Optimal 1D Ly α forest power spectrum estimation – I. DESI-lite spectra

Naim Göksel Karaçaylı [ID],[1]★ Andreu Font-Ribera[2,3] and Nikhil Padmanabhan [ID][1,4]

[1]*Department of Physics, Yale University, New Haven, CT 06511, USA*
[2]*Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*
[3]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, E-08193 Bellaterra (Barcelona), Spain*
[4]*Department of Astronomy, Yale University, New Haven, CT 06511, USA*

ABSTRACT

The 1D Ly α forest flux power spectrum P_{1D} is sensitive to scales smaller than a typical galaxy survey, and hence ties to the intergalactic medium's thermal state, suppression from neutrino masses, and new dark matter models. It has emerged as a competitive framework to study new physics, but also has come with various challenges and systematic errors in analysis. In this work, we revisit the optimal quadratic estimator for P_{1D}, which is robust against the relevant problems such as pixel masking, time evolution within spectrum, and quasar continuum errors. We further improve the estimator by introducing a fiducial power spectrum, which enables us to extract more information by alleviating the discreteness of band powers. We meticulously apply our method to synthetic Dark Energy Spectroscopic Instrument (DESI) spectra and demonstrate how the estimator overcomes each challenge. We further apply an optimization scheme that approximates the Fisher matrix to three elements per row and reduces computation time by 60 per cent. We show that we can achieve per cent precision in P_{1D} with 5-yr DESI data in the absence of systematics and provide forecasts for different spectral qualities.

Key words: methods: data analysis – intergalactic medium – quasars: absorption lines.

1 INTRODUCTION

Through absorption lines in quasar spectra, the Ly α forest technique can probe matter in vast volumes far into the past and at smaller scales than galaxy surveys, that are shaped by the thermal state of the gas and reionization history of the universe (Hui & Gnedin 1997; Gnedin & Hui 1998). Connecting flux fluctuations in quasar spectra to physical parameters relies on multiple demanding steps in a typical analysis. The first step is to summarize the statistical information contained in millions of pixels across thousands of spectra using the correlation function or the power spectrum. Second, one relies on numerical simulations to relate the matter fluctuations to the neutral hydrogen that reionizes until $z \sim 6$ to obtain mock quasar spectra. The physical parameters are then mapped to the statistics using large numbers of these mocks with different parameters. The final step constrains physical parameters by performing a likelihood analysis on the observed statistics using this mapping and a prior.

The Ly α forest technique already proved to be fruitful. The Extended Baryon Oscillation Spectroscopic Survey (eBOSS) (Dawson et al. 2016) and its predecessors successfully measured baryon acoustic oscillations in large-scale 3D correlations of the Ly α forest (Slosar et al. 2011, 2013; Busca et al. 2013; Font-Ribera et al. 2014; Delubac et al. 2015; Bautista et al. 2017; du Mas des Bourboux et al. 2017; Blomqvist et al. 2019; de Sainte Agathe et al. 2019). It has also emerged as a promising tool to investigate intergalactic medium (IGM) thermal evolution (Boera et al. 2019; Walther et al. 2019), to constrain neutrino masses (Croft, Hu & Dave' 1999; Seljak et al.

2006; Palanque-Delabrouille et al. 2015a, b; Yeche et al. 2017) and to probe the nature of dark matter (Boyarsky et al. 2009; Viel et al. 2013; Baur et al. 2016; Iršič et al. 2017a; Garzilli et al. 2019).

The line-of-sight flux power spectrum of the Ly α forest has been at the frontier in new physics by being sensitive to medium to small scales. McDonald et al. (2006), Palanque-Delabrouille et al. (2013), and Chabanier et al. (2019) measured this 1D power spectrum at large to medium scales ($0.001 \text{ s km}^{-1} \leq k \leq 0.02 \text{ s km}^{-1}$) using thousands of quasar spectra, whereas Viel et al. (2013), Iršič et al. (2017b), Walther et al. (2017), Boera et al. (2019), and Day, Tytler & Kambalur (2019) pushed the measurement to smaller scales ($k \leq 0.1 \text{ s km}^{-1}$) using few but high-resolution spectra.

These recent works utilize three different methods in power spectrum estimation. Palanque-Delabrouille et al. and Chabanier et al. applies fast Fourier transforms (FFT), which require all pixels to be present, equally spaced, and have uniform noise and resolution. These conditions are rarely met in real spectra due to masking of sky emission lines, high column density absorbers (HCD), bad pixels, and sometimes metal contamination. As a result, Chabanier et al. apply up to 20 per cent corrections to their power spectrum estimates. In their small-scale measurement, Walther et al. and Day et al. use Lomb–Scargle periodogram (Lomb 1976; Scargle 1982) as this method allows for masking. However, neither method is able to weight pixels with respective noise estimates and cannot account for the time evolution within a spectrum without splitting the spectrum into multiple chunks; and both are limited by S/N and resolution in their quasar samples. Third method is the likelihood maximization. Palanque-Delabrouille et al. also implemented a direct maximization of the likelihood function, which they found sensitive to the implementation details and noise in spectra. Our

★ E-mail: naimgoksel.karacayli@yale.edu

focus in this paper is the faster and more stable quadratic maximum likelihood estimator (QMLE), which McDonald et al. applied while splitting the spectra into two chunks because of computational limitations. However, with recent developments in computing, QMLE promises its full strength by careful application to 1D power spectrum.

Our work focuses on developing and applying an improved QMLE to measure the 1D power spectrum. We meticulously apply the formalism to maximize the information extraction. Our method readily handles the redshift evolution by interpolating pixel pairs to two redshift bins. This enables us to keep the full forest and all pixel pairs. We also introduce a baseline estimate that improves the accuracy by alleviating discrete band powers. As The Dark Energy Spectroscopic Instrument (DESI; Levi et al. 2013; DESI Collaboration 2016) comes online to find abundance of quasar spectra, our improved QMLE can exploit the most information and make the best measurement.

This paper is organized under six sections. We summarize relevant general and specific formula for QMLE in Section 2. This section also briefly discusses a continuum limit to clarify what QMLE actually yields. Section 3 outlines the algorithm and further provides details on our implementation, including validation with synthetic spectra. We then move to applying our method to simulated simple DESI data in Section 4. In this section, we examine the effects of gaps and continuum errors, the advantage of fiducial power, Fisher matrix approximation, and 5-yr forecasts for DESI with different spectral qualities. Section 5 discusses the finer details of our method, such as interpretation of QMLE results, its advantages and possible problems. Finally, we summarize this work in Section 6.

2 METHOD

The optimal power spectrum estimator was extensively studied by Hamilton (1997), Tegmark, Taylor & Heavens (1997), Tegmark et al. (1998), and Seljak (1998). We first assign a Gaussian probability distribution that depends on parameters θ_α for measuring a data set x. We define the likelihood function $\mathcal{L}$ as twice the logarithm of this probability and pursue the most likely parameters $\hat{\theta}_\alpha$ for a fixed x

$$\mathcal{L}(x; \theta_\alpha) = -\ln \det \mathbf{C} - x^\mathrm{T}\mathbf{C}^{-1}x, \tag{1}$$

where $\mathbf{C} = \mathbf{C}(\theta_\alpha) \equiv \langle xx^\mathrm{T} \rangle$. The most likely parameters $\hat{\theta}$ can be found by maximizing this likelihood function: $\mathcal{L}_{,\alpha}(\hat{\theta}) = 0$, where comma represents a partial derivative, which can be iteratively solved using the Newton–Raphson method:

$$\hat{\theta}_\alpha^{(X+1)} = \hat{\theta}_\alpha^{(X)} - \sum_{\alpha'} \left\langle \mathcal{L}_{,\alpha\alpha'} \right\rangle^{-1} \Big|_{\hat{\theta}^{(X)}} \mathcal{L}_{,\alpha'}(\hat{\theta}^{(X)}), \tag{2}$$

where X is the iteration number. Note that we compute the ensemble average of the second derivative (which is the Fisher matrix) instead of using the full curvature matrix as has been the convention.

