Measuring Galactic Redshift

John Gillan

1 Introduction

The aim of this project is to calculate the redshift of a galaxy from its spectrum. The spectrum is from the Sloan Digital Sky Survey (SDSS), a spectroscopic survey of mostly distant galaxies and quasars. As the universe is expanding, distant objects are receding from us, and the further an object is the faster it is receding. When the light emitted by these objects reaches us, even though it is still travelling at the same constant speed, it has been stretched out by the intervening expanding space, causing the light to be shifted to the redder end of the spectrum.

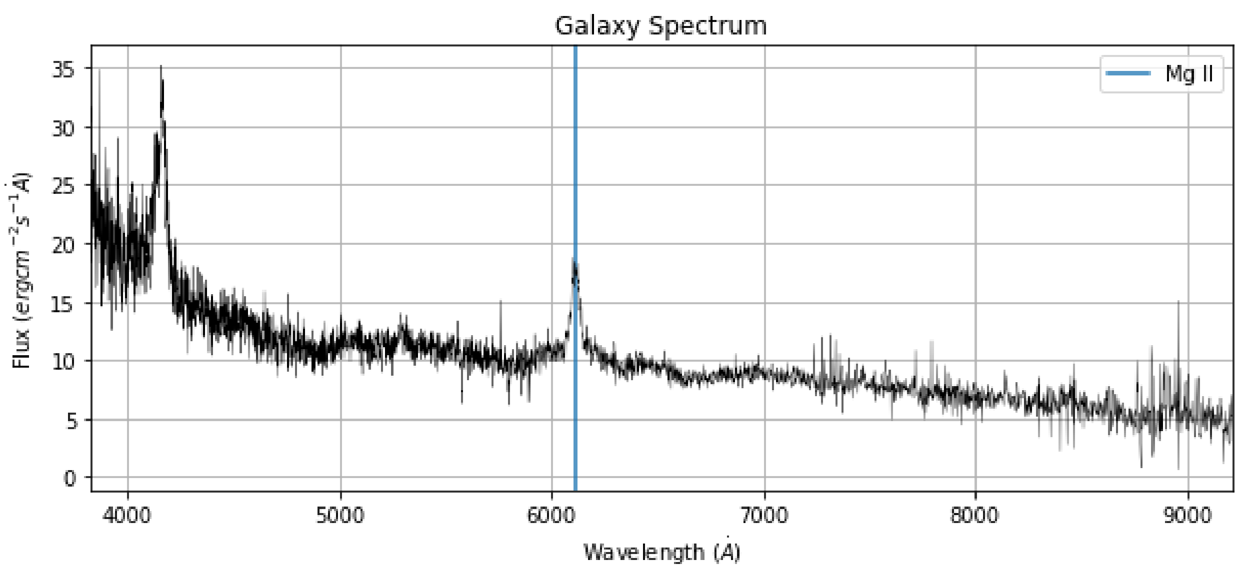

Figure 1: Plot of entire spectrum of data under investigation. The spectrum is from data set 2 provided. The Mg II line is marked in blue.

The redshift is found with:

$$z = \frac{\lambda_{obs} - \lambda_{lab}}{\lambda_{lab}} \quad (1)$$

where λ_{obs} is the wavelength of a prominent feature in the spectrum, and λ_{lab} its corresponding wavelength observed in controlled conditions on earth. One such prominent feature in the spectrum is the Mg II emission line. It is measured in the lab as occurring at 2800.3 Å, and comparing this value to the one measured in the galaxy will allow the calculation of the redshift. However, the measured spectrum of the galaxy will have associated uncertainties which could affect the result of the calculation. To overcome these uncertainties and ensure that the calculated redshift is accurate to a satisfactory degree, it will be required to analyse and minimise the errors of the spectrum around the Mg II emission line. If a section of the spectrum containing just an emission line and background continuum either side is chosen, then it can be modelled by:

$$f(\lambda|\theta) = \theta_1 + \theta_2(\lambda - \lambda_{ref}) + \theta_3 L\left(\frac{\lambda - \theta_4}{\theta_5}\right) \tag{2}$$

