

## Light in the dark forest. Part I. An efficient optimal estimator for 3D Lyman-alpha forest power spectrum

Naim Göksel Karaçaylı  and Christopher M. Hirata 

*Center for Cosmology and AstroParticle Physics, The Ohio State University,  
191 West Woodruff Avenue, Columbus, OH 43210, U.S.A.*

*Department of Physics, The Ohio State University,  
191 West Woodruff Avenue, Columbus, OH 43210, U.S.A.*

*Department of Astronomy, The Ohio State University,  
140 W 18th Avenue, Columbus, OH 43210, U.S.A.*

*E-mail:* [karacayli.1@osu.edu](mailto:karacayli.1@osu.edu), [hirata.10@osu.edu](mailto:hirata.10@osu.edu)

**ABSTRACT:** The highly anisotropic nature of the Lyman-alpha ( $\text{Ly}\alpha$ ) forest data introduces a complex survey window function that complicates the measurement of the three-dimensional power spectrum ( $P_{3\text{D}}$ ). In this paper, we present the first fully optimal estimator for  $P_{3\text{D}}$ , which exactly deconvolves the survey window function and marginalizes contaminated modes that distort the power spectrum. Our approach adapts optimal estimator techniques developed for the 2D cosmic microwave background data to the 3D case. To achieve computational feasibility, we employ the conjugate gradient method and implement the  $\text{P}^3\text{M}$  formalism to handle large-scale and small-scale operations separately and efficiently. We validate our estimator using Monte Carlo mocks and Gaussian simulations, demonstrating its accuracy and computational efficiency. We confirm that mode marginalization eliminates distortions arising from quasar continuum errors and delivers robust power spectrum estimation, though it also inflates errors at large scales. This first implementation works in the flat-sky case; we discuss the remaining steps needed to generalize it to the curved-sky case. This formalism offers a foundation for the  $\text{Ly}\alpha$  forest  $P_{3\text{D}}$  measurements and a new path toward cosmological constraints from the  $\text{Ly}\alpha$  forest data.

**KEYWORDS:** Lyman alpha forest, power spectrum

**ARXIV EPRINT:** [2503.15619](https://arxiv.org/abs/2503.15619)

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Setup</b>	<b>3</b>
2.1	P3D model	3
2.2	Geometry	5
<b>3</b>	<b>Optimal estimation of P3D</b>	<b>6</b>
3.1	Conjugate gradient solution of $\mathbf{C}^{-1}\mathbf{x}$	6
3.2	Computing $\mathbf{S}\mathbf{z}$	7
3.3	Redshift evolution	9
3.4	Continuum marginalization	9
3.5	Multiplication with derivative matrices	10
3.6	Estimating the bias and the Fisher matrix	11
3.7	Some numerical details	14
<b>4</b>	<b>Validation</b>	<b>14</b>
4.1	Self-consistency of Monte Carlo simulations	15
4.2	Gaussian simulations	17
<b>5</b>	<b>Discussion</b>	<b>21</b>
<b>A</b>	<b>Matrix square root <math>\mathbf{Q} = \mathbf{A}^{1/2}</math></b>	<b>23</b>
A.1	Newton iteration	23
A.2	Newton-Schulz iteration	23
A.3	Padé iteration	24
<b>B</b>	<b>Quasi-optimal cross-forest estimator</b>	<b>24</b>
<b>C</b>	<b>Hermite interpolation details</b>	<b>25</b>

---

**1 Introduction**

Neutral hydrogen is a strong absorber, capable of completely blocking light from extragalactic sources, even in trace amounts. Photons emitted blueward of the  $\text{Ly}\alpha$  line are redshifted as they travel through the universe and can be subjected to resonant absorption with neutral hydrogen atoms they encounter. Gunn and Peterson first demonstrated that the observed flux of 3C 9 blueward of  $\text{Ly}\alpha$  indicated a highly ionized intergalactic medium (IGM) [1, 2]. This finding implied that the universe underwent another major phase transition (reionization) following the earlier recombination phase transition. Over the following decades, the study of these absorption features, known as the  $\text{Ly}\alpha$  forest, became a powerful tool for mapping vast volumes of the universe between redshifts 2 and 5, achieving spatial resolutions as fine as kiloparsec scales.

On large scales, the Ly $\alpha$  forest technique provided measurements of the standard ruler, baryon acoustic oscillation (BAO) scale, using 3D correlation function statistics ( $\xi_{3D}$ ) [3–11]. Recently, the measurement of the Alcock-Paczyński effect enabled a “full-shape” analysis and measurement of  $\Omega_m$  from the Ly $\alpha$  forest data alone [12]. On smaller scales, statistics of interest have been the 1D power spectrum ( $P_{1D}$ ), which has access to small-scale physics such as the thermal state of the IGM [13–15], the suppression due to the mass of neutrinos [16–18] and dark matter particles [19–26].

Current and next-generation surveys will significantly increase the amount of Ly $\alpha$  forest data. The Dark Energy Spectroscopic Instrument [DESI, 27] became operational in 2021 with a mission to collect 40 million spectra from galaxies and quasars in five years. The five-year DESI data will observe the Ly $\alpha$  forest from approximately 700,000 quasar spectra at  $z \gtrsim 2.1$  [28, 29], which will be four times larger than the preceding experiment, the Extended Baryon Oscillation Spectroscopic Survey [eBOSS, 30].

This great increase rekindled efforts to measure the 3D power spectrum of the Ly $\alpha$  forest ( $P_{3D}$ ) [31–34]. In principle, the large-scale 3D power spectrum contains a wealth of information about ionizing background fluctuations [35–38], hydrogen [39, 40] and helium [41] reionization, pre-reionization thermal history [42], and primordial non-Gaussianity [43, 44]. Even though the theoretical  $P_{3D}(\mathbf{k})$  and  $\xi_{3D}(\mathbf{r})$  are Fourier transform pairs and equivalent in information content, the same is not true for the estimated  $\hat{P}_{3D}(\mathbf{k})$  and  $\hat{\xi}_{3D}(\mathbf{r})$  from noisy observations [45]. For example, the errors of  $\hat{P}_{3D}(\mathbf{k})$  are significantly less correlated between  $\mathbf{k}$  modes compared to the errors of  $\hat{\xi}_{3D}(\mathbf{r})$  between  $\mathbf{r}$  bins. The survey window function mixes different  $\mathbf{k}$  modes in  $\hat{P}_{3D}(\mathbf{k})$ , whereas  $\hat{\xi}_{3D}(\mathbf{r})$  is robust against this effect. In fact, the strongly anisotropic nature of Ly $\alpha$  forest data — sparse sampling in the transverse direction and dense (but noisy) sampling in the line-of-sight direction — introduces a highly complex, challenging survey window function to  $P_{3D}$  estimation methods. The workaround to this problem has been to (1) propose an alternative hybrid statistic called  $P_{\times}(k_{\parallel}, \Delta\theta)$  that is in Fourier space in the line-of-sight direction and configuration space in the transverse direction [31, 32] or (2) sub-optimally estimate  $P_{3D}$  using a weighted pair-count estimator [33, 46].