Our goal is to estimate the power spectrum of the observed spectra, so we take θ_α to be the power spectrum estimates. Furthermore, we would like to estimate deviations from a fiducial power spectrum such that $P(k, z) = P_\mathrm{fid}(k, z) + \sum_\alpha w_\alpha(k, z)\theta_\alpha$, where $w_\alpha(k, z)$ are the functional forms for deviations and θ_α are the amplitudes (Font-Ribera, McDonald & Slosar 2018). Then, the covariance matrix is the sum of signal and noise as usual, $\mathbf{C} = \mathbf{S} + \mathbf{N}$; and by extension $\mathbf{S} = \mathbf{S}_\mathrm{fid} + \sum_\alpha \mathbf{Q}_\alpha \theta_\alpha$, where $\mathbf{Q}_\alpha = \partial \mathbf{C}/\partial \theta_\alpha$. The fiducial power needs to be subtracted from the estimate. We denote its contribution by t_α

below, which can be immediately calculated by substituting $\mathbf{N} \rightarrow \mathbf{N} + \mathbf{S}_\mathrm{fid}$ to the quadratic estimator equation in the references.

$$\hat{\theta}_\alpha^{(X+1)} = \sum_{\alpha'} \frac{1}{2} F_{\alpha\alpha'}^{-1}(d_{\alpha'} - b_{\alpha'} - t_{\alpha'}), \tag{3}$$

where

$$d_\alpha = x^\mathrm{T}\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}x, \tag{4}$$

$$b_\alpha = \mathrm{Tr}(\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}\mathbf{N}), \tag{5}$$

$$t_\alpha = \mathrm{Tr}(\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}\mathbf{S}_\mathrm{fid}), \tag{6}$$

and the estimated Fisher matrix is

$$F_{\alpha\alpha'} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_\alpha\partial\theta_{\alpha'}} \right\rangle = \frac{1}{2}\mathrm{Tr}(\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}\mathbf{Q}_{\alpha'}). \tag{7}$$

The covariance matrices in the right-hand side of equation (3) are computed using parameters from the previous iteration $\theta_\alpha^{(X)}$.

2.1 Ly α forest specifics

In the Ly α forest analysis, our data set x is a collection of pixels representing the normalized flux fluctuations δ_F. Assuming different quasar spectra are uncorrelated, the covariance matrix becomes block diagonal, where only the correlations within a spectrum are non-zero.[1] For example, stacking three quasar spectra $\delta_{1,2,3}$ yields

$$\delta_F = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{pmatrix} \rightarrow \mathbf{C} = \begin{pmatrix} \mathbf{C}_1 & 0 & 0 \\ 0 & \mathbf{C}_2 & 0 \\ 0 & 0 & \mathbf{C}_3 \end{pmatrix}. \tag{8}$$

This block diagonal structure simplifies equation (3) as well: The Fisher matrix $F_{\alpha\alpha'}$ and the expression in parentheses can be computed for each quasar, then accumulated, i.e. $\mathbf{F} = \sum_q \mathbf{F}_q$, etc.

The correlation (signal) between pixels depends on their velocity separation, underlying power spectrum, and spectrograph window function. We convert a pixel's wavelength to velocity using logarithmic spacing.

$$v_i = c\ln(\lambda_i/\bar{\lambda}) \tag{9}$$

$$z_i = (1 + \bar{z})e^{v_i/c} - 1, \tag{10}$$

where $\bar{\lambda}$ and $\bar{z}$ are the median wavelength and the median redshift of the spectrum, respectively.[2] In general, the signal is multiplied with a resolution matrix $\mathbf{R}$, such that $\tilde{s} = \mathbf{R}s$ and therefore $\tilde{\mathbf{S}} = \mathbf{R}\mathbf{S}\mathbf{R}^\mathrm{T}$ (Bolton & Schlegel 2010). DESI will provide this resolution matrix in its pipeline, which will be one of QMLE's strengths for future analyses. For the rest of the paper, we make the approximation that the resolution does not change with wavelength. Then, the signal becomes the correlation function convolved with the spectrograph resolution, which is the power spectrum multiplied with the spectrograph window function $W(k)$ in Fourier space.

$$S_{ij}^\mathrm{fid} = \int_0^\infty \frac{\mathrm{d}k}{\pi} \cos(kv_{ij})W^2(k)P_\mathrm{fid}(k, z_{ij}), \tag{11}$$

---

[1] The 3D analysis inherently needs these correlations between spectra, so this would not hold true.

[2] The pivot point does not matter as long as $v = c\ln\lambda$, and therefore the Ly α rest-frame wavelength can also be used instead of the median wavelength of the spectrum.

where $v_{ij} \equiv v_i - v_j$ and $1 + z_{ij} \equiv \sqrt{(1 + z_i)(1 + z_j)}$. The spectrograph window function is given by

$$W(k) = e^{-k^2 R^2/2} \text{sinc}(k \triangle v/2),\tag{12}$$

where R is the 1σ resolution and $\triangle v$ is the pixel width, both in velocity units.

We now move on to defining power spectrum-related specifics from data-related specifics. In theory, $w_\alpha(k, z)$ can be any function, but we adopt top-hat k bands with k_n as bin edges and linear interpolation for z bins with z_m as bin centres: $w_{(mn)}(k, z) = H(k - k_n)H(k_{n+1} - k)I_m(z)$, where $\alpha \equiv (mn)$, $H(x)$ is the Heaviside step function, and $I_m(z)$ is the interpolation kernel. We chose the linear interpolation for its smoothness over top hats. However, this mandates distributing pixel pairs into two redshift bins. One can imagine using higher order terms (such as cubic interpolation) to make the function smoother, and hence more accurate, but this will obviously make the calculation more complex. In the end, the linear interpolation is a good compromise between accuracy and complexity.

$$I_m(z) = \begin{cases} \frac{z - z_{m-1}}{z_m - z_{m-1}} & , z_{m-1} < z < z_m \\ \frac{z_{m+1} - z}{z_{m+1} - z_m} & , z_m < z < z_{m+1} \\ 0 & , \text{otherwise} \end{cases}\tag{13}$$

Note that this is 1 when $z = z_m$ and 0 when $z = z_{m \pm 1}$. The derivative matrix for redshift bin m and wavenumber bin n is then

$$Q_{ij}^{(mn)} = I_m(z_{ij}) \int_{k_n}^{k_{n+1}} \frac{dk}{\pi} \cos(k v_{ij}) W^2(k).\tag{14}$$

We compute these matrices for as many redshift bins as necessary for a given spectrum.

Finally, we assume that the noise of every pixel is independent. This results in a diagonal noise matrix with $N_{ii} = \sigma_i^2$, where σ_i is the pipeline noise divided by the continuum and the mean normalized flux $\bar{F}(z)$.

2.2 Continuum limit

It is sufficient to take the estimated power as $P_{est}(k_n^c, z_m) = P_{fid}(k_n^c, z_m) + \theta_{(n,m)}$ in our analysis, where k_n^c is the bin centre. However, in order to further improve our intuition, let us discuss what equation (3) constructs in detail. First, we should differentiate between the underlying true power P_{true} of data and the fiducial power P_{fid} of the estimator; these two are not necessarily the same. For simplicity, let us ignore redshift dependence, spectrograph resolution and noise, and adopt band powers for $w_n(k) = H(k - k_n)H(k_{n+1} - k)$. In the continuum limit, matrix multiplications can be converted into integrals. Then, equation (3) becomes the following at iteration X:

$$\theta_n^{(X+1)} = \int_{k_n}^{k_{n+1}} dk \, \gamma_n^{(X)}(k)[P_{true}(k) - P_{fid}(k)],\tag{15}$$

$$\gamma_n^{(X)}(k) = \frac{1}{P_{(X)}^2(k)} \left[\int_{k_n}^{k_{n+1}} \frac{dk}{P_{(X)}^2(k)} \right]^{-1},\tag{16}$$

where $P_{(X)}$ is exactly the fiducial power at the first iteration, and approaches to the true underlying power P_{true} by the last iteration. Thus, this estimator gives us an inverse variance weighted average of the residuals at each iteration. As these residuals gets smaller, the effect of averaging gets smaller as well.

3 IMPLEMENTATION

QMLE implementation presents challenges in memory, CPU time, numerical stability, and confident validation of the results. In this section, we clarify our implementation decisions in order to overcome such challenges.

We will refer Palanque-Delabrouille et al. (2013) as PD13, Walther et al. (2017) as W17, and McDonald et al. (2006) as M06 from now on.

3.1 Algorithm

At every iteration, the algorithm for each spectrum is as follows:

(i) Compute the covariance matrix using the previous iteration's $\hat{\theta}$ estimates, then invert the covariance matrix. Note that the fiducial signal matrix stays fixed.

(ii) Compute weighted data vector $y = C^{-1} \delta_F$, then $d_\alpha = y^T Q_\alpha y$.

(iii) Compute $C^{-1} Q_\alpha C^{-1}$, then b_α and t_α using equations (5) and (6).

(iv) Compute Fisher matrix using equation (7). Note that this needs to consider all redshift bins that the spectrum spans.

We then sum d_α, b_α, t_α, and $F_{\alpha\alpha'}$ of every quasar. Finally, we invert F and find $\hat{\theta}$ estimates using equation (3). We check for convergence by comparing these results to the previous iteration using the expressions in Section 3.3.

