PS6

Shihong Pan
https://github.com/PSH-hub24/phys-ga2000

October 16, 2024

1 part a

I plotted the first four galaxies as shown in Fig 1. I noticed that all 4 subplots have a peak around $x = 3.82$, which translates to about $6562\mathring{A}$. This is the red line in the hydrogen spectrum.

2 part b

Standard normalization. Please see the codes (I divided the codes into the corresponding parts).

3 part c

Please see the codes.

4 part d

Fig 2 plots the first five eigenvectors of the covariance matrix (the x-axis is the entry index of the vectors).

5 part e

Similar to the fact that V is composed of the right eigenvectors of $R^T R$, the matrix U is composed of the right eigenvectors of $RR^T = C$, which gives the answers equivalent to part d. Fig 3 prints the computation time of the two different methods. Finding the eigenpairs of the covariance matrix is faster.

6 part f

The SVD is numerically more stable. As shown in Fig 4, C has a much higher condition number than R, implying larger numerical instability.

7 part g

I rotated the spectra and record the first 20 coefficients for the later uses. Please see the codes for the implementation.

8 part h

Fig 5 plots the first three coefficients.

9 part i

Fig 6 plots the squared residuals as a function of N_c. It is indeed declining.

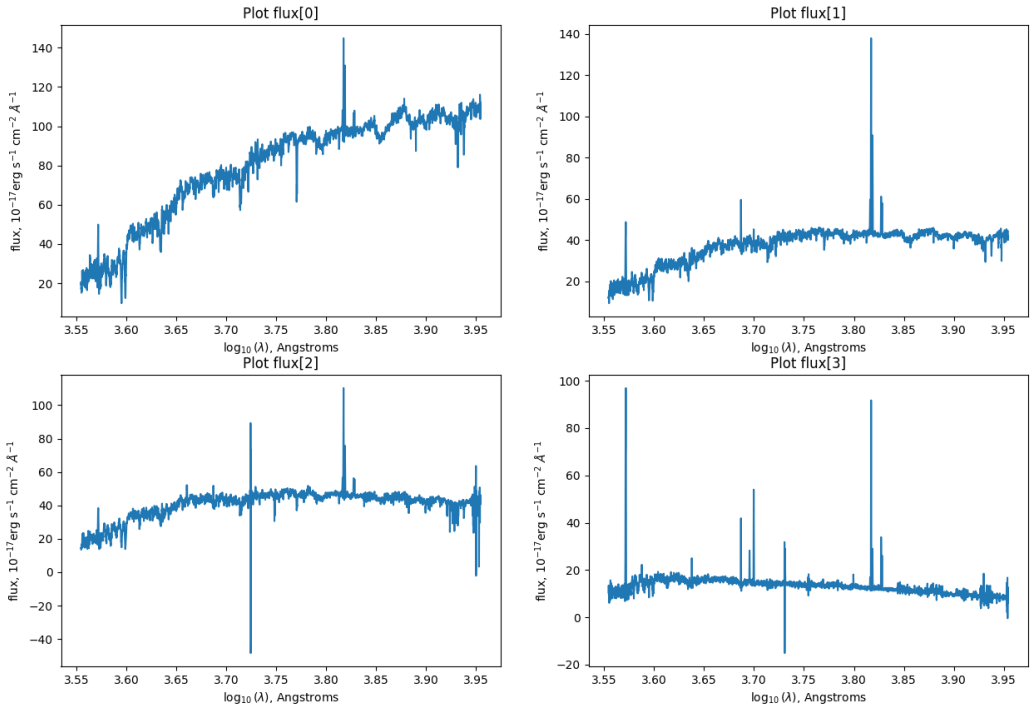

Figure 1: The first four galaxies.

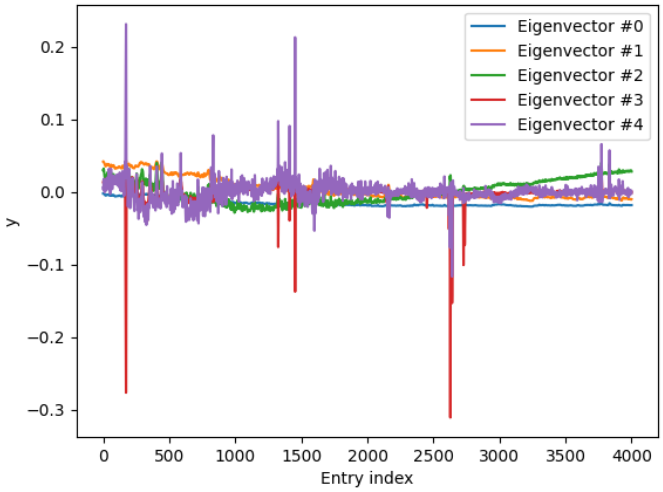

Figure 2: The first five eigenvectors of the covariance matrix.

```
The covariance matrix method takes 29.1403232.
The SVD matrix method takes 48.4955746.
[Finished in 80.9s]
```

Figure 3: Computation time (in seconds) of the covariance matrix method and the SVD method.

```
The condition number of C is 155593952.0.
The condition number of R is 12429.4951171875.
[Finished in 36.7s]
```

Figure 4: Condition numbers of C and R.