In other words, the field has lacked an efficient, optimal estimator technique for  $P_{3D}$  until now. The optimal estimator formalism has been developed for cosmic microwave background (CMB) maps and fruitfully applied to those data sets [47–49].  $P_{1D}$  of the Ly $\alpha$  forest has been optimally estimated by refs. [50, 51] thanks to its reduced dimensionality. These optimal estimators minimize error bars, exactly deconvolve the survey window function, and marginalize contaminated modes from the data, making them highly desirable. Specifically, in the current Ly $\alpha$  forest analyses, the errors in the quasar continuum distort the two-point statistics and significantly complicate the cosmological inferences. Our optimal estimator not only yields the smallest error bars but also eliminates these distortions from the measurement.

At its core, the optimal estimator is a weighted average, where the weights are given by the inverse covariance matrix of the entire data vector. This makes the core computational operation  $\sim \mathbf{C}^{-1}\mathbf{x}$ . However, for data vectors of size  $\mathcal{O}(10^{7-9})$ , directly computing this inverse is infeasible. Instead of explicitly inverting the covariance matrix, we solve the linear system  $\mathbf{C}\mathbf{y} = \mathbf{x}$  for  $\mathbf{y}$  using the conjugate gradient method, as is commonly done in CMB analyses [48]. Moreover, the operation  $\mathbf{C}\mathbf{y}$  becomes computationally intractable if all matrix-vector multiplications are explicitly computed for every forest pair. We make

this operation feasible with the P<sup>3</sup>M formalism, originally developed for numerical gravity solvers and later applied to weak lensing surveys in optimal estimator formalism [52]. In this approach, the large-scale matrix-vector multiplication ( $\mathbf{C}_L \mathbf{y}$ ) is performed using a mesh-based method, capitalizing on fast Fourier-space convolution (particle-mesh, PM). Meanwhile, the small-scale matrix-vector multiplication ( $\mathbf{C}_S \mathbf{y}$ ) is computed exactly through direct summation between neighboring forests (particle-particle, PP).

In this work, we describe our proposed optimal estimator algorithm and its implementation in a Cartesian grid while focusing on overcoming Ly $\alpha$  specific challenges. We address mode marginalization, redshift evolution along the radial direction, and generating Monte Carlo simulations with proper correlations between lines of sight. The treatment of spherical geometry is left for future work. Our approach provides a robust framework for estimating the Ly $\alpha$  forest  $P_{3D}$  with good accuracy and high efficiency, and paves the way for precise cosmological measurements.

This algorithm can optimally extract information from both large, quasi-linear scales and small, non-linear scales. In this work, however, we focus on the large scales, where the estimator proves most impactful for sparsely sampled, large-scale structure surveys such as DESI. Our method can recover full-shape information in the power spectrum, revealing Alcock-Paczyński distortions, ionizing background fluctuations, and signatures of primordial non-Gaussianity. Conversely, a densely sampled survey like the Ly $\alpha$  Tomography IMACS Survey (LATIS, [53]) is better suited for measuring small-scale  $P_{3D}$ , though the benefits of the optimal estimator formalism may be limited unless mode marginalization becomes necessary.

This paper is organized as follows. Section 2 presents the setup of our problem, including the geometry and the input  $P_{3D}$  model to the optimal estimator, where a fitting function is sufficient since the estimator only needs to approximate the underlying signal. Section 3, the central section of this paper, provides a detailed description of our optimal estimator. Validation tests using Monte Carlo mocks and Gaussian simulations are discussed in section 4. The advantages and limitations of our method, and directions for future work, are examined in section 5.

## 2 Setup

We now set up the power spectrum estimation problem addressed in this paper. This includes a model for the fiducial Ly $\alpha$  power spectrum (section 2.1) and the simplified geometry within the flat-sky approximation (section 2.2).

### 2.1 P3D model

In linear theory, the Ly $\alpha$  forest power spectrum can be modeled with a linear bias term and a Kaiser term for the redshift space distortions. A major complication of modeling  $P_{3D}$  is the non-linear corrections  $D_{NL}(k, \mu)$  to this linear model. We further need to introduce an additional suppression to prevent small-scale modes from aliasing.

$$P_{3D}(k, \mu) = b_F^2 (1 + \beta_F \mu^2)^2 P_L(k) D_{NL}(k, \mu) e^{-k^2/k_A^2}, \quad (2.1)$$

where  $P_L(k)$  is the linear matter power spectrum at redshift  $z$ , and the additional exponential suppression at a characteristic scale  $k_A = 0.4 \text{ Mpc}^{-1}$  is for numerical stability. To be

consistent across our comparisons, we include this numerical term in our model. Empirical fitting functions of  $D_{\text{NL}}(k, \mu)$  aim to capture three major deviations from the linear theory: non-linear enhancement, pressure smoothing, and line-of-sight broadening. Non-linear collapse enhances the power. The thermal broadening of line profiles and non-linear peculiar velocities smooth the power spectrum only in the line-of-sight direction. Lastly, the gas pressure isotropically suppresses the power at a characteristic scale known as the filtering scale or Jeans scale [54]. As we will see below, this effect is negligible compared to the smoothing required for numerical stability and to prevent aliasing.