One can bootstrap QMLE results easily by saving $F_{\alpha\alpha'}$ and $(d_\alpha - b_\alpha - t_\alpha)$ of each spectrum to a file. Since each spectrum is treated independently, there is no need to recompute these for every realization. One can generate as many bootstrap realizations as needed on spectrum level, and then simply add these saved quantities with repetition to find a bootstrapped estimate. This treatment is exact at the first iteration, and should be a good approximation at convergence.

We find that using a smooth weighted spline in step (i) makes the algorithm numerically stable. This smoothing spline is performed on $(k, P = P_{fid} + \theta)$, and it can be more reliable if performed on $(\ln k, \ln P)$ while non-positive values are removed.

This algorithm can be implemented using only three matrices for each quasar: One holds the covariance matrix and its inverse, and the other two temporarily hold the derivative and fiducial signal matrices. Holding all matrices in memory for a quasar at a time can improve computation time when memory is the lesser concern, which we find possible in our tests.

This method demands substantial CPU time, and therefore some optimization mechanisms are noteworthy. First, instead of integrating S_{fid} and Q_α when needed, we create lookup tables once and interpolate. Moreover, every spectrum can be computed in parallel as they are assumed independent. Each CPU should have near-equal workload for efficient parallelization. Matrix operations scale as $\mathcal{O}(N_D^3)$, where N_D is the number of pixels in a spectrum. Let us define N_B as the number of total bins to which a spectrum contributes. Then, the Fisher matrix calculation will require $N_B(N_B + 1)/2$ matrix multiplications for that spectrum. Hence, we use $T_{cpu} = N_D^3 N_B(N_B + 1)$ as an estimate for computation time, and distribute spectra accordingly.

We use GSL[3] for interpolation, integration, matrix inversion; Intel's MKL library[4] for matrix multiplication; and we compile with

---

[3] https://www.gnu.org/software/gsl/
[4] https://software.intel.com/en-us/mkl

Table 1. Top: PD13 BOSS likelihood fitting parameters. Bottom: This work fitting combined PD13 and W17 data with Lorentzian decay added.

A	n	α	B	β	k_1 (km s^{-1})
0.06	-2.55	-0.10	3.55	-0.28	–
0.066	-2.685	-0.22	3.59	-0.18	0.53

OPEN-MPI[5] for parallelization. Smooth weighted spline is constructed using SCIPY.[6]

3.2 Fiducial power spectrum

We exploit a baseline estimate of power spectrum in our analysis as discussed in Section 2. PD13 provide a fitting function with best-fitting parameters. We further modify their fitting function with a Lorentzian decay:

$$\frac{k\,P(k,z)}{\pi} = A\,\frac{(k/k_0)^{3+n+\alpha\ln k/k_0}}{1+(k/k_1)^2}\left(\frac{1+z}{1+z_0}\right)^{B+\beta\ln k/k_0}, \tag{17}$$

where $k_0 = 0.009$ km s^{-1} and $z_0 = 3.0$. However, PD13 measures power spectrum up to 0.02 km s^{-1}, so their parameters are not valid on small scales. We combine W17's power spectrum estimates, and fit the resulting data set. Our modification reduces χ_ν^2 from 16.5 to 5.6. Even though the fit should not be used for scientific purposes, it should be sufficient for a baseline estimate. Table 1 shows our fitting parameters.

Our implementation is also equipped with taking a tabulated fiducial power as input. This feature gives greater freedom in the choice of fiducial. We use this to eliminate any discord between synthetic data and the estimator; and to investigate different choices as fiducial.

3.3 Convergence

We choose a convergence criterion that summarizes the overall fluctuations between iterations. For this purpose, we calculate the weighted average of the changes between iterations using estimated Gaussian errors. In other words, we define convergence when

$$\triangle\chi = \sqrt{\frac{1}{N}\sum_\alpha \frac{(\triangle\hat{\theta}_\alpha)^2}{F_{\alpha\alpha}^{-1}}} \tag{18}$$

or

$$\triangle\chi_F = \sqrt{\frac{1}{N}(\boldsymbol{\Delta\hat{\theta}})^{\mathrm{T}}\mathbf{F}(\boldsymbol{\Delta\hat{\theta}})} \tag{19}$$

becomes smaller than $\chi_c = 0.01$, where $\triangle\hat{\theta}_\alpha \equiv \hat{\theta}_\alpha^{(X+1)} - \hat{\theta}_\alpha^{(X)}$ and N is the total number of bins. Both expressions return close values and reach convergence at the same iteration in almost all cases.

3.4 Validation

We use lognormal mocks for validation. These semirealistic spectra are crucial to examine the accuracy, precision, and efficiency of our method.

We generate lognormal mocks by using a modified version of M06. These realizations approximately produce theoretically expected

mean flux redshift evolution (Faucher-Giguére et al. 2008; Becker et al. 2013) and power spectra similar to PD13 and W17.

(i) Generate a long high-resolution Gaussian random grid with equal spacing in velocity v, zero mean, and unit variance.

(ii) FFT this grid and multiply with $\sqrt{P(k)/\mathrm{d}v}$ to obtain $\tilde{\delta}_b(k)$, where $\mathrm{d}v$ is the grid spacing in velocity units and the power spectrum is

$$P(k) = \frac{(k/k_0)^{n-\alpha\ln(k/k_0)}}{1+(k/k_1)^\gamma}, \tag{20}$$

where $k_0 = 0.001$ s km^{-1}, $k_1 = 0.04$ s km^{-1}, $n = 0.5$, $\alpha = 0.26$, and $\gamma = 1.8$. Inverse FFT and save the variance of this grid σ^2. This is a crude Gaussian base for baryon fluctuations $\delta_b(v)$ with defined power spectrum at $z_0 = 3$.

(iii) Multiply with a redshift evolution factor $a(z)$. Such that $\delta_b(z) = a(z)\delta_b$ and $\sigma^2(z) = a^2(z)\sigma^2$.

$$a^2(z) = 58.6\left(\frac{1+z}{1+z_0}\right)^{-2.82} \tag{21}$$

(iv) Apply a squared lognormal transformation to approximate the non-linear and non-Gaussian H I column density field.

$$n(z) = \mathrm{e}^{2\delta_b(z)-\sigma^2(z)} \tag{22}$$

(v) Transform this to optical depth τ by multiplying with another redshift-dependent function.

$$\tau(z) = 0.55\left(\frac{1+z}{1+z_0}\right)^{5.1}n(z) \tag{23}$$

(vi) Finally, the flux is $F(z) = \mathrm{e}^{-\tau(z)}$.

(vii) Smoothing $F(z)$ with a Gaussian kernel and resampling it on to the observed wavelength grid will result in a spectrograph function in equation (12).

The mean flux and power spectrum of these mocks can be analytically computed. Using one-point probability of the base Gaussian random field δ, we can write down the following integral for the mean flux:

$$\bar{F}(z) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty \exp\left[-\frac{\delta^2}{2\sigma^2} - x(z)\mathrm{e}^{2a(z)\delta}\right]\mathrm{d}\delta, \tag{24}$$

where we have defined

$$x(z) \equiv 0.55\left(\frac{1+z}{1+z_0}\right)^{5.1}\mathrm{e}^{-a^2(z)\sigma^2}. \tag{25}$$

Although, we have found a closed form using saddle point integration (see Appendix A), it deviates from the truth for $z \lesssim 3$ up to 4 per cent. It is more accurate to just integrate this expression; and Gauss–Hermite quadrature provides a reliable fast integration.

Power spectrum expression needs two-point probability. We start with flux fluctuations $\delta_F(v,z) = F(v,z)/\bar{F}(z) - 1$, then express the correlation function ξ_F as an integral over two Gaussian random δs assuming all pixels are at the same redshift. We drop z for clarity.

$$1 + \xi_F(v_{ij}) = \int \frac{\mathrm{e}^{-\boldsymbol{\delta}^T\mathbf{C}^{-1}\boldsymbol{\delta}/2}}{2\pi\sqrt{\det\mathbf{C}}}\frac{F_i F_j}{\bar{F}^2}\mathrm{d}\boldsymbol{\delta}, \tag{26}$$