where θ_1 is the flux at a point chosen to represent the continuum of the spectrum without the emission line, θ_2 is the slope of the continuum at this point, θ_3 is the flux of the emission line relative to the continuum, θ_4 is the wavelength of the emission line, and θ_5 is the width of the spectral feature. $L(x)$ is the line profile which can be Gaussian or Lorentzian. Using this expression and minimising the errors should give a more accurate peak wavelength for the feature, and therefore a more accurate redshift. The equation to be minimised is:

$$\chi^2(\theta) = \sum_{i=1}^{n} \left(f_i - f(\lambda_i|\theta)\right)^2 w_i \tag{3}$$

where f_i is the flux of the spectrum, $f(\lambda_i|\theta)$ the model from equation (2), and w_i the inverse variance, equal to σ^{-2}.

2 Methods

The spectrum from data set 2 was used. A section of the spectrum was chosen that centred on the Mg II emission line and had a width of 1000 Å, containing a sample of the background continuum of the spectrum. A region of the continuum free of other spectral features was required to obtain accurate results. The initial guesses for the minimising function were $\theta_1 = 30$, $\theta_2 = 0$, $\theta_3 = 15$, $\theta_4 = 5200$, and $\theta_5 = 50$. For the first test the line profile was chosen to be Gaussian. `scipy.optimize.minimize` was used to calculate values of $\hat{\theta}$ that minimised the function. The redshift $\hat{z}$ was then calculated using equation (1), with the returned value of $\hat{\theta}_4$ as λ_{obs}, and 2800.3 Å as λ_{lab}.

Monte Carlo simulation was used to estimate the uncertainty. 100 random data set were generated using:

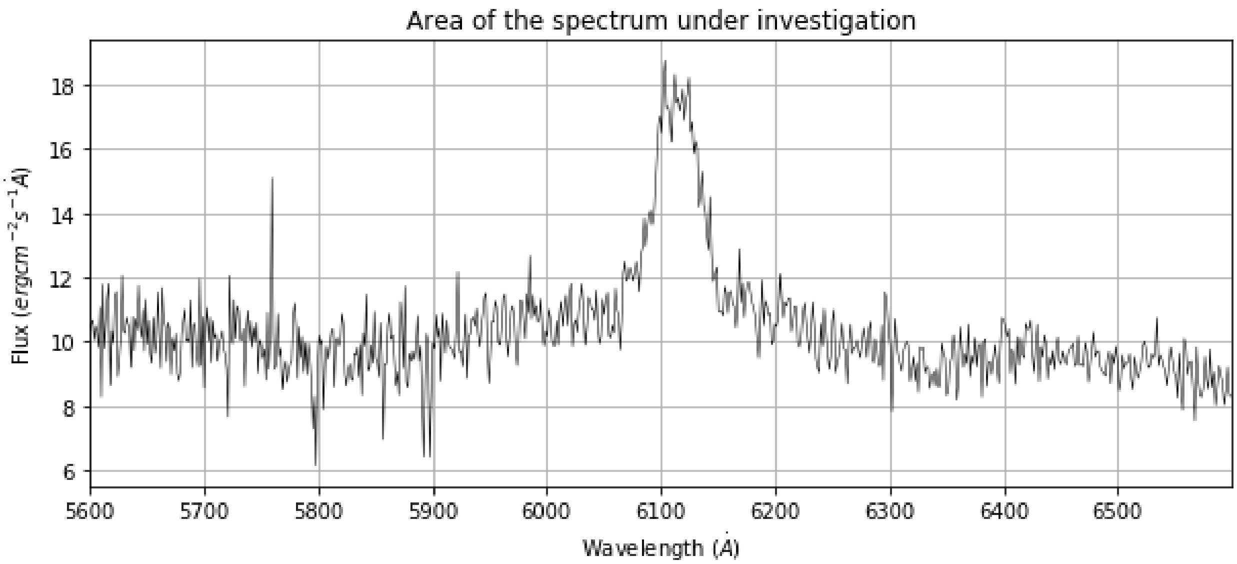

Figure 2: Plot of a section of the given spectrum of width 1000 Å centred on the Mg II emission line.