We use the fitting function for  $D_{\text{NL}}(k, \mu)$  presented in ref. [55] based on simulations, which is inspired by and improved upon the work in ref. [56] by reducing the number of free parameters to 5.

$$\ln D_{\text{NL}}(k, \mu) = q_1 \Delta^2(k) \left[ 1 - (k/k_\nu)^{a_\nu} \mu^{b_\nu} \right] - (k/k_p)^2, \quad (2.2)$$

where  $\Delta^2(k) = k^3 P_L(k)/2\pi^2$ . The non-linear enhancement is quantified by the parameter  $q_1$ . The pressure smoothing is captured by the parameter  $k_p$ . The anisotropic line-of-sight broadening is described with additional parameters that multiply the non-linear enhancement term.

Ref. [57] has simulated the Ly $\alpha$  forest using state-of-the-art techniques and provides the most recent best-fitting parameters. To quantify the redshift evolution of the bias,  $b_F(z)$ , we fit a power law  $b_F(z) = b_F^{\text{pivot}} \left( \frac{1+z}{1+z_{\text{pivot}}} \right)^{\alpha_F}$  to the values listed in table A3 of ref. [57], with  $z_{\text{pivot}} = 2.4$ . The remaining parameters are linearly interpolated to  $z_{\text{pivot}}$ . The values used in the optimal estimator are:  $b_F^{\text{pivot}} = -0.1196$ ,  $\alpha_F = 3.38$ ,  $\beta_F = 1.66$ ,  $q_1 = 0.796$ ,  $k_\nu = 0.392 \text{ Mpc}^{-1}$ ,  $a_\nu = 0.383$ ,  $b_\nu = 1.65$ . The reported value of  $k_p = 16.8 \text{ Mpc}^{-1}$  can be ignored since  $1/k_A \gg 1/k_p$ .

The linear power spectrum is computed using the Boltzmann solver CAMB [58], assuming a flat  $\Lambda$ CDM universe with Planck 2018 cosmology [59] at the pivot redshift.

### 2.1.1 High-column density systems

High-column density (HCD) systems trace highly overdense regions of the underlying matter field and follow a different bias relation, while their saturated absorption profiles strongly affect forest fluctuations along the line of sight. Since it is nearly impossible to completely identify, mask, and correct for these saturated absorption regions, HCDs inevitably modify the correlation function and power spectrum.

This modification follows a simple form at linear order [11, 60, 61]. Using the definitions  $\tilde{b}_F = b_F(1 + \beta_F \mu^2)$  and  $\tilde{b}_{\text{HCD}} = b_{\text{HCD}}(1 + \beta_{\text{HCD}} \mu^2) F_{\text{HCD}}(k_{\parallel})$ , the total 3D power spectrum is expressed as

$$P_{3\text{D}}(k, \mu) = P_L(k) \left( \tilde{b}_F^2 D_{\text{NL}}(k, \mu) + A_{\text{halo}}(k) (2\tilde{b}_F \tilde{b}_{\text{HCD}} + \tilde{b}_{\text{HCD}}^2) \right), \quad (2.3)$$

where  $b_{\text{HCD}}$  and  $\beta_{\text{HCD}}$  are the bias parameters for the absorption caused HCDs. We initially introduced the term  $A_{\text{halo}}$  to capture the decorrelation between the Ly $\alpha$  forest and haloes at  $1.5 \text{ Mpc}^{-1}$  [62].<sup>1</sup> However, aliasing occurs at larger scales than this decorrelation. So

<sup>1</sup>This decorrelation should not be included in the autocorrelation term  $\tilde{b}_{\text{HCD}}^2$ .

$A_{\text{halo}}$  is repurposed to suppress small-scale power to prevent aliasing.<sup>2</sup> Lastly, we use the approximation  $F_{\text{HCD}}(k_{\parallel}) = \exp(-L_{\text{HCD}}k_{\parallel})$ , which is the Fourier transform of the Lorentzian broadening in the Voigt profile, and  $L_{\text{HCD}}$  is the typical length of the HCD absorption [11].

We cut off contributions from HCDs at  $k_{\text{max}} = 0.4 \text{ Mpc}^{-1}$ , to mitigate aliasing and inaccurate modelling at small scales. To smoothly suppress these small-scale contributions, we apply an apodizing function identical to that in section 3.2:

$$A_{\text{halo}}(k) = \begin{cases} 1 & k < k_{\text{max}}/2 \\ \cos^2\left(\frac{\pi}{2}\left(\frac{k}{k_{\text{max}}/2} - 1\right)\right) & k_{\text{max}}/2 \leq k \leq k_{\text{max}} \\ 0 & k_{\text{max}} < k \end{cases}, \quad (2.4)$$

Our HCD model adopts best-fit parameters from ref. [11]:  $b_{\text{HCD}} = -0.05$ ,  $\beta_{\text{HCD}} = 0.7$ , and  $L_{\text{HCD}} = 14.8 \text{ Mpc}$ .

### 2.1.2 Metals

The effect of non-Ly $\alpha$  absorbers in the forest is small but non-negligible. Ionized silicon has transition wavelengths close to the Ly $\alpha$  wavelength and adds oscillations (peaks) to the power spectrum (correlation function) through redshift confusion. Let us illustrate how these systems can be included in our  $P_{3\text{D}}$  estimation for the strongest line, Si III (1206.52 Å).

First, the redshift confusion places Si III systems at an offset  $\Delta r = \chi(z_{\text{Si III}}) - \chi(z_{\text{Ly}\alpha})$ , where  $(1 + z_{\text{Si III}})\lambda_{\text{Si III}} = (1 + z_{\text{Ly}\alpha})\lambda_{\text{Ly}\alpha}$ . With cosmological parameters above,  $\Delta r \approx 30 \text{ Mpc}$  at  $z_{\text{Ly}\alpha} = 2.4$ . In general,  $\Delta r$  changes with redshift, but this change is negligible since  $\Delta r(z = 2.7) - \Delta r(z = 2.1) \approx 2 \text{ Mpc}$ . The peak in the correlation function at this separation is smeared due to this effect, redshift space distortions, and clustering. As these ions reside in the densest regions of the universe, we again refer to the Ly $\alpha$  forest-halo cross-correlation study of ref. [62]. We model the redshift space distortions with a Lorentzian decay  $D_{\text{FoG}} = 1/(1 + (k_{\parallel}\sigma_v)^2)$  and suppress the contribution of metals with the same  $A_{\text{halo}}(k)$  function. Similarly, defining  $\tilde{b}_{\text{Si III}} = b_{\text{Si III}}(1 + \beta_{\text{Si III}}\mu^2)$ , the Si III contribution to the total  $P_{3\text{D}}$  is

$$P_L(k)A_{\text{halo}}(k) \left( 2\tilde{b}_F\tilde{b}_{\text{Si III}} \cos(k_{\parallel}\Delta r) \sqrt{D_{\text{NL}}D_{\text{FoG}}} + \tilde{b}_{\text{Si III}}^2 D_{\text{FoG}} \right), \quad (2.5)$$

where we use  $\beta_{\text{Si III}} = 0.5$  and  $\sigma_v = 5 \text{ Mpc}$  [11, 63].<sup>3</sup>

