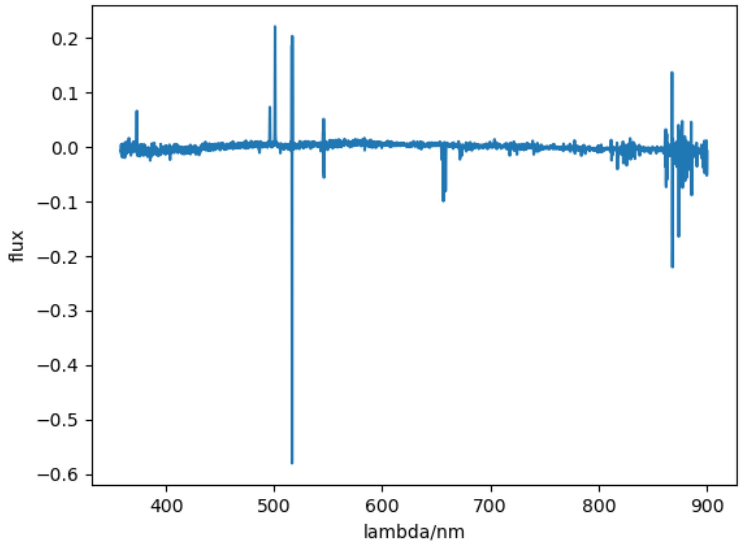
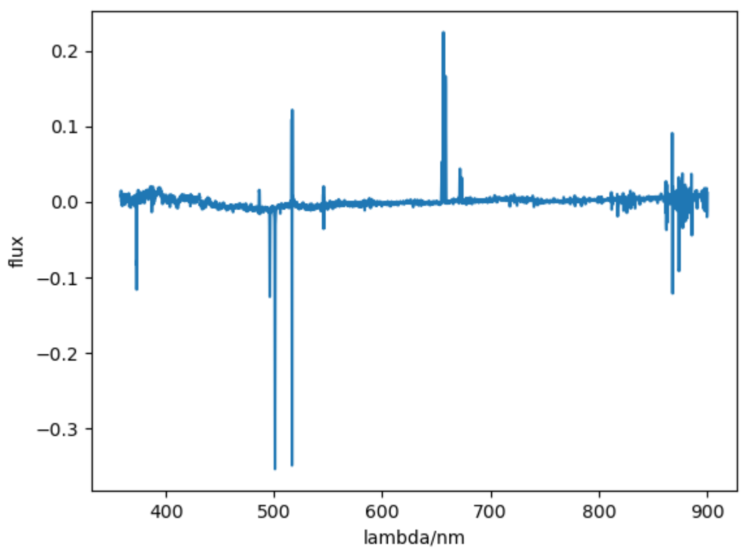


Table 10: 5th PCA eigenstate



2.4 E

We can similarly calculate the eigenstate by doing SVD directly to R matrice. We've explicitly test that the eigenstate by SVD is equvalent to what's found before(we've calculated the cosine of related angle of two states.)

We found that the SVD method cost 37.32s while diagonialize a $R^T R$ matrice in PCA cost 25.4s. It seems that doing SVD cost slightly more time.

2.5 F

We also found that the conditional number for doing a diagonalization for $R^T R$ is much larger than doing SVD to R(which is very easy to understand, since SVD is somehow a squareroot of diagonalization). The conditional number for SVD is 5205738 while for PCA diagonalization is 27099608379355. When the data set is even larger, doing SVD, although cannot make computation faster, do enhance the stablization of the computation.

2.6 G

We can explicitly calculate the coefficient on first 5 eigenstates by doing rotation. Then we can approximate the real data by only storing several coeffieicnt before those 5 eigenstates. We've plotted the approximated data and real flux data in the following diagram:

Table 11: flux 0 vs approximation

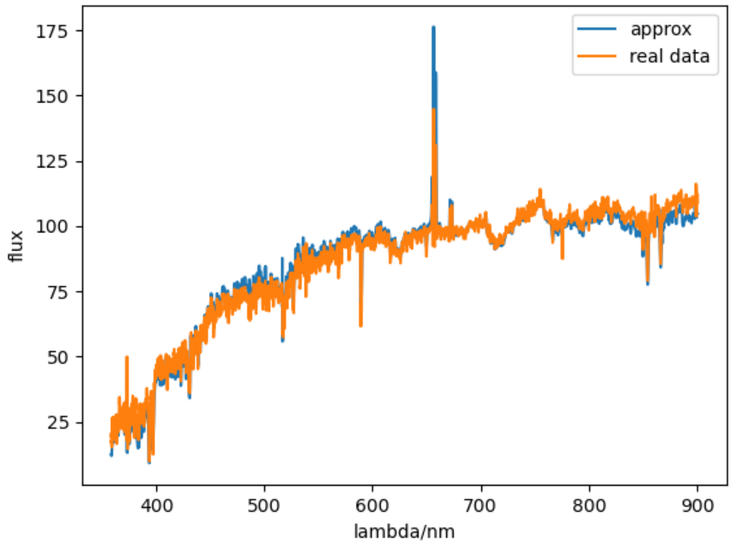


Table 12: flux 1 vs approximation

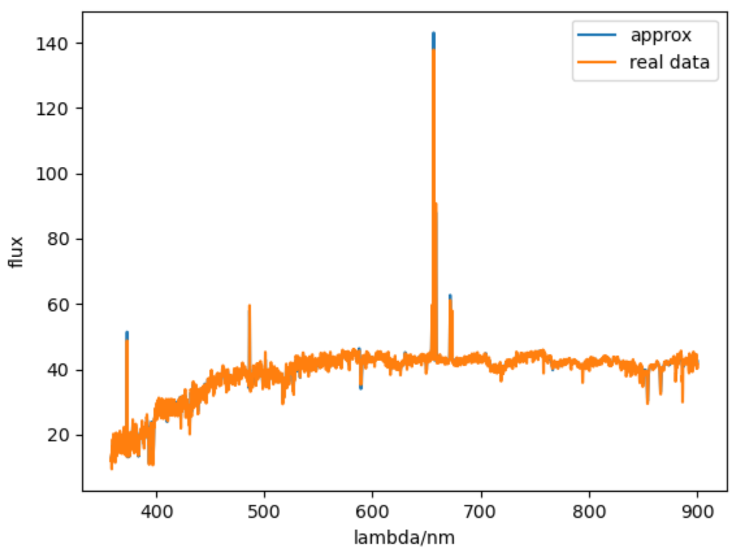


Table 13: flux 2 vs approximation

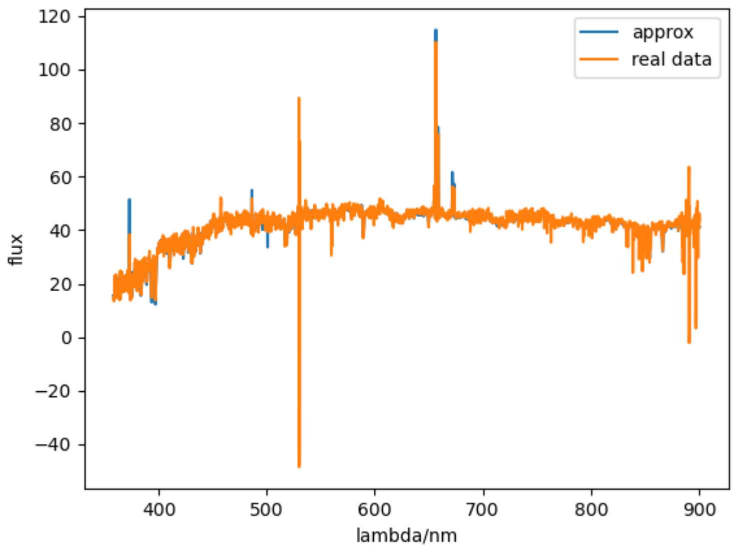


Table 14: flux 3 vs approximation

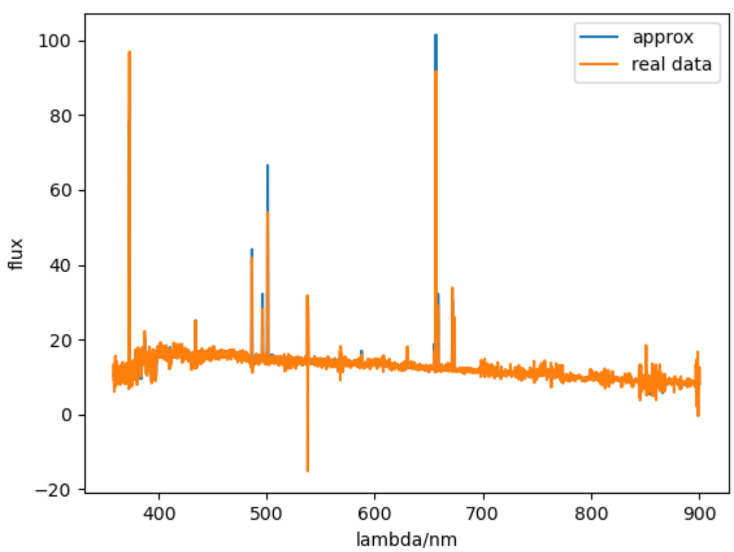
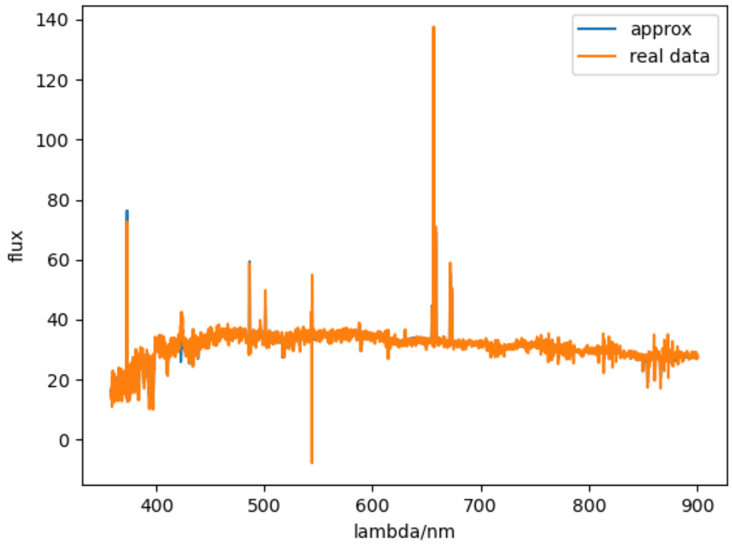


Table 15: flux 4 vs approximation



2.7 H

We've plotted C0 vs C1 and C1 vs C2:

Table 16: C0 vs C1