$$f_i^{(synt)} = f(\lambda_i|\hat{\theta}) + \sigma_i \times N(0,1) \tag{4}$$

where $\sigma_i = \frac{1}{w}$, and $N(0,1)$ is a random number between 0 and 1 drawn from a normal distribution. Each of these random sets of data were then used to minimise equation (3) in turn, resulting in 100 different sets of values for each of $\hat{\theta}$, and then using equation (1), 100 different values for $\hat{z}$. The values of $\hat{z}$ were then used to calculate the sample standard deviation, and this was then taken as the uncertainty of $\hat{z}$. The entire process was then repeated using a Lorentzian line profile.

3 Results

A fitted line generated from a Gaussian line profile is shown in Figure 3, using nine different methods provided by `scipy.optimize.minimize`. All the different methods provide reasonable fits except for `COBYLA`, which doesn't reach the full strength of the peak flux. `TNC` also falls slightly short of the full height of the peak flux when inspected closely. The residuals are shown in Figure 5. The plots generated from a Lorentzian line profile are shown in Figure 4. Overall the Lorentzian plots seem to show a better fit to the data. In particular they show a more gradual transition from continuum to peak than the Gaussian plot, although it is not clear that this would affect the location of the peak, the main object of interest here. All methods seem to work equally effectively, with the exceptions this time of `COBYLA` and `L-BFGS-B`, with `COBYLA` producing in this case even more pronounced deviations from the