In the correlation function analysis of DESI, all three Si II lines and a Si III line at 1206.52 Å are modeled, including their cross-correlations with each other [63]. We include all these in our metal contamination model with the following parameters:  $b_{\text{Si III}(1207)} = 9.8 \times 10^{-3}$ ,  $b_{\text{Si II}(1190)} = 4.5 \times 10^{-3}$ ,  $b_{\text{Si II}(1193)} = 3.05 \times 10^{-3}$ , and  $b_{\text{Si II}(1260)} = 4.0 \times 10^{-3}$ .

## 2.2 Geometry

For this first implementation, we work in Cartesian geometry with a flat sky. This is the first step in understanding the fundamental linear algebra issues before proceeding to the spherical geometry of a real survey.

<sup>2</sup>In hindsight, this aliasing term could have been combined with the one above to have a single exponential suppression.

<sup>3</sup>This  $\sigma_v$  is approximately the best-fitting Lorentzian parameter value for the quasar fingers of god effect in ref. [63] and may be non-negligibly influenced by quasar redshift errors. We only need to approximate this smoothing in our model, so it is sufficient for our purposes.

Unrolling spherical shells into Cartesian space inevitably distorts the geometry. To prevent sightlines from diverging, we fix  $(x, y)$  coordinates at an effective comoving distance  $\chi_{\text{eff}}$  such that:

$$x = \chi_{\text{eff}}(\phi + \phi_0), \quad y = \chi_{\text{eff}}(\theta - \theta_0), \quad z = \chi(z_r) - \chi_0 \quad (2.6)$$

where  $\theta$  is declination,  $\phi$  is right ascension of the quasar, and  $\chi(z_r)$  is the comoving distance to redshift  $z_r$ .  $\chi_0$  is the initial distance in the radial direction where the mesh begins and is chosen to be large enough to minimize the periodicity effects from FFT.  $\theta_0$  is chosen as the median declination to further minimize distortions due to projection.

### 3 Optimal estimation of P3D

Our main task is to optimally estimate the 3D power spectrum of the Ly $\alpha$  forest on large, weakly non-linear scales from noisy and spatially sparse transmitted flux fluctuations. The quasar spectra with the Ly $\alpha$  forest data form unevenly sampled, infinitesimal lines in space. The transmitted flux fluctuations contain nonuniform noise and are biased due to the quasar continuum determination procedure that fits the underlying large-scale fluctuations. The data vector can be expressed as  $\mathbf{d} = \boldsymbol{\delta}_F + \mathbf{n} + \mathbf{u}$ , where  $\boldsymbol{\delta}_F$  are the transmitted flux fluctuations convolved by the spectrograph window function in the line of sight,  $\mathbf{n}$  is noise, and we denote the continuum error templates by vector  $\mathbf{u}$ .

The quadratic (optimal) estimator formalism has been a highly successful technique in power spectrum measurements from the cosmic microwave background radiation, galaxy surveys, and weak lensing [47, 48, 52, 64–66]. We use this formalism to build a  $P_{3D}$  estimator that maximizes the likelihood function for the data  $\mathbf{d}$  under the Gaussianity assumption. We use  $\mathcal{L}$  to denote the negative logarithm of the probability,  $2\mathcal{L}(P_{3D}(k_{\perp}, k_{\parallel})|\mathbf{d}) = \ln \det \mathbf{C} + \mathbf{d}^T \mathbf{C}^{-1} \mathbf{d}$ , where  $\mathbf{C} = \langle \mathbf{d} \mathbf{d}^T \rangle = \mathbf{S} + \mathbf{N}$  is the full covariance matrix. The most likely  $P_{3D}$  can be found by maximizing this likelihood function:  $\partial \mathcal{L} / \partial P_k = 0$  (we use subscript  $k$  to denote one  $(k_{\perp}, k_{\parallel})$  bin for clarity from here on), which can be iteratively solved using the Newton-Raphson method. However, with a good initial estimate of the covariance matrix, a single iteration is sufficient [51]. The maximum likelihood estimator is then

$$\hat{P}_k^{3D} = P_k^{3D, \text{fid}} + \sum_{k'} \frac{1}{2} F_{kk'}^{-1} \left[ \mathbf{d}^T \mathbf{C}^{-1} \mathbf{C}_{,k'} \mathbf{C}^{-1} \mathbf{d} - \text{Tr}(\mathbf{C}^{-1} \mathbf{C}_{,k'}) \right], \quad (3.1)$$

where  $\mathbf{C}_{,k} \equiv \partial \mathbf{C} / \partial P_k^{3D}$  and  $\text{Tr}(\mathbf{C}^{-1} \mathbf{C}_{,k})$  is the bias term, and the estimated Fisher matrix is

$$F_{kk'} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial P_k^{3D} \partial P_{k'}^{3D}} \right\rangle = \frac{1}{2} \text{Tr}(\mathbf{C}^{-1} \mathbf{C}_{,k} \mathbf{C}^{-1} \mathbf{C}_{,k'}). \quad (3.2)$$

Thus, computing  $\mathbf{C}$ , its inverse, and its derivatives with respect to  $P_{3D}$  is the focal point of this method.

#### 3.1 Conjugate gradient solution of $\mathbf{C}^{-1} \mathbf{x}$

The data vector  $\mathbf{d}$  has  $3.4 \times 10^7$  pixels for coarse-grained eBOSS data and is expected to grow to  $6 \times 10^8$  for full-resolution DESI data.<sup>4</sup> The covariance matrix  $\mathbf{C}$  cannot be stored

<sup>4</sup>This large increase in number of pixels cannot be explained by improved resolution of approximately 20–30% in DESI and likely comes from improvements in the rest of the pipeline.

in memory or inverted trivially. We instead solve  $\mathbf{y} = \mathbf{C}^{-1}\mathbf{x}$  iteratively using the conjugate gradient method.