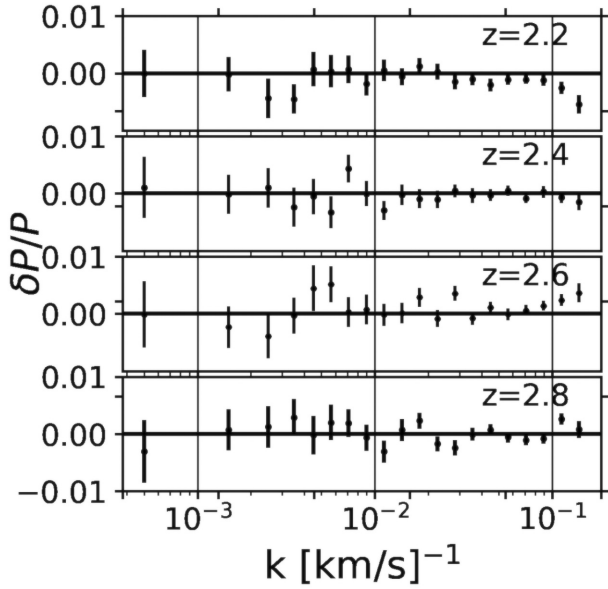where

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & \xi_G(v_{ij}) \\ \xi_G(v_{ij}) & \sigma^2 \end{pmatrix}, \tag{27}$$

and $\xi_G(v)$ is the correlation function of the base Gaussian field. To convert this expression into Gauss–Hermite quadrature with

Figure 1. The relative error in our measurements from 100 lognormal catalogues where each has 1000 spectra in every redshift bin. Our method reaches sub per cent level precision in the absence of any systematic. Bins go up to the Nyquist frequency.

two variables, we apply Cholesky decomposition $\mathbf{C} = \mathbf{L}\mathbf{L}^T$ and transform $\boldsymbol{\delta} = \sqrt{2}\mathbf{L} \cdot \mathbf{y}$, where

$$\mathbf{L} = \begin{pmatrix} \sigma & 0 \\ \frac{\xi_G(v)}{\sigma} & \sigma\sqrt{1 - \frac{\xi_G(v)^2}{\sigma^4}} \end{pmatrix}. \tag{28}$$

Here is the final expression for completeness:

$$1 + \xi_F(v) = \int e^{-y^2} \frac{1}{\pi \bar{F}^2(z)} \exp\left[-x(z)\left(e^{2a\delta_1} + e^{2a\delta_2}\right)\right] d\mathbf{y}. \tag{29}$$

We perform an initial test in the absence of redshift distribution and resolution effects. We start with a long fine grid ($dv = 0.4\bar{3}$ km s^{-1}, $N = 2^{20}$), and generate mocks that are discretely distributed in redshift. For validation, we pick four redshifts (2.2, 2.4, 2.6, 2.8), and simulate 100 catalogues where each catalogue has 1000 spectra in every redshift bin. These spectra are exactly centred at these redshifts. We set the spectrograph resolution $R = 71\,600$ to diminish its effect, limit the spectral length to half of the bin size $\Delta z = 0.1$, and finally resample to pixel size of $\Delta v = 20.8$ km s^{-1}. Fig. 1 shows our results. Our method reaches sub per cent level accuracy in the absence of any systematic.

4 DESI-LITE SPECTRA

To test the feasibility of our method for future DESI spectra, we generate semirealistic data set using the following simplifying assumptions:

(i) We set the observed wavelength grid between 3600 and 9800 Å. This means the closest forest pixel is at $z = 1.96$.

(ii) We create a logarithmically spaced wavelength grid with $\Delta v = 30$ km s^{-1}.[7] Our grid spacing corresponds to a Nyquist frequency of 0.1 s km^{-1}.

---

[7]DESI will use linear wavelength spacing in its pipeline, but QMLE does not require equal velocity spacing.

(iii) We assume a constant resolution power of $R = 3200$ (≈ 94 km s^{-1} FWHM in velocity units) for all spectra at all wavelengths.

(iv) We add Gaussian random errors with $\sigma = 0.7$ to F, which is approximately S/N = 2 Å^{-1}.

(v) We set the minimum redshift of an Ly α quasar to 2.1 and the maximum to 4.4, and pick random redshifts from a distribution $n(z)$ (DESI Collaboration 2016; Palanque-Delabrouille et al. 2016). We always limit the forest to [1050 Å, 1180 Å] range in quasar's rest frame.

(vi) Finally, we assume DESI will observe 800 000 quasar spectra.

We choose 12 redshift bins between 2.0 and 4.2 with $\Delta z = 0.2$. First 5 k bins are linearly spaced with $\Delta k_{\mathrm{lin}} = 0.001$ s km^{-1}, and the following 14 bins are logarithmically spaced with $\Delta k_{\mathrm{log}} = 0.1$. Hence, 0.0005 s km$^{-1} \leq k \leq 0.112$ s km^{-1}.

Fig. 2 shows power spectrum estimates for each redshift bin with error bars from the diagonal elements of the inverse Fisher matrix. The chi-square using the full Fisher matrix is $\chi^2/\nu = 248/228$,[8] which implies a valid agreement with the truth. We measure the power spectrum to sub per cent accuracy at lower redshifts, but the accuracy and precision get progressively worse towards high redshifts due to declining quasar numbers. We also expected noise and window function corrections to dominate at high k as the noise power crosses the signal at $k_N \approx 0.025$ s km^{-1}.[9] This constitutes our foundation as validation of our method.

We would like to stress a subtle point here. First of all, the estimator constructs correct covariance matrices in the first iteration by using the true power as fiducial input. This means we also have the correct Fisher matrix and expect $\theta \approx 0$ after the iteration. However, the convergence criteria in equation (18) will still yield nearly 1 due to the statistical fluctuations of the power spectrum (note Var$[\theta_\alpha] \sim F_{\alpha\alpha}^{-1}$). Therefore, when the estimator goes into the second iteration, it misidentifies these intrinsic fluctuations as corrections to the covariance matrices and readjusts the Fisher matrix. The smoothing spline ameliorates this digression, though imperfectly.

Fisher matrix calculation consumes the most time as it requires calculating $\mathbf{Q}_\alpha$ matrices and multiplying them $\mathcal{O}(N^2)$ many times. We identified that at least 46 per cent of the total time goes into the Fisher matrix (see Fig. 3). Moreover, we also found that the Fisher matrix is a band matrix that is prominently tridiagonal for individual k and z bins. We normalize the Fisher matrix with respect to its diagonal elements for a clear representation in Fig. 4, which is for $k = 0.0071$ s km^{-1} bin, but represents a typical redshift dependence.[10] This tridiagonal shape is due to keeping the full forest and distributing pixel pairs into two redshift bins. This structure weakly persists between k bins for a given z bin as well. Given Fisher matrix is the longest step, limiting its calculation to only these terms will speed up the estimation significantly by decreasing the number of matrix multiplications to $\mathcal{O}(N)$. We consider this optimization scheme and how well it performs in Section 4.4. It is also worth pointing out that the Fisher matrix does not depend on the data, but only on the input power spectrum. Therefore, one could also choose a common Fisher matrix for analysis with many simulations.

---

[8]We performed nine independent runs and found χ^2 fluctuating around 228 as expected. The value 248 is from our first run, and not a special case.

[9]This k_N roughly corresponds to $P_N/\sqrt{N_{\mathrm{qso}}} = P_{1D}$, where the noise power is $P_N = \left(\sigma \Delta v/\bar{F}(z)\right)^2$ with $N_{\mathrm{qso}} = 100\,000$ for all redshift bins and P_{1D} is obtained analytically.