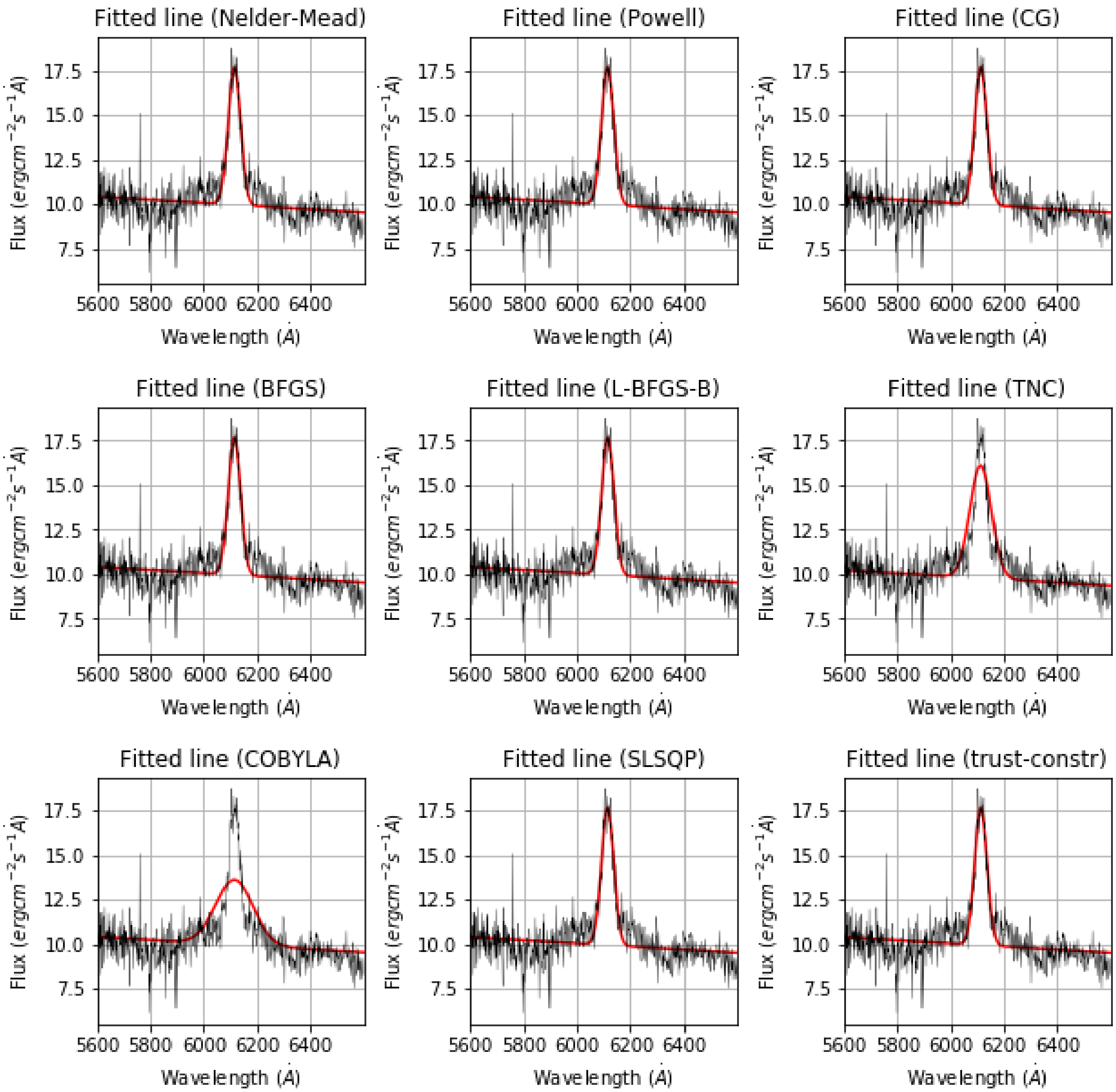

Figure 3: Spectrum of galaxy overlaid with fitted line with values generated from Gaussian line profile. Fitted line shown in red.

original data. For both the calculations using the Gaussian line profile and the Lorentzian line profile several of the methods provided good fits, however only `Nelder-Mead` was used for the Monte Carlo simulation of the synthetic data as it was one of the fastest, and it seemed unnecessary to repeat those more intensive computations with all the different methods.