We first put the equation  $(\mathbf{S} + \mathbf{N})\mathbf{y} = \mathbf{x}$  into a more stable form while retaining the symmetric, positive-definite structure:

$$(\mathbf{I} + \mathbf{N}^{-1/2}\mathbf{S}\mathbf{N}^{-1/2})\mathbf{N}^{1/2}\mathbf{y} = \mathbf{N}^{-1/2}\mathbf{x} \quad (3.3)$$

$$(\mathbf{I} + \mathbf{N}^{-1/2}\mathbf{S}\mathbf{N}^{-1/2})\mathbf{z} = \mathbf{m}. \quad (3.4)$$

where we start with an initial guess  $\mathbf{z}^{(0)}$ , and iterate until  $\mathbf{z}$  converges. Then, the solution is  $\mathbf{y} = \mathbf{N}^{-1/2}\mathbf{z}$  at convergence. We underscore that we only use the numerically stable  $\mathbf{N}^{-1/2}$ , which is zero for masked pixels.

The preconditioned conjugate gradient method requires at least three additional vectors to be stored: the residual vector  $\mathbf{r}$ , the search direction  $\mathbf{p}$ , and, of course, the output vector. After the initial guess  $\mathbf{z}^{(0)}$ , the residual vector is set to  $\mathbf{r}^{(0)} = \mathbf{m} - \mathbf{A}\mathbf{z}^{(0)}$ , where  $\mathbf{A}$  is the matrix in parentheses in eq. (3.4), and the search direction is set to  $\mathbf{p}^{(0)} = \mathbf{E}\mathbf{r}^{(0)}$ , where  $\mathbf{E}$  is the preconditioner. For iteration  $(i + 1)$ , we first calculate  $r_2^{(i)} = \mathbf{r}^{(i)\text{T}}\mathbf{E}\mathbf{r}^{(i)}$ ,  $\tilde{\mathbf{p}}^{(i)} = \mathbf{A}\mathbf{p}^{(i)}$  and  $p_2^{(i)} = \tilde{\mathbf{p}}^{(i)\text{T}}\mathbf{p}^{(i)}$ . Then, we update the solution  $\mathbf{z}^{(i+1)} = \mathbf{z}^{(i)} + \alpha^{(i)}\mathbf{p}^{(i)}$  and the residual vector  $\mathbf{r}^{(i+1)} = \mathbf{r}^{(i)} - \alpha^{(i)}\tilde{\mathbf{p}}^{(i)}$ , where  $\alpha^{(i)} = r_2^{(i)}/p_2^{(i)}$ . If convergence has not been achieved, we update the search direction  $\mathbf{p}^{(i+1)} = \mathbf{E}\mathbf{r}^{(i+1)} + \beta^{(i)}\mathbf{p}^{(i)}$ , where  $\beta^{(i)} = r_2^{(i+1)}/r_2^{(i)}$ . We end the iterations if  $\|\mathbf{r}^{(i+1)}\|_2 < \epsilon \cdot \|\mathbf{r}^{(0)}\|_2$ , where we pick  $\epsilon = 10^{-6}$ – $10^{-5}$ . The algorithm typically achieves convergence between 30 and 50 iterations.

We use the block diagonal approximation as the preconditioner,  $\mathbf{E} = \mathbf{A}_{\text{BD}}^{-1}$ , which is also used for the initial guess  $\mathbf{z}^{(0)} = \mathbf{A}_{\text{BD}}^{-1}\mathbf{m}$ . Better preconditioners can be constructed by formulating the noise matrix in Fourier space [48, 49].

### 3.2 Computing $\mathbf{S}\mathbf{z}$

We use a P<sup>3</sup>M algorithm to efficiently construct and perform required matrix operations following a similar technique applied to weak lensing surveys [52]. We decompose the power spectrum into large- and small-scale components by an adjustable large-scale radius parameter  $R_L$ :

$$P_L = P_{3\text{D}}(k, k_{\parallel})e^{-k^2 R_L^2} \quad \text{and} \quad P_S = P_{3\text{D}}(k, k_{\parallel})(1 - e^{-k^2 R_L^2}). \quad (3.5)$$

We perform the large-scale signal matrix  $\mathbf{S}_L$  multiplication as a convolution on the mesh in Fourier space using FFTs and perform the small-scale signal matrix  $\mathbf{S}_S$  multiplications only for the neighboring lines of sight for each quasar within  $fR_L$ , where  $f$  is the neighboring radius factor that is also adjustable. In this work, we adopt values of  $R_L = 6$  and 8 Mpc and  $f = 8$  and 12.

#### 3.2.1 Large-scale component

The large-scale part  $\mathbf{S}_L$  uses harmonic space, and as such, it is the step that depends most sensitively on the choice of the Cartesian/flat sky (with Fourier modes) versus the curved sky (with spherical harmonics). We first test our method with Cartesian grids, which are more straightforward to implement and interpret, and leave the spherical sky treatment to future work.

We reverse interpolate vectors using the clouds-in-cell (CIC) assignment function onto a Cartesian grid that is an equirectangular projection. We then perform FFT, multiplication of each Fourier mode by  $P_L(\mathbf{k})$ , FFT back to real space, and do trilinear interpolation to the sightlines. The operation  $\mathbf{S}_L$  computed this way is semi-positive definite since  $P_L(\mathbf{k}) \geq 0$ , and exactly symmetric since the CIC operation from sightline pixels to Cartesian grid points is the transpose of trilinear interpolation.

### 3.2.2 Small-scale component

The small-scale correlation function is given by

$$\xi_S(r_\perp, r_\parallel) = A(r) \int_0^\infty \frac{k_\perp dk_\perp}{2\pi} J_0(k_\perp r_\perp) \int_0^\infty \frac{dk_\parallel}{\pi} \cos(k_\parallel r_\parallel) P_S(k_\perp, k_\parallel), \quad (3.6)$$

where an apodizing function  $A(r)$  is applied to reduce ringing in the Fourier space of a sharp transition to zero at  $r = fR_L$ . This spherical top-hat induced ringing can lead to negative eigenvalues in the covariance matrix. A smooth transition function suppresses this ringing, although it does not eliminate it. The apodizing function used for this purpose is given by

$$A(r) = \begin{cases} 1 & r < r_1 \\ \cos^2\left(\frac{\pi}{2} \frac{r-r_1}{fR_L-r_1}\right) & r_1 \leq r \leq fR_L \\ 0 & fR_L < r, \end{cases} \quad (3.7)$$

which smoothly transitions from one to zero starting at the breakpoint  $r_1 = (3/4)fR_L$ .