[10]The lowest k bin is weakly coupled (~ 1 per cent) to an additional redshift bin.
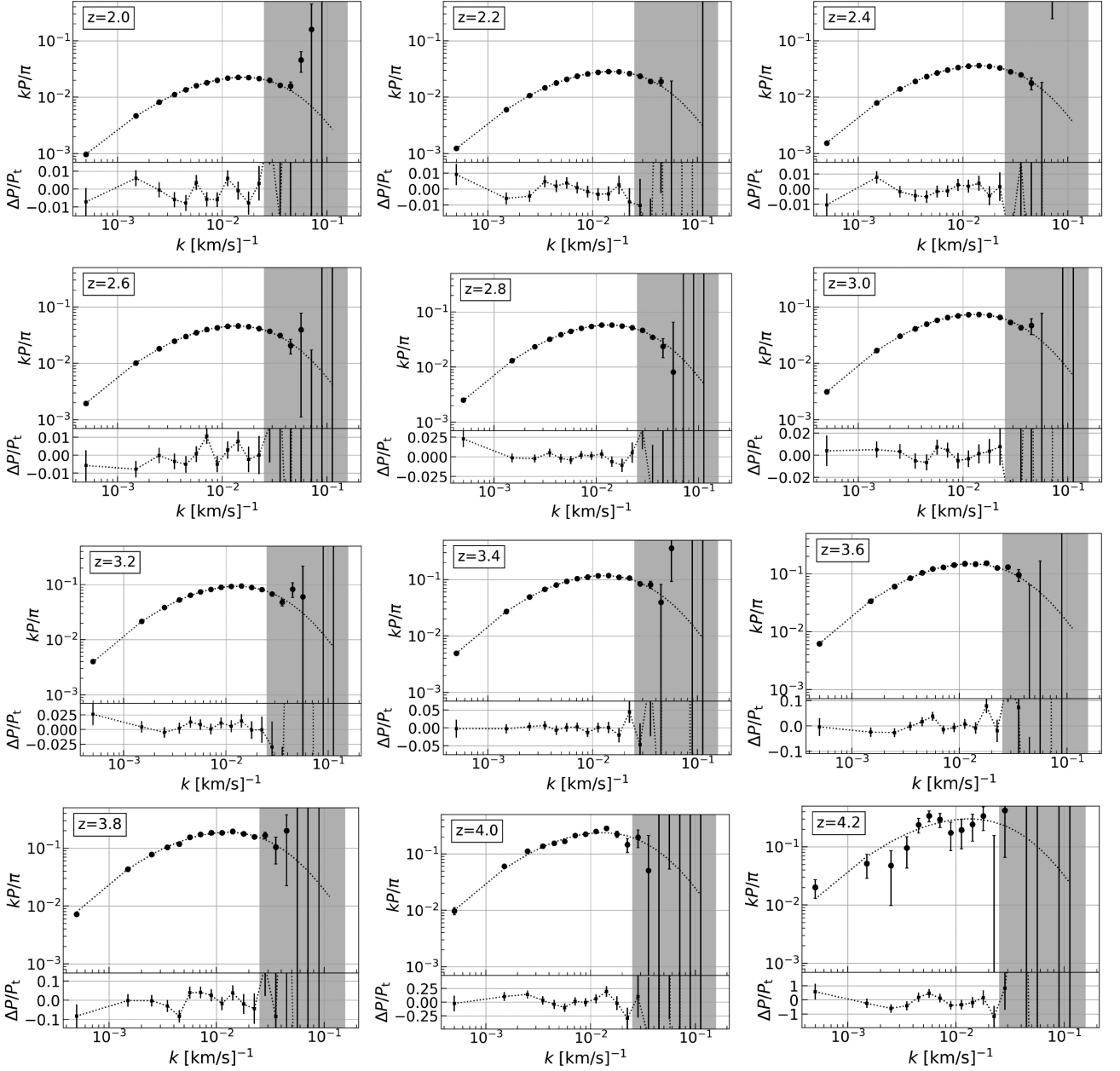
Figure 2. Power spectrum estimates from 800 000 DESI-lite spectra. Error bars from the diagonal of the inverse Fisher matrix. Dotted line in each upper panel is the true power. Noise begins to dominate in grey shaded area, $k > k_N = 0.025$ s km^{-1}. The estimates reach per cent-level accuracy in lower redshift bins. However, last redshift bins are poorly estimated due to significant decline in statistics. The χ^2 calculated from diagonal (all) elements is 240 (248) for 228 degrees of freedom.

We also computed the power spectrum covariance matrix with 5000 bootstrap realizations using the results of the first iteration as discussed in Section 3.1. We found that the diagonals of the covariance matrix estimated from this bootstrap procedure agreed within per cent level with the formal estimates from the Fisher matrix except for the high-redshift bin where we do not have enough quasars to compute robust bootstrap errors. This bootstrapping procedure therefore provides a straightforward test of the assumptions underlying the error estimates from the Fisher matrix.

4.1 Gaps in spectra

An advantage of our estimator is that it works in pixel space and can therefore robustly handle missing data in the spectrum. In order to

test how masking affects our results, we remove continuous regions from some spectra. We assign 15 per cent probability of having a high HCD in a spectrum, matching Noterdaeme et al. (2012). If a spectrum has an HCD, we randomly pick a central wavelength and mask 12.5 A$^\circ$ on each side by setting the flux to the mean value ($\delta_F = 0$). We apply this random masking procedure to 100 000 spectra. We perform a run where masked pixels are removed by assigning large errors, and another run without removing these pixels for comparison.

Fig. 5 shows our results in two redshift bins. The masking suppresses power at small scales, but propagates to all scales. The run without any correction yields extremely poor results with $\chi^2 \sim 4000$. When we correct for these pixels by assigning large errors, the power spectrum estimates yield $\chi^2 = 234$, close to the
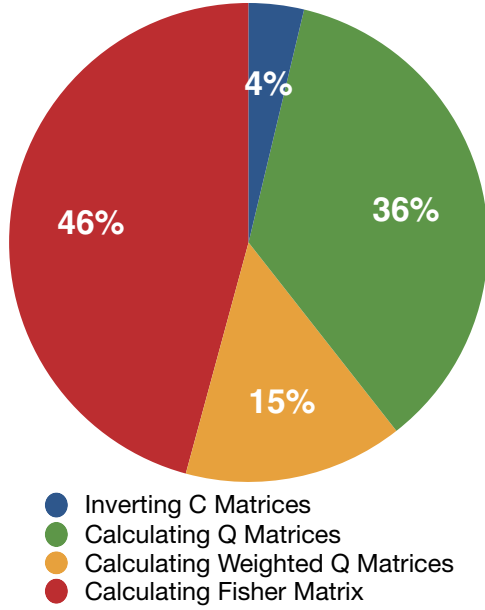
Figure 3. Percentage of time spent in different steps for 800 000 DESI-lite spectra. The weighted Q matrices are $\mathbf{C}^{-1}\mathbf{Q}_\alpha\mathbf{C}^{-1}$. Fisher matrix calculation consumes the most time, which also requires calculating $\mathbf{Q}_\alpha$ matrices as prerequisite. The time for remaining steps is insignificant.
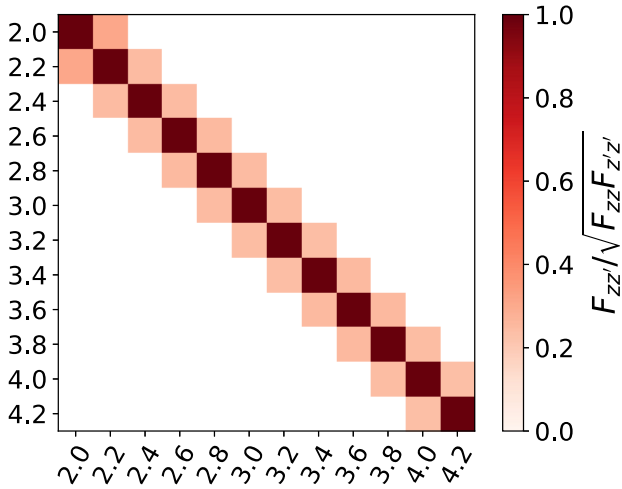

Figure 4. The normalized Fisher matrix between redshift bins for $k = 0.0071$ s km^{-1} bin. The off-diagonal terms are approximately 40 per cent. Only neighbouring redshift bins dominate the Fisher matrix, since a pixel pair is split into two bins.

original case. This confirms our method is robust against gaps in spectra.

4.2 Continuum marginalization

In a typical pipeline, the real observable flux $f(\lambda)$ is divided by the quasar continuum $C(\lambda)$ to obtain the normalized flux $F(\lambda)$ for each spectrum. Furthermore, flux fluctuations are obtained by diving this with the mean normalized flux $\bar{F}(\lambda)$. The errors in this process propagate to mostly large scales and are called the continuum errors in Ly α nomenclature.

Our quadratic estimator is armed with marginalization capability to suppress these offsets (appendix B, Slosar et al. 2013). By
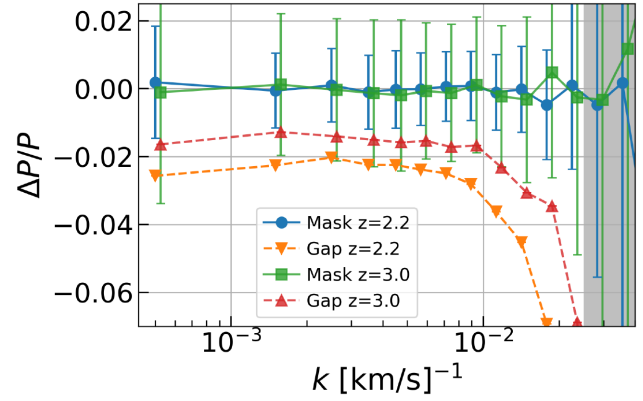
Figure 5. The effect of gaps in P_{1D} for two redshift bins $z = 2.2$ and $z = 3.0$. Points labelled 'Mask' are corrected measurements by large pixel errors, whereas 'Gap' points have unchanged errors. For clarity, we slightly shift $z = 3.0$ points and omit error bars from uncorrected results. The masking error propagates to all scales.

modifying the covariance matrix to $\mathbf{C}' = \mathbf{C} + N\mathbf{t}\mathbf{t}^{\mathrm{T}}$, where N is large and $\mathbf{t}$ is the mode we want to marginalize out, one can show that $\mathbf{C}'^{-1}(\boldsymbol{\delta}'_F + \alpha\mathbf{t}) \approx \mathbf{C}^{-1}\boldsymbol{\delta}'_F$, where the new data vector $\boldsymbol{\delta}'_F$ is orthogonal to $\mathbf{t}$. This effectively removes any information from data that is in mode $\mathbf{t}$.