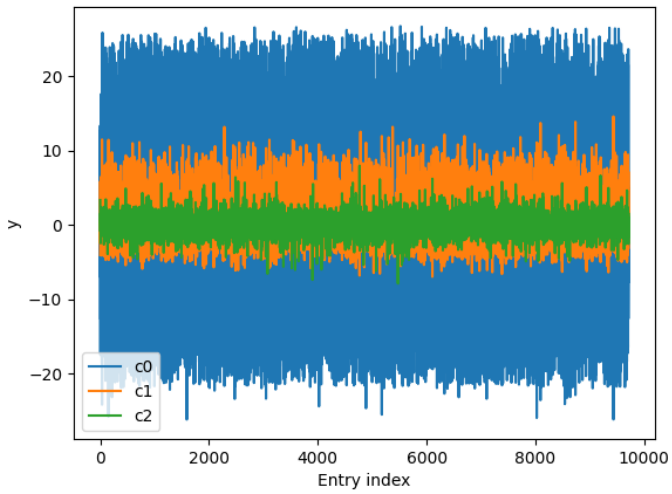

Figure 5: c_0, c_1, and c_2 of the PCA.

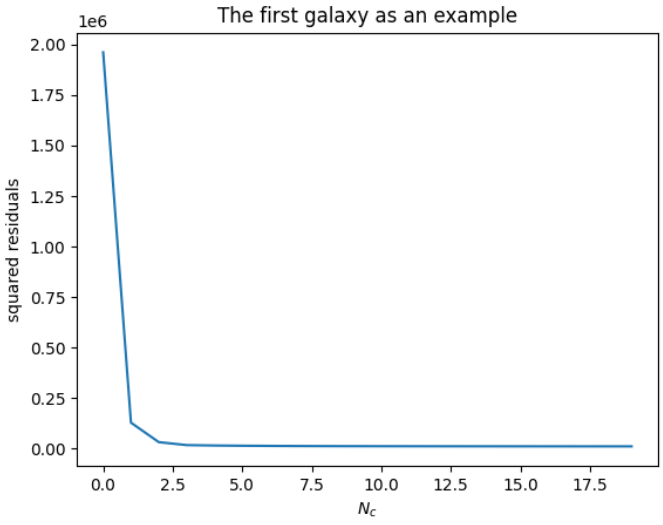

Figure 6: The squared residuals between the spectra and the reconsituted, approximate spectra as a function of N_c.

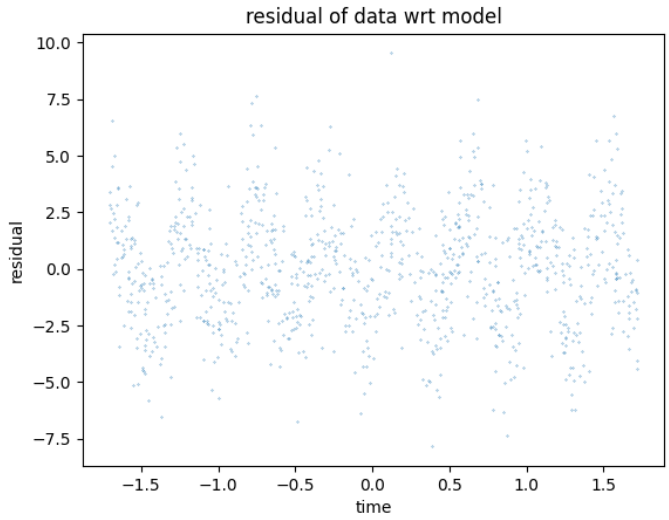

Figure 7: Q3(c): The residuals of the data wrt the model in part (b).