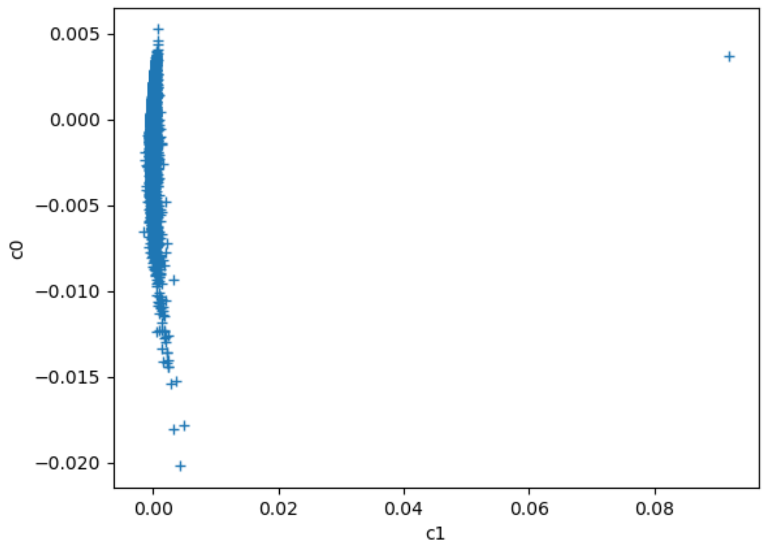
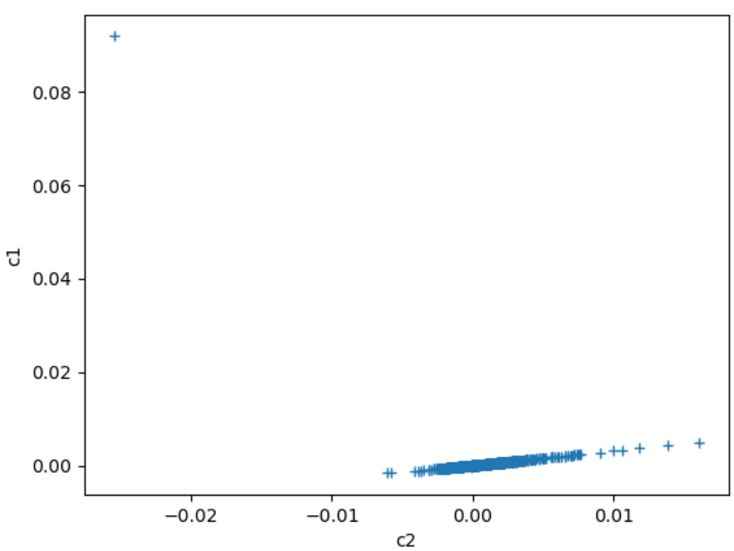


Table 17: C1 vs C2



2.8 I

We've calculated the Error of approximation with different number of principal components from 0 to 20. Here's our result and we've explicitly found that the error goes down when Nc grows higher. When Nc equals 20 the error(sum of the square of difference between real data and approximation) goes down to 8675.

Table 18: 5th PCA eigenstate