We will focus our attention to the continuum fitting Slosar et al. (2013) uses. They fit the flux of each quasar q with $f_q(\lambda_o) = A_q(\lambda_o)C(\lambda_r)\bar{F}(\lambda_o)(1 + \delta_{F,q})$, where λ_o is the observed wavelength and λ_r is the rest-frame wavelength. $C(\lambda_r)$ and $\bar{F}(\lambda_o)$ are determined globally, whereas the function A_q has two free parameters for each quasar:

$$A_q(\lambda) = a_q + b_q \frac{\ln \lambda - \ln \lambda_1}{\ln \lambda_2 - \ln \lambda_1}, \tag{30}$$

where $\lambda_{1,2}$ are the beginning and the end of the forest in a given spectrum. These two quasar specific parameters source most of the error assuming the global functions are more robust.

To replicate these continuum errors, we add an error $\eta(\lambda)$ to each quasar's flux fluctuations $\delta_F(\lambda)$. We limit the form of $\eta(\lambda)$ to the equation above in order to perform a controlled test. We randomly generate the two parameters from a Gaussian distribution with $\sigma = 0.1$. We run eight independent sets of 100 000 spectra.

We add a large constant N_0 to all elements of the covariance matrix to marginalize the amplitude, and $N_1\mathbf{t}\mathbf{t}^{\mathrm{T}}$ to marginalize the slope, where $\mathbf{t} = \ln(\lambda/\lambda_{\mathrm{Ly}\alpha})$.

Fig. 6 presents the average of eight runs. The continuum errors mostly contaminate large scales. The largest scale $k = 5 \times 10^{-4}$ s km^{-1} bins show big offsets, whereas the following k bin is less affected. When we marginalize these two continuum terms with $N_0 = N_1 = 1000$, results go back to the expected values. As a side effect, marginalization increases the error in the first bin by 25–45 per cent.

4.3 Choice of fiducial

We already established in the previous sections that using the true power as fiducial yields correct results. In this section, we investigate how different fiducial power influences results by considering two additional cases. First, we run the estimator without any fiducial power. Second, imagining a realistic pipeline, we fit equation (17) to these no fiducial results, and employ this best-fitting function as
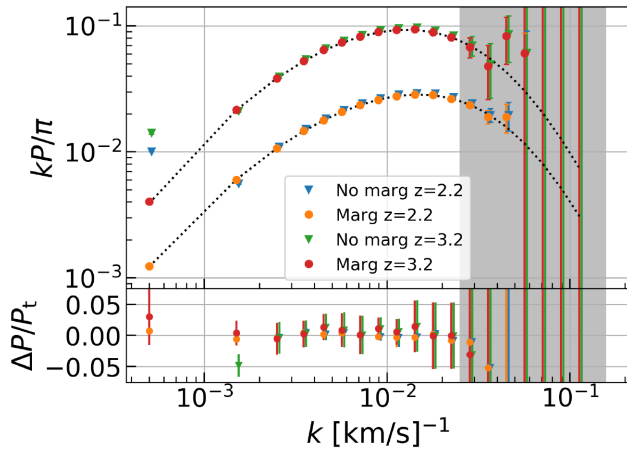
Figure 6. Average power spectra of eight runs at $z = 2.2$ and $z = 3.2$ with and without marginalizing continuum errors. The continuum errors highly contaminate the first two k bins at $k = 5 \times 10^{-4}$ s km^{-1} and $k = 1.5 \times 10^{-3}$ s km^{-1}. The remaining scales are less affected. By marginalizing two continuum modes, we are able to correctly estimate the power at these scales. Average errors of eight runs at $k = 5 \times 10^{-4}$ s km^{-1} bin increase by 25–45 per cent. The change in the second bin is insignificant.
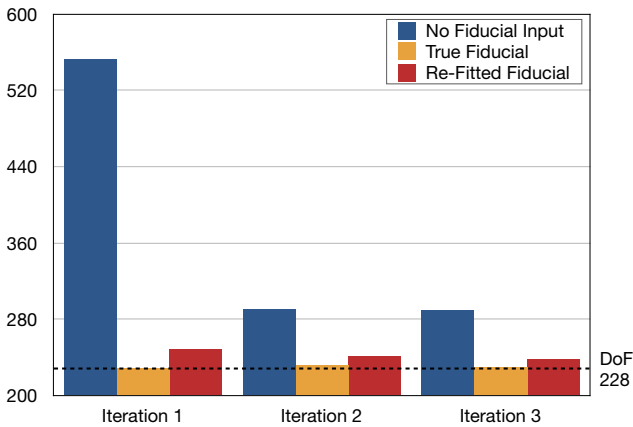
Figure 7. Comparing average χ^2 of eight runs for three different fiducials using respective Fisher matrices. True fiducial yields the correct answer in one iteration and outperforms other cases as expected. Not using any prior naturally starts away from the truth, but converges to the correct power immediately. Using the results from no prior to construct a better fiducial (re-fitted fiducial) decisively surpasses no fiducial case. Repeating this analysis with true Fisher matrix yields the same conclusion. In that case, χ^2 stays constant over three iterations with values close iteration two and three of this figure. This implies that the first iteration yields good P_{1D} and subsequent iterations correct the error estimates.

fiducial. In order to run multiple independent realizations, we use eight subsamples of 100 000 spectra.

We compare χ^2 for every case using respective Fisher matrices, which can be seen in Fig. 7. True fiducial yields the correct answer in one iteration and is the best estimate as expected. Not using any prior naturally starts away from the truth, but converges to the correct power immediately while being relatively a poor fit. Using the results from no prior to construct a better fiducial decisively outperforms no fiducial case, which we found to hold for all eight runs.

The χ^2 analysis above weights absolute errors with respective Fisher matrices. The differences in these matrices partially source the χ^2 trend in Fig. 7. In order to decouple error accuracy from P_{1D}

Table 2. Our model power spectrum and precision forecasts for 5-yr DESI survey for fixed $k_f = 2.5 \times 10^{-3}$ s km^{-1}. The precision is defined as error divided by signal: e/P_{1D}. For reference, metal contamination constitutes 5–15 per cent of the Ly α power in reality.

z	$P_{1D}(k_f)$ (km s^{-1})	Precision S/N = 1 Å^{-1}	Precision S/N = 2 Å^{-1}	Precision S/N = 4 Å^{-1}
2.0	10.2	1.31 per cent	0.44 per cent	0.22 per cent
2.2	13.4	1.08 per cent	0.37 per cent	0.20 per cent
2.4	17.5	1.12 per cent	0.40 per cent	0.23 per cent
2.6	22.8	1.24 per cent	0.46 per cent	0.27 per cent
2.8	29.5	1.46 per cent	0.56 per cent	0.34 per cent
3.0	37.8	1.87 per cent	0.74 per cent	0.45 per cent
3.2	48.3	2.50 per cent	0.99 per cent	0.62 per cent
3.4	61.4	3.34 per cent	1.32 per cent	0.82 per cent
3.6	77.6	4.82 per cent	1.89 per cent	1.15 per cent
3.8	97.7	8.67 per cent	3.32 per cent	1.98 per cent
4.0	123.0	22.00 per cent	8.14 per cent	4.63 per cent
4.2	153.0	95.30 per cent	33.30 per cent	18.10 per cent

accuracy, we also calculate all χ^2s using the true Fisher matrix. We find this quantity still distinguishes different fiducials as before with values close to the last two iterations of the initial analysis, and stays approximately flat over three iterations. This indicates that QMLE yields good P_{1D} at the first iteration even for no fiducial case, but it does not improve at subsequent iterations. Evidently, these iterations mostly correct the error estimates. This behaviour is reasonable given the weights in equation (15) stay approximately constant across bins, and therefore are unaffected by band power corrections. The only way to overcome the band power discretization is through integrating a fiducial power.

To summarize, the key points are as follows: (1) even with no fiducial, the power spectrum is correctly recovered, (2) a simple iteration quickly gets the correct Fisher matrix, and (3) a good fiducial yields better power spectrum and error estimates.

4.4 Fisher matrix approximation

Each quasar's Fisher matrix requires $N_B(N_B + 1)/2$ matrix multiplications, where N_B is the number of total bins to which this spectrum contributes. However, Fig. 4 tells us that only few elements are dominant in the Fisher matrix. Moreover, we see this structure in all cases, i.e. different fiducial powers, masking, and continuum marginalization.