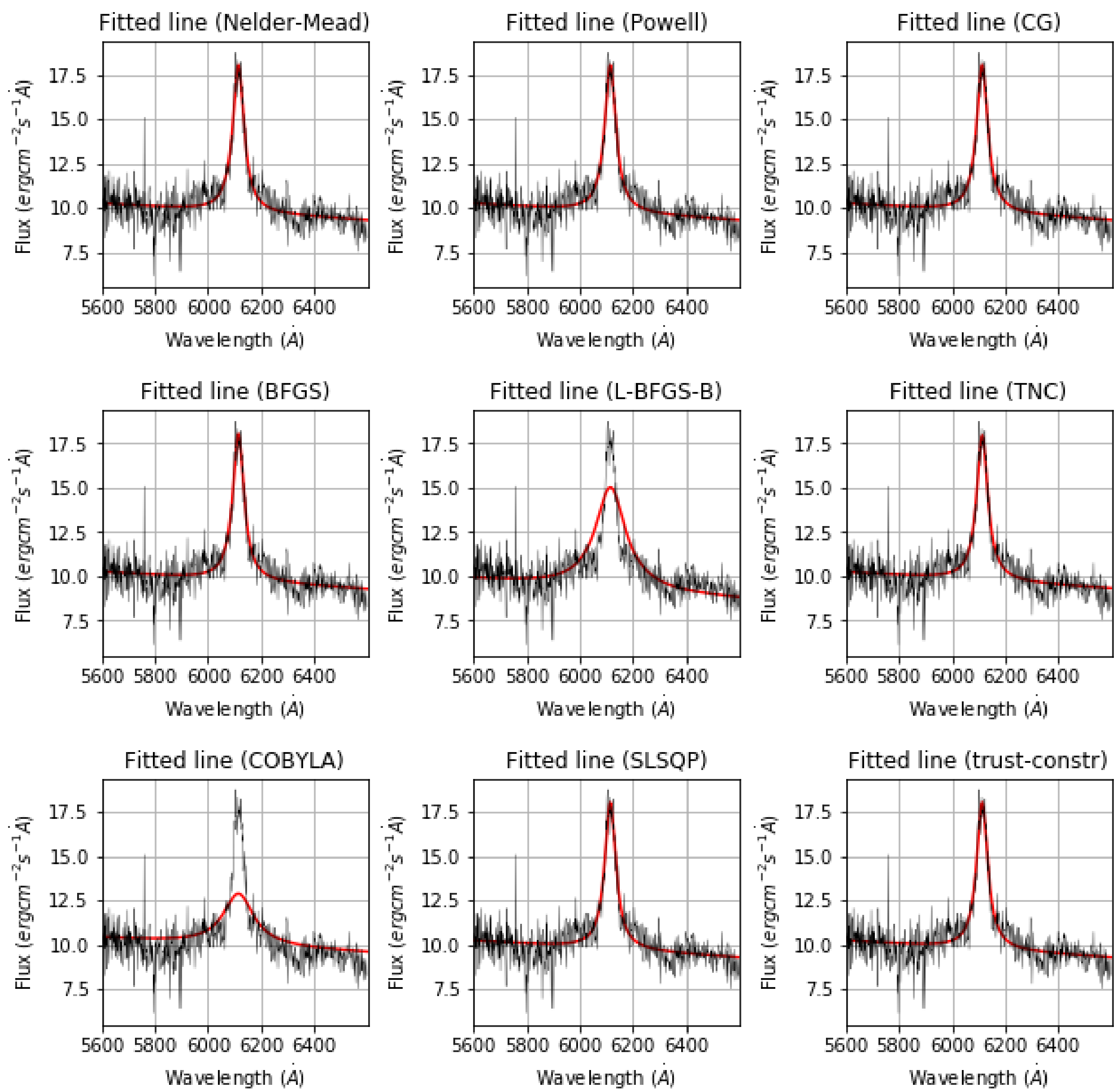

Figure 4: Spectrum of galaxy overlaid with fitted line with values generated from Lorentzian line profile. Fitted line shown in red.