We compute the integrations using the FFTLog algorithm [67, 68], noting that  $\cos(x) = J_{-1/2}(x) \sqrt{\pi x/2}$ . To reduce FFTLog ringing, the transformations are performed over a wide  $k$  range of  $10^{-6}$ – $10^6$  Mpc $^{-1}$  with 2048 points,<sup>5</sup> and only the range  $r \in [5 \times 10^{-4}$ – $10^3]$  Mpc is kept.<sup>6</sup> This is a highly conservative choice — a smaller number of points in a shorter  $k$  span can still achieve low ringing. Since we calculate  $\xi_S$  once and interpolate later, we opt for the high accuracy in the FFTLog calculation.

### 3.2.3 Spectrograph resolution

As a further complication, both of these expressions will have a wavelength-dependent spectrograph window function contribution  $W(k_\parallel, \lambda)$ . In general, each forest falls onto a different part of the spectrograph and can have a unique spectrograph window function. This is formally treated in ref. [31] using an alternative summary statistics  $P_\times(\Delta\theta, k_\parallel)$ . For simplicity, we instead use an average window function of the form:

$$W(k_\parallel) = e^{-k_\parallel^2 \sigma_\parallel^2 / 2} \text{sinc}\left(k_\parallel \Delta r_\parallel / 2\right), \quad (3.8)$$

where  $\sigma_\parallel = 69 \text{ km s}^{-1} \approx 1 \text{ Mpc}$  and  $\Delta r_\parallel \approx 2.9 \text{ Mpc}$  for eBOSS Ly $\alpha$  forest data at  $z = 2.4$  since eBOSS Ly $\alpha$  data is coarse-grained by a factor of 3.

The spectrograph resolution can be much finer than the mesh grid spacing, so its contributions to the large-scale component can be ignored in the P<sup>3</sup>M formalism. Still, we use

<sup>5</sup> $10^6$  Mpc = 1 Tpc.

<sup>6</sup>Note that there are two sources of ringing, one from the top-hat cutoff and one from FFTLog.

an average window function in the convolution as it does not cost additional computational time. In theory, the small-scale contribution can incorporate unique spectrograph window functions for each quasar by computing  $\xi_S(r_\perp, r_\parallel)$  using  $P_S \rightarrow P_S W_a(k_\parallel) W_b(k_\parallel)$  for quasars  $a$  and  $b$  similar to ref. [31]. However, this is extremely costly as it requires the recalculation of  $\xi_S(r_\perp, r_\parallel)$  for every quasar pair.

Lastly, DESI’s spectral extraction pipeline provides a resolution matrix  $\mathbf{R}_q$  for each quasar [69, 70]. Under the resolution matrix formalism, the spectrograph line smearing can be represented with a block-diagonal matrix  $\mathbf{R}$  operating on the entire data vector  $\mathbf{d}$  such that  $\mathbf{S} \rightarrow \mathbf{R}\mathbf{S}\mathbf{R}^T$ . This can be implemented per forest matrix-vector multiplications for both components without modifying eqs. (3.5) and (3.6) or approximating the window function.

### 3.3 Redshift evolution

The 3D map of the Ly $\alpha$  forest spans a large volume between  $2.0 \leq z \leq 3.8$ . The translational symmetry is strongly broken in the line-of-sight direction, and the FFT is not well-defined. We solve this problem by calculating the signal matrix at a pivot redshift  $z_{\text{pivot}} = 2.4$  and applying the redshift growth function  $G(z)$  separately from the convolution with FFTs.

For a pixel pair  $ij$ , the redshift-evolved signal matrix is given by  $S_{ij} = G(z_{ij})S_{ij}^{\text{pivot}}$ , where  $1 + z_{ij} = \sqrt{(1 + z_i)(1 + z_j)}$ . This is an element-wise operation and not trivial in the general case. However, for separable redshift growth functions,  $G(z_{ij}) = \sqrt{G(z_i)} \times \sqrt{G(z_j)}$ , this element-wise operation becomes

$$\mathbf{S} = \mathbf{G}^{1/2} \mathbf{S}^{\text{pivot}} \mathbf{G}^{1/2}, \quad (3.9)$$

where  $\mathbf{G}^{1/2}$  is a diagonal matrix and now trivial to implement. This is a valid assumption since the redshift growth function will be separable for power law growth functions of  $\sim (1 + z_{ij})^\gamma$ . We further propagate this growth function formalism to derivative matrices by  $\mathbf{C}_{,k} \rightarrow \mathbf{G}^{1/2} \mathbf{C}_{,k} \mathbf{G}^{1/2}$  and estimate  $P_{3\text{D}}$  at a known pivot redshift.

We use  $G^{1/2}(z) = b_F(z)D(z)/b_F^{\text{pivot}}D(z_{\text{pivot}})$ , where  $D(z)$  is the linear growth function of the matter field and  $b_F(z)$  is the bias of the Ly $\alpha$  forest. As discussed in section 2.1,  $z_{\text{pivot}} = 2.4$  and  $b_F(z)/b_F^{\text{pivot}} = [(1 + z)/3.4]^{\alpha_F}$ , where  $\alpha_F = 3.38$ .

These simplifying assumptions are sufficient for our work. For more complicated treatments of the redshift evolution, one could use multiple grids at different redshifts and interpolate the results into a single vector. A critical shortcoming of the single-grid formalism is that when metals or DLAs are included in the covariance matrix model, they must be assumed to have the same growth as Ly $\alpha$  forest.

### 3.4 Continuum marginalization

Another critical challenge in computing  $\mathbf{C}^{-1}\mathbf{x}$  arises from the missing mode matrix. This is handled in many 3D correlation function studies by applying a distortion matrix to the theory [11]. In this work, we include it as part of the power spectrum estimation.