While the sparsity structure of the Fisher matrix appears to be quite generic, the exact details (as to the number of non-zero elements) do appear to depend on the total signal to noise of the sample. For the DESI-like data considered in this work, we find that including a single off-diagonal element in the $k-$ and $z-$ directions works very well. Specifically, for every (k_n, z_m) pair we calculate (k_{n+1}, z_m) and (k_n, z_{m+1}) elements besides the diagonal. That reduces the number of matrix multiplications to $3N_B$, which will boost the speed. Note that this should also be taken into account in load balancing.

We test this optimization scheme on 800 000 spectra. We find that the resulting χ^2 is not changed while the average time per iteration decreases by 64 per cent and the time spent in Fisher matrix calculation per iteration drops by 95 per cent.

Even though we do not have an exact prescription for computing the number of non-zero elements, the above discussion suggests a procedure by which one starts to fill in the Fisher matrix from the diagonals outwards and stops when the overall relative contribution is below a chosen threshold.

4.5 5-yr forecasts

We would like to make simple predictions for future DESI performance using our simple spectra. We run two additional cases where S/N = 1 Å^{-1} and S/N = 4 Å^{-1} on 800 000 spectra. In order to speed up our analysis, we turn on the Fisher optimization scheme. We perform only one iteration with true power as fiducial.

We compare P_{1D} estimates for different redshifts at a fixed wavenumber $k_f = 2.5 \times 10^{-3}$ s km^{-1}. Our results are summarized in Table 2. This prediction is under ideal circumstances, hence further complications should be added for accuracy. For example, metal contamination constitutes 5–15 per cent of the Ly α forest power (Day et al. 2019).

We provide the model power spectrum and the full Fisher matrices in the electronic form. A summary of the Fisher matrices is in Table 3.

In the noise dominated limit, the covariance matrix is $\mathbf{C} \approx \mathbf{N}$, and so the Fisher matrix scales as $\mathbf{N}^{-2}$. Assuming the noise is uncorrelated and Gaussian with standard deviation σ, this means that the Fisher matrix is proportional to $\mathbf{F} \propto \sigma^{-4} = (\text{S/N})^4$ (or the precision scales as $(\text{S/N})^2$), where we substituted the definition S/N$\equiv 1/\sigma$. We confirmed that this scaling holds true for high k values in the Fisher matrix.

5 DISCUSSION

We would like to start our discussion by highlighting the differences between QMLE and FFT. As we have shown in Section 2.2, QMLE finds inverse variance weighted averages across bins. Measuring deviations from a baseline power further lessens the averaging, so that QMLE yields near exact $P_{1D}(k)$. On the other hand, FFT estimator computes simple averages. For wide enough bins, this does not equal to $P_{1D}(k)$. Therefore, it is important to note that these two methods will not fully agree unless special circumstances are met, because they compute mathematically different quantities.

Systematic errors are the significant uncertainty source in P_{1D} analysis. We have considered three such error sources: (1) HCD absorbers, (2) metal contamination, and (3) continuum fitting. Using our method, we showed that HCDs can be masked without further complications. Modelling their contribution can now be solely a theoretical question as they are biased tracers in reality. However, metal contamination is more complicated as they are not easily separable from data like HCDs. Metal contamination estimates will inevitably bring statistical errors; and these errors can be fairly higher than our simple forecasts. Nevertheless, all methods face this challenge. Third, continuum errors can be more complicated than our simple model, but we believe these errors can still be marginalized with careful investigation. Even though marginalization washes out some information, it still enables us to keep the largest scales in P_{1D}. Furthermore, DESI will supply plenty of spectra to study and better understand these errors in the future.

Two additional error sources not considered here are uncertainties in the estimates of the resolution and the noise. Although, we used a simple Gaussian form of the spectrograph resolution and a diagonal noise matrix, the QMLE formalism allows us trivially extend this to more complicated forms, such as described in Bolton & Schlegel (2010). We defer a complete analysis, including a sensitivity analysis to resolution and noise misestimations to future work.

A disadvantage of our method over a direct FFT method is the additional computational time. However, we demonstrated that these calculations are now practical even for surveys of the scope of DESI, as we were able to perform multiple runs of full DESI-like data in a small cluster of 30 nodes with 24 cores each. We also introduced an optimization scheme that brings down the CPU time significantly. We think this cost is reasonable given QMLE's capacity to overcome Ly α forest specific challenges. Further computational speedups such as using GPUs are likely possible; we defer these to later work.

6 SUMMARY

The Ly α forest has emerged as a unique and competitive tool to investigate the large-scale structure of the universe. This technique can probe cosmological parameters on large scales, while being sensitive to the thermal state of the IGM, neutrino masses, and new dark matter models on small scales. This small-scale physics enriches the 1D and 3D power spectrum of the Ly α forest.

Table 3. Our Fisher matrix forecasts for 5-yr DESI survey for different spectral qualities. The subscript in *F* refers to the S/N value. The full table and model power spectrum can be found in the electronic submission of this article.

(z_i, k_i) [, (s km^{-1})]	(z_j, k_j) [, (s km^{-1})]	F_1 (s^2 km^{-2})	F_2 (s^2 km^{-2})	F_4 (s^2 km^{-2})
$(2.0, 5.00 \times 10^{-4})$	$(2.0, 5.00 \times 10^{-4})$	6.02867×10^1	6.51197×10^2	3.66215×10^3
$(2.0, 5.00 \times 10^{-4})$	$(2.0, 1.50 \times 10^{-3})$	6.63624×10^0	6.52238×10^1	3.02865×10^2
$(2.0, 5.00 \times 10^{-4})$	$(2.2, 5.00 \times 10^{-4})$	2.12509×10^1	2.30833×10^2	1.33759×10^3
$(2.0, 1.50 \times 10^{-3})$	$(2.0, 1.50 \times 10^{-3})$	6.54364×10^1	6.00179×10^2	2.47973×10^3
$(2.0, 1.50 \times 10^{-3})$	$(2.0, 2.50 \times 10^{-3})$	5.99838×10^0	5.43360×10^1	2.19731×10^2
$(2.0, 1.50 \times 10^{-3})$	$(2.2, 1.50 \times 10^{-3})$	1.97400×10^1	1.75264×10^2	6.90261×10^2
$(2.0, 2.50 \times 10^{-3})$	$(2.0, 2.50 \times 10^{-3})$	6.38029×10^1	5.69068×10^2	2.25086×10^3
$(2.0, 2.50 \times 10^{-3})$	$(2.0, 3.50 \times 10^{-3})$	5.83666×10^0	5.28085×10^1	2.13079×10^2
$(2.0, 2.50 \times 10^{-3})$	$(2.2, 2.50 \times 10^{-3})$	1.92028×10^1	1.65626×10^2	6.24662×10^2
$(2.0, 3.50 \times 10^{-3})$	$(2.0, 3.50 \times 10^{-3})$	6.29898×10^1	5.69389×10^2	2.29742×10^3
$(2.0, 3.50 \times 10^{-3})$	$(2.0, 4.50 \times 10^{-3})$	5.74635×10^0	5.31461×10^1	2.21748×10^2
$(2.0, 3.50 \times 10^{-3})$	$(2.2, 3.50 \times 10^{-3})$	1.89681×10^1	1.66002×10^2	6.39374×10^2
$(2.0, 4.50 \times 10^{-3})$	$(2.0, 4.50 \times 10^{-3})$	6.21816×10^1	5.77621×10^2	2.43005×10^3
$(2.0, 4.50 \times 10^{-3})$	$(2.0, 5.65 \times 10^{-3})$	5.85085×10^0	5.58077×10^1	2.44502×10^2
$(2.0, 4.50 \times 10^{-3})$	$(2.2, 4.50 \times 10^{-3})$	1.87450×10^1	1.68881×10^2	6.78999×10^2
$(2.0, 5.65 \times 10^{-3})$	$(2.0, 5.65 \times 10^{-3})$	8.19711×10^1	7.90781×10^2	3.53432×10^3
$(2.0, 5.65 \times 10^{-3})$	$(2.0, 7.11 \times 10^{-3})$	6.07740×10^0	6.05452×10^1	2.85148×10^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

In this work, we studied the optimal quadratic estimator (QMLE) for P_{1D} by first formulating its generic and Ly α specific expressions, offered an analytic continuum limit formula to depict what QMLE calculates and a straightforward way to bootstrap its results. Under the simplest terms, QMLE finds an inverse variance weighted average power for each bin. We then outlined the implementation in detail by providing our step-by-step algorithm. We here underline two of our decisions that mitigates the systematics and numerical instabilities: (1) we fed smoothed estimates to the covariance matrix, and (2) we picked a convergence criteria that weights changes between iterations by the error estimates. Then, we described our synthetic spectra and provided analytic expressions for what they construct. These analytic expressions were crucial to ascertain the performance of QMLE.