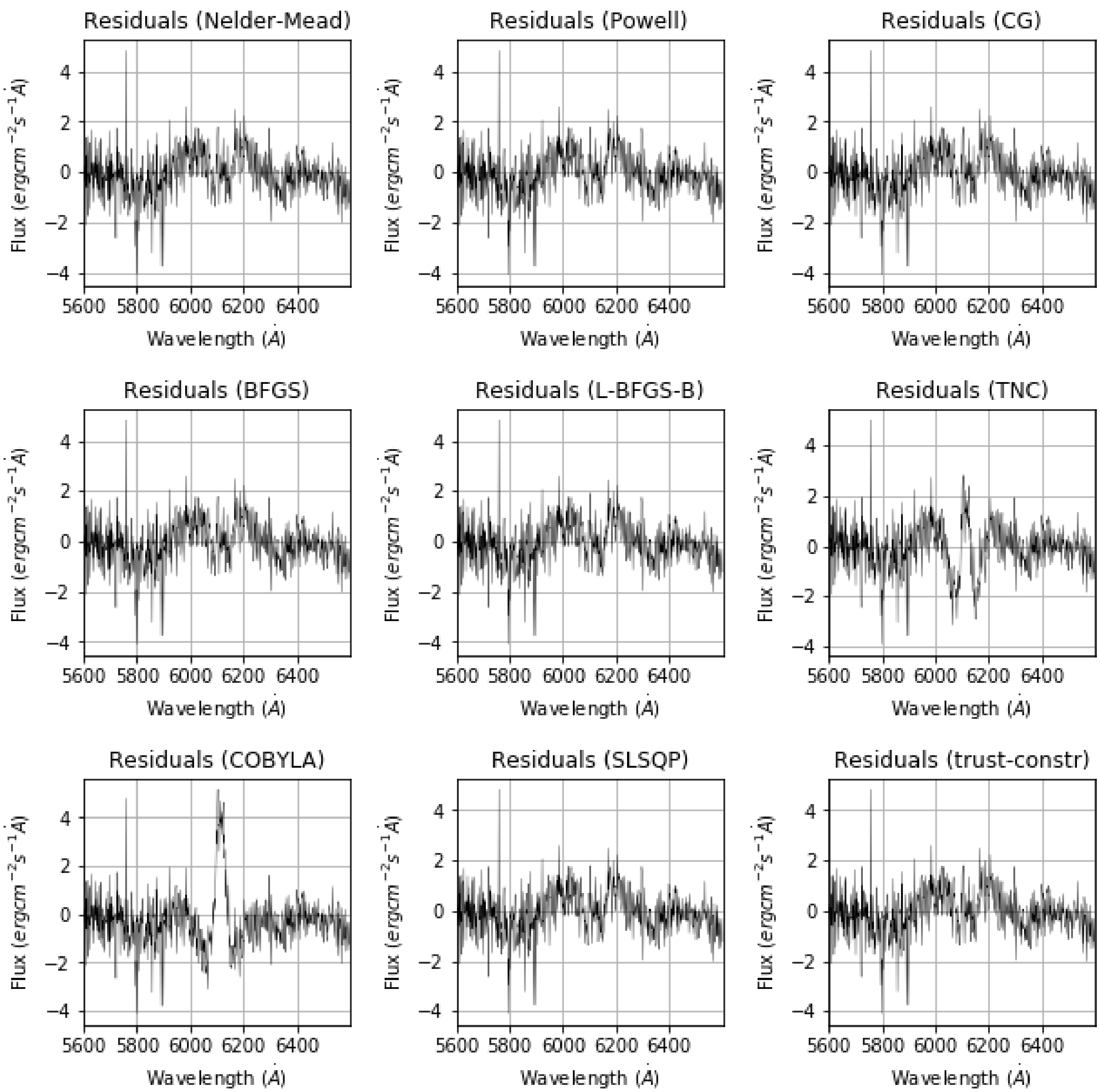

Figure 5: Residuals from Gaussian fit showing that all methods except `TNC` and `COBLYA` provide reasonably good fits to the data.