The missing mode matrix arises because some components in the spectrum are lost at the quasar continuum fitting stage. Usually,  $n_C$  continuum modes — typically polynomials of  $\ln \lambda$  — are fitted out for each quasar, which also removes real signal from the Ly $\alpha$  forest. We denote these directions in data vector space by  $\mathbf{u}^{(A)}$ , where  $A = 1, \dots, n_C N_Q$  and  $n_C$  is

one plus the order of the continuum fitting polynomial. Typically,  $n_C = 2$ , so for each quasar an overall mean and tilt are projected out. The  $A$ th component of this vector is non-zero for sightline  $\lfloor A/n_C \rfloor$ , and zero for the other sightlines. So we have

$$\boldsymbol{\Sigma} = \sigma \sum_{A=1}^{n_C N_Q} \mathbf{u}^{(A)} \mathbf{u}^{(A)\text{T}}, \quad \sigma \rightarrow \infty. \quad (3.10)$$

Working with matrices that formally go to  $\infty$  is undesirable, so we use the limiting form of the Woodbury formula,

$$(\mathbf{N} + \boldsymbol{\Sigma})^{-1} = \mathbf{N}^{-1} - \sum_{A,B} [\boldsymbol{\Gamma}^{-1}]_{AB} \mathbf{t}^{(A)} \mathbf{t}^{(B)\text{T}}, \quad \Gamma_{AB} = \mathbf{u}^{(A)\text{T}} \mathbf{N}^{-1} \mathbf{u}^{(B)}, \quad \mathbf{t}^{(A)} = \mathbf{N}^{-1} \mathbf{u}^{(A)}, \quad (3.11)$$

which no longer involves infinite entries. Since  $\mathbf{N}$  is diagonal, the matrix  $\boldsymbol{\Gamma}$  breaks into  $N_Q$  blocks of size  $n_C \times n_C$  each, and  $(\mathbf{N} + \boldsymbol{\Sigma})^{-1}$  breaks into  $N_Q$  blocks of size  $n_\lambda \times n_\lambda$  each. We can further define  $\tilde{\mathbf{u}}^{(A)} = \mathbf{N}^{-1/2} \mathbf{u}^{(A)}$  to write  $\Gamma_{AB} = \tilde{\mathbf{u}}^{(A)\text{T}} \cdot \tilde{\mathbf{u}}^{(B)}$  and

$$(\mathbf{N} + \boldsymbol{\Sigma})^{-1} = \mathbf{N}^{-1/2} \left( \mathbf{I} - \sum_{A,B} [\boldsymbol{\Gamma}^{-1}]_{AB} \tilde{\mathbf{u}}^{(A)} \tilde{\mathbf{u}}^{(B)\text{T}} \right) \mathbf{N}^{-1/2}. \quad (3.12)$$

There are several paths forward to consider here. We find that the square root of the matrix in parentheses is easier to implement:

$$\mathbf{I} - \sum_{A,B} [\boldsymbol{\Gamma}^{-1}]_{AB} \tilde{\mathbf{u}}^{(A)} \tilde{\mathbf{u}}^{(B)\text{T}} = \boldsymbol{\Pi}^{1/2} \boldsymbol{\Pi}^{1/2}, \quad (3.13)$$

such that  $(\mathbf{N} + \boldsymbol{\Sigma})^{-1} = \mathbf{N}^{-1/2} \boldsymbol{\Pi}^{1/2} \boldsymbol{\Pi}^{1/2} \mathbf{N}^{-1/2}$ . We compute  $\boldsymbol{\Pi}^{1/2}$  once using eigenvalue decomposition for each quasar. These matrices are stored in files and read when needed, as they significantly increase the memory usage.

The covariance matrix-vector multiplication  $\mathbf{C}\mathbf{y} = (\mathbf{S} + \mathbf{N} + \boldsymbol{\Sigma})\mathbf{y} = \mathbf{x}$  can be written as

$$(\mathbf{S} + \mathbf{N}^{1/2} \boldsymbol{\Pi}^{-1/2} \boldsymbol{\Pi}^{-1/2} \mathbf{N}^{1/2}) \mathbf{y} = \mathbf{x} \quad (3.14)$$

or

$$(\mathbf{I} + \boldsymbol{\Pi}^{1/2} \mathbf{N}^{-1/2} \mathbf{S} \mathbf{N}^{-1/2} \boldsymbol{\Pi}^{1/2}) \boldsymbol{\Pi}^{-1/2} \mathbf{N}^{1/2} \mathbf{y} = \boldsymbol{\Pi}^{1/2} \mathbf{N}^{-1/2} \mathbf{x}, \quad (3.15)$$

such that  $\mathbf{m} = \boldsymbol{\Pi}^{1/2} \mathbf{N}^{-1/2} \mathbf{x}$ , and  $\mathbf{y} = \mathbf{N}^{-1/2} \boldsymbol{\Pi}^{1/2} \mathbf{z}$ . As an additional advantage of this formalism, random noise realizations in MC simulations remain well-defined as matrix multiplications on finite values of  $\mathbf{N}^{-1/2} \mathbf{n} \sim \mathcal{N}(\mu = 0, \sigma = 1)$ .

### 3.5 Multiplication with derivative matrices

Another core operation required by the optimal estimator is of the form  $\mathbf{x}^{\text{T}} \mathbf{C}_{,k} \mathbf{x}$ . As noted in the previous section, the derivative matrix includes a redshift evolution  $\mathbf{C}_{,k} = \mathbf{G}^{1/2} \mathbf{C}_{,k}^{\text{pivot}} \mathbf{G}^{1/2}$ . Therefore, devolving redshift evolution from the vector  $\mathbf{x}$  entails a simple  $\mathbf{x} \rightarrow \mathbf{G}^{1/2} \mathbf{x}$  operation. The P<sup>3</sup>M methodology still allows us to divide this operation into large- and small-scale components:  $\mathbf{x}^{\text{T}} (\mathbf{C}_{L,k} + \mathbf{C}_{S,k}) \mathbf{x}$ .

The power spectrum can be estimated as  $P(k_\perp, k_\parallel)$  or in multipoles  $P_\ell(k)$  such that  $P(k, \mu) = \sum_{\ell=0}^{\ell_{\text{max}}} P_\ell(k) L_\ell(\mu)$  where  $L_\ell$  are the Legendre polynomials. While we describe both approaches below, our analysis focuses on estimating multipoles.