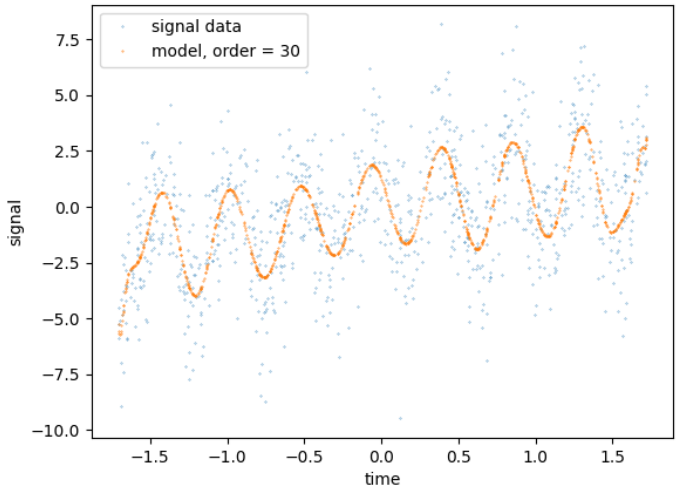

Figure 8: Q3(d): The signal data and the model with the order of polynomial = 30.

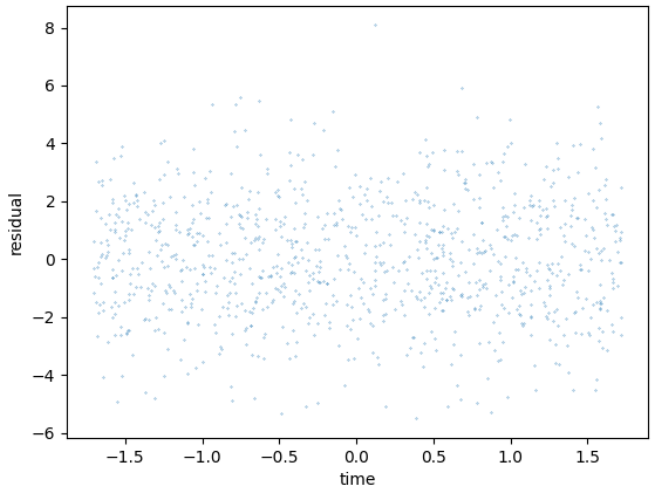

Figure 9: Q3(d): The residuals of the data wrt the model with the order of polynomial = 30.

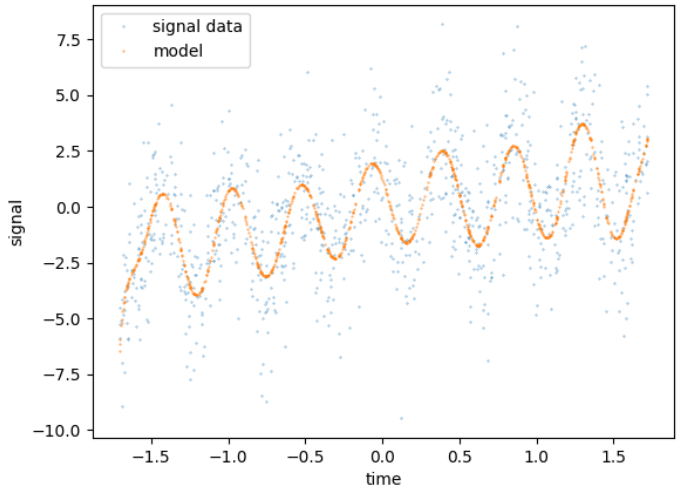

Figure 10: Q3(e): Use the Lomb-Scargle model to fit the data.

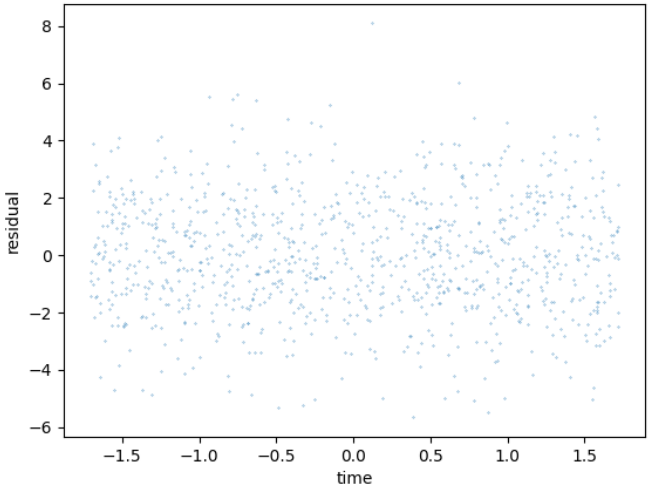

Figure 11: Q3(e): The residuals of the data wrt the Lomb-Scargle model.