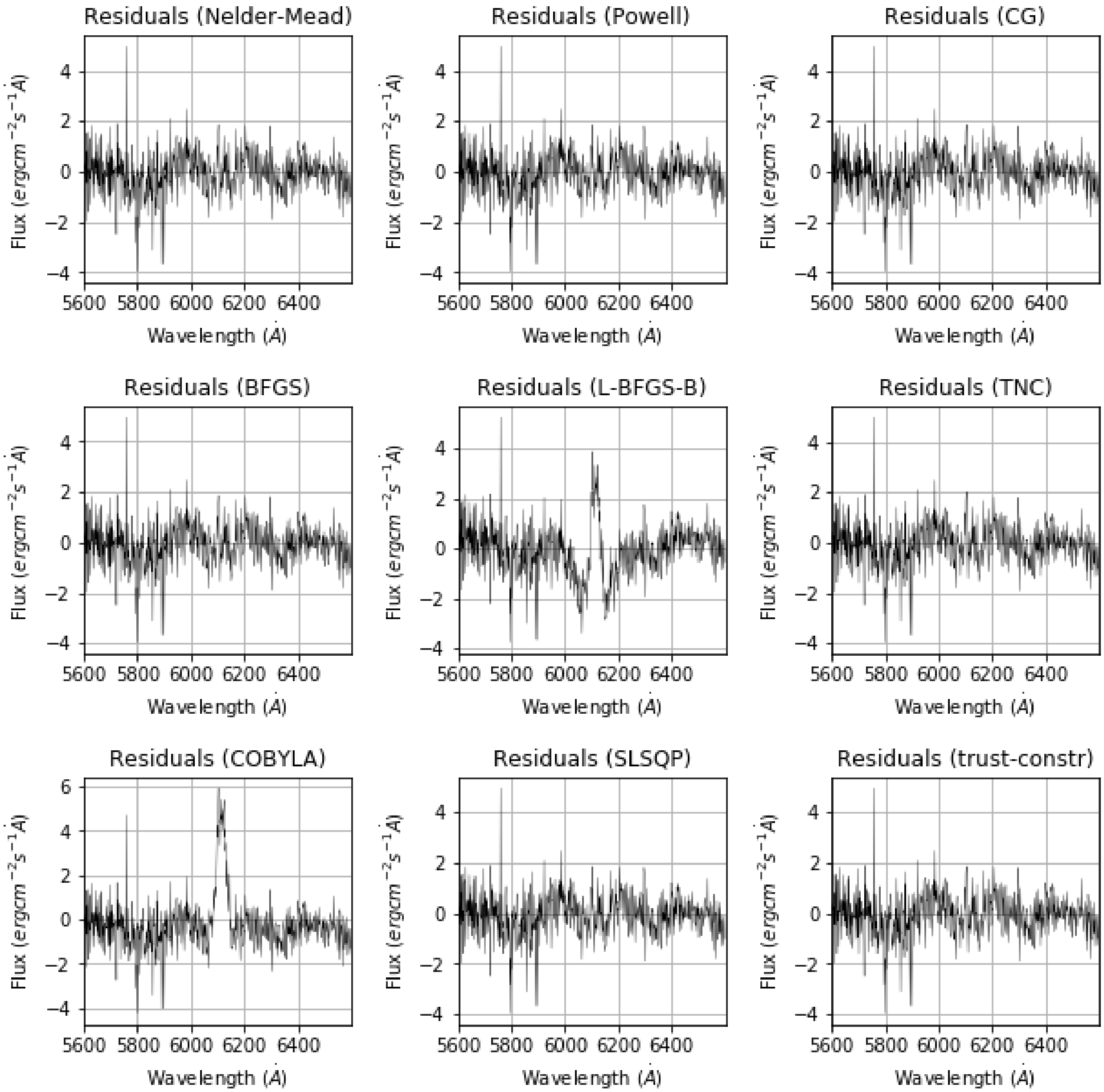

Figure 6: Residuals from Lorentzian fit showing that all methods except `L-BFGS-B` and `COBLYA` provide reasonably good fits to the data.

	Reduced chi-square (Gaussian)	Reduced chi-square (Lorentzian)
Nelder-Mead	1.6097091177	1.3665029262
Powell	1.6097091177	1.3665029264
CG	1.6097858612	1.3667562675
BFGS	1.6097091177	1.3665029262
L-BFGS-B	1.6444046435	1.4966701874
TNC	1.6097097066	1.3807541414
COBYLA	2.839767724	3.2200697441
SLSQP	1.6097091179	1.3665029262
trust-constr	1.6097091178	1.3665029262

Table 1: Reduced chi-square for both the Gaussian and Lorentzian line profiles and nine different methods provided by `scipy.optimize.minimize`.

	Redshift (Gaussian)	Redshift (Lorentzian)
Nelder-Mead	1.1833831565	1.1834270863
Powell	1.1833831567	1.1834270801
CG	1.183384168	1.1834276401
BFGS	1.1833831576	1.183627086
L-BFGS-B	1.1824810598	1.18165433
TNC	1.1833833806	1.1834058798
COBYLA	1.1836350877	3.1836350349
SLSQP	1.1833831789	1.1834270905
trust-constr	1.1833831414	1.1834270691

Table 2: Calculated redshift for both the Gaussian and Lorentzian line profiles and nine different methods provided by `scipy.optimize.minimize`.

The sample standard deviation of the synthetic redshifts was 2.8533346×10^{-5} for the Gaussian line profile, and 1.9798639×10^{-5} for the Lorentzian line profile, both found utilizing the `Nelder-Mead` method of minimization.

4 Discussion

Comparing the plots from the Gaussian line profile and those from the Lorentzian line profile, it seems as though the Lorentzian provides better fits to the data. The sample standard deviation found using the Lorentzian is likewise smaller than that from the Gaussian, showing that it does indeed provide a more precise value for the redshift. When the Monte Carlo simulation was repeated several times for the Lorentzian with 1000 synthetic data sets each time, the sample standard deviation was always close to 2×10^{-5}. In Table 2 it can be seen that the redshift varies slightly between the two line profiles, being slightly larger for the Lorentzian for most of the optimization methods that provided good fits to the data. However, these differences are small, and when used in cosmological calculations are unlikely to have a large affect.