We generated DESI-oriented synthetic spectra in order to perform comprehensive tests. These mocks assumed constant resolution ($R = 3200$) and noise at all wavelengths, and limited the quasar redshift range to $z_{\mathrm{qso}} \in [2.1, 4.4]$ with pixel width $c\Delta\ln\lambda = 30$ km s^{-1}. A summary of our findings is as follows:

(i) Using 800 000 of these spectra with S/N = 2 Å^{-1}, we first showed that the power spectrum could be accurately measured with 1 per cent precision. This number declined towards higher redshifts due to diminishing statistics. This proved the absence of biases in our algorithm.

(ii) We randomly masked $\Delta\lambda = 25$ Å region of 15 per cent of spectra, and showed that this masking badly affected the measurement when untreated, and proved QMLE was robust against its effects on 100 000 mocks. As real analyses would mask bad pixels and high-density absorbers, this strength is invaluable to P_{1D} analysis.

(iii) We introduced continuum errors by adding wavelength-dependent error $\eta(\lambda) \sim 10$ per cent to flux fluctuations δ_F. This error function $\eta(\lambda)$ had two independent parameters for each quasar: amplitude and slope. Using eight independent runs with 100 000 spectra each, we demonstrated that these continuum errors contaminate first two k bins, but QMLE could marginalize out the noisiest modes, and recover these scales.

(iv) We presented, both analytically and by using 100 000 spectra, how a baseline model improved the accuracy. This fiducial power enables us to integrate over bins and to lessen the discretization of band powers. Our proposed unbiased procedure to find this fiducial power is to first measure the power spectrum without any input, then find the best-fitting analytic function on these results. We showed this feature significantly boosted our measurements.

(v) Finally, we found that computation was mostly spent on Fisher matrix calculation. We also found that Fisher matrix had a simple structure, which allowed us to come up with an optimization scheme that reduced the computation time by 60 per cent.

This work represents an initial step towards analysing the upcoming Lyman α data sets with surveys like DESI. The quadratic estimator formalism allows us to optimally use all the available data (even with varying S/N) and to effectively mitigate systematic errors like those arising from an imperfect continuum estimate. We address several practical issues with implementing a QMLE power spectrum code for the Lyman α forest. Future work will use these results and the codes described here to analyse both existing high-resolution data as well as upcoming DESI data.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

REFERENCES

Baur J., Palanque-Delabrouille N., Yéche C., Magneville C., Viel M., 2016, J. Cosmol. Astropart. Phys., 2016, 012
Bautista J. E. et al., 2017, A&A, 603, A12
Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067
Blomqvist M. et al., 2019, A&A, 629, A86
Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, ApJ, 872, 101
Bolton A. S., Schlegel D. J., 2010, PASP, 100119133735095
Boyarsky A., Lesgourgues J., Ruchayskiy O., Viel M., 2009, J. Cosmol. Astropart. Phys., 2009, 012
Busca N. G. et al., 2013, A&A, 552, A96
Chabanier S. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 017
Croft R. A. C., Hu W., Dave' R., 1999, Phys. Rev. Lett., 83, 1092
Dawson K. S. et al., 2016, AJ, 151, 44
Day A., Tytler D., Kambalur B., 2019, MNRAS, 489, 2536
de Sainte Agathe V. et al., 2019, A&A, 629, A85
Delubac T. et al., 2015, A&A, 574, A59
DESI Collaboration, 2016, preprint (arXiv:1611.00036)
du Mas des Bourboux H. et al., 2017, A&A, 608, A130
Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 681, 831
Font-Ribera A. et al., 2014, J. Cosmol. Astropart. Phys., 05, 027
Font-Ribera A., McDonald P., Slosar A., 2018, J. Cosmol. Astropart. Phys., 2018, 003
Garzilli A., Magalich A., Theuns T., Frenk C. S., Weniger C., Ruchayskiy O., Boyarsky A., 2019, MNRAS, 489, 3456
Gnedin N. Y., Hui L., 1998, MNRAS, 296, 44
Hamilton A. J. S., 1997, MNRAS, 289, 285
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Iršič V., Viel M., Haehnelt M. G., Bolton J. S., Becker G. D., 2017a, Phys. Rev. Lett., 119, 031302
Iršič V. et al., 2017b, MNRAS, 466, 4332
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Lomb N. R., 1976, Ap&SS, 39, 447
McDonald P. et al., 2006, ApJS, 163, 80 (M06)
Noterdaeme P. et al., 2012, A&A, 547, L1
Palanque-Delabrouille N. et al., 2013, A&A, 559, A85 (PD13)
Palanque-Delabrouille N. et al., 2015a, J. Cosmol. Astropart. Phys., 2015, 011
Palanque-Delabrouille N. et al., 2015b, J. Cosmol. Astropart. Phys., 2015, 045
Palanque-Delabrouille N. et al., 2016, A&A, 587, A41
Scargle J. D., 1982, ApJ, 263, 835
Seljak U., 1998, ApJ, 506, 64
Seljak U., Makarov A., McDonald P., Trac H., 2006, Phys. Rev. Lett., 97, 191303
Slosar A. et al., 2011, J. Cosmol. Astropart. Phys., 2011, 001
Slosar A. et al., 2013, J. Cosmol. Astropart. Phys., 2013, 026
Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22
Tegmark M., Hamilton A. J. S., Strauss M. A., Vogeley M. S., Szalay A. S., 1998, ApJ, 499, 555

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502

Walther M., Hennawi J. F., Hiss H., Oñorbe J., Lee K.-G., Rorai A., O'Meara J., 2017, ApJ, 852, 22 (W17)

Walther M., Oñorbe J., Hennawi J. F., Lukic Z., 2019, ApJ, 872, 13

Yeche C., Palanque-Delabrouille N., Baur J., BourBoux H. D. M. D., 2017, J. Cosmol. Astropart. Phys., 2017, 047

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

fiducial_p1d.txt
fisher_matrix_table.txt

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: SADDLE POINT APPROXIMATION FOR THE MEAN FLUX OF LOGNORMAL MOCKS

We can find a closed analytic form for the mean flux in equation (24) using saddle point approximation assuming the integrand is large around its maximum. We start by finding the derivatives of the function in the exponential:

$$\phi(\delta) \equiv -\frac{\delta^2}{2\sigma^2} - xe^{2a\delta}, \tag{A1}$$

$$\phi'(\delta) = -\frac{\delta}{\sigma^2} - 2axe^{2a\delta}, \tag{A2}$$

$$\phi''(\delta) = -\frac{1}{\sigma^2} - 4a^2xe^{2a\delta}, \tag{A3}$$

where the z dependence of $x(z)$ and $a(z)$ is suppressed for clarity. Solving for the maximum $\phi'(\delta_*) = 0$ yields $-\delta_* = 2a\sigma^2xe^{2a\delta_*}$, for which the solution is given by the Lambert $\mathcal{W}$ function.

$$\delta_*(z) = -\frac{1}{2a}\mathcal{W}\left(4a^2\sigma^2x\right) \tag{A4}$$

Note that $\delta_* < 0$. Substituting this solution back into ϕ and ϕ'' yields

$$\phi(\delta_*) = \frac{\delta_*}{2a\sigma^2}(1 - a\delta_*), \tag{A5}$$

$$\phi''(\delta_*) = -\frac{1 - 2a\delta_*}{\sigma^2}. \tag{A6}$$

Now we can approximate the integration around the maximum.

$$\bar{F}(z) \approx \frac{1}{\sigma\sqrt{2\pi}}\int_{\delta_*-\epsilon}^{\delta_*+\epsilon} \exp\left[\phi(\delta_*) + \frac{(\delta-\delta_*)^2}{2}\phi''(\delta_*)\right]d\delta, \tag{A7}$$

$$\approx \frac{e^{\phi(\delta_*)}}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \exp\left[\phi''(\delta_*)\frac{\delta^2}{2}\right]d\delta = \frac{e^{\phi(\delta_*)}}{\sqrt{-\phi''(\delta_*)\sigma^2}}, \tag{A8}$$

and note that $-\phi''(\delta_*)\sigma^2 = 1 - 2a(z)\delta_*(z)$. We can further define $d_*(z) = a(z)\delta_*(z)$ and $\sigma(z) = a(z)\sigma$ to simplify this expression.

$$\bar{F}(z) \approx [1 - 2d_*(z)]^{-1/2} \exp\left\{\frac{d_*(z)[1 - d_*(z)]}{2\sigma^2(z)}\right\}. \tag{A9}$$

As noted in the main section, this approximation starts to deviate from the truth at $z \lesssim 3$ up to 4 per cent. It would be interesting to use this function as a fitting template for the real data.

This paper has been typeset from a TeX/LaTeX file prepared by the author.