### 3.5.1 Measuring $P(k_\perp, k_\parallel)$

In Fourier space,  $\tilde{\mathbf{C}}_{L,k}$  is a diagonal matrix where the diagonals are  $\exp(-k^2/R_L^2)W^2(k_\parallel)$  if the Fourier mode is in  $(k_\perp, k_\parallel)$  bin, and zero otherwise. Therefore,  $\tilde{\mathbf{x}}^\dagger \tilde{\mathbf{C}}_{L,k} \tilde{\mathbf{x}}$  is a simple calculation that can be done purely on the mesh. The small-scale derivative matrices can be constructed using the derivative of eq. (3.6):

$$\begin{aligned} \mathbf{C}_{S,k}(r_\perp, r_\parallel) &= \int_{k_\perp^{(1)}}^{k_\perp^{(2)}} \frac{k_\perp dk_\perp}{2\pi} J_0(k_\perp r_\perp) \int_{k_\parallel^{(1)}}^{k_\parallel^{(2)}} \frac{dk_\parallel}{\pi} \cos(k_\parallel r_\parallel) W^2(k_\parallel) (1 - e^{-k^2/r_L^2}) \\ &= q_\perp^{T,k}(r_\perp) q_\parallel^{T,k}(r_\parallel) - q_\perp^{L,k}(r_\perp) q_\parallel^{L,k}(r_\parallel), \end{aligned} \quad (3.16)$$

where superscripts (1) and (2) denote the lower and upper bin edges, respectively, and we completely separated  $k_\perp$  and  $k_\parallel$  integrations and defined the integrals:

$$\begin{aligned} q_\perp^{T,k}(r_\perp) &= \int_{k_\perp^{(1)}}^{k_\perp^{(2)}} \frac{k_\perp dk_\perp}{2\pi} J_0(k_\perp r_\perp), & q_\perp^{L,k}(r_\perp) &= \int_{k_\perp^{(1)}}^{k_\perp^{(2)}} \frac{k_\perp dk_\perp}{2\pi} J_0(k_\perp r_\perp) e^{-k_\perp^2/r_L^2}, \\ q_\parallel^{T,k}(r_\parallel) &= \int_{k_\parallel^{(1)}}^{k_\parallel^{(2)}} \frac{dk_\parallel}{\pi} \cos(k_\parallel r_\parallel) W^2(k_\parallel), & \text{and} & \quad q_\parallel^{L,k}(r_\parallel) = \int_{k_\parallel^{(1)}}^{k_\parallel^{(2)}} \frac{dk_\parallel}{\pi} \cos(k_\parallel r_\parallel) W^2(k_\parallel) e^{-k_\parallel^2/r_L^2}. \end{aligned} \quad (3.17)$$

The  $q_\perp^{T,k}$  integration can be done exactly:

$$q_\perp^{T,k}(r_\perp) = \frac{k_\perp^{(2)} J_1(k_\perp^{(2)} r_\perp) - k_\perp^{(1)} J_1(k_\perp^{(1)} r_\perp)}{2\pi r_\perp}; \quad \text{note } \lim_{r_\perp \rightarrow 0} q_\perp^{T,k}(r_\perp) = \frac{k_{\perp,(2)}^2 - k_{\perp,(1)}^2}{4\pi}. \quad (3.18)$$

### 3.5.2 Measuring multipoles $P_\ell(k)$

In practice, most of the  $P_{3D}$  information from the Ly $\alpha$  forest alone (based on the best-fitting values to the model of ref. [55]) is contained in the  $\ell = 0, 2, 4$  multipoles at linear scales [33, 71]. However, metal lines exhibit distinct oscillation signatures in  $\ell \geq 6$  multipoles, so we also estimate higher-order multipoles. The reduced dimensionality of the multipole approach greatly helps the precision of the estimated covariance matrix.

In Fourier space,  $\tilde{\mathbf{C}}_{L,k}$  remains a diagonal matrix, with diagonal elements given by  $\exp(-k^2/R_L^2) W^2(k_\parallel) L_\ell(\mu)$  if the Fourier mode falls within a  $k$  bin, and zero otherwise. This introduces an additional  $L_\ell(\mu)$  weighting compared to Cartesian binning. We apply linear interpolation across  $k$  bins in estimating multipoles, introducing another term — triangular binning function [51]. This correlates neighboring  $k$  bins, but as shown in section 4, the bins are already fairly correlated.

Small-scale corrections can be derived following the steps above. However, we ignore these contributions here since the measurement precision does not necessitate their inclusion.

## 3.6 Estimating the bias and the Fisher matrix

We estimate the bias term  $b_k = \text{Tr}(\mathbf{C}^{-1} \mathbf{C}_{,k})$  using the stochastic trace estimation method. We generate Monte Carlo (MC) realizations of the data vector  $\mathbf{q}$ , such that  $\langle \mathbf{q} \mathbf{q}^\text{T} \rangle = \mathbf{C}$ , and

$$b_k = \text{Tr}(\mathbf{C}^{-1} \mathbf{C}_{,k} \mathbf{C}^{-1} \mathbf{C}) = \langle \mathbf{q}^\text{T} \mathbf{C}^{-1} \mathbf{C}_{,k} \mathbf{C}^{-1} \mathbf{q} \rangle. \quad (3.19)$$

As noted in Oh et al. [48], these MC simulations give the Fisher matrix for free:

$$F_{kk'} = \frac{1}{4} \langle (b_k - \langle b_k \rangle)(b_{k'} - \langle b_{k'} \rangle) \rangle. \quad (3.20)$$

We find this to be the fastest method to obtain the Fisher matrix, albeit  $N_{\text{MC}} \gtrsim 2,000$  realizations are required to suppress noise in the off-diagonal terms. We use the estimated variance on the mean to quantify the accuracy of our MC simulations  $\mathbf{b}^{(i)}$ , which is given by  $\sum_i (\mathbf{b}^{(i)} - \langle \mathbf{b} \rangle)^2 / N_{\text{MC}}(N_{\text{MC}} - 1)$ , where  $\langle \mathbf{b} \rangle = \sum_i \mathbf{b}^{(i)} / N_{\text{MC}}$ . After 2,880 realizations, we reach an average  $6 \times 10^{-5}$  deviation.

The primary challenge in this front arises from incorporating correlated small-scale fluctuations between sightlines. To address this, we construct a composite random vector  $\mathbf{q} = \mathbf{q}_L + \mathbf{q}_S$ , where the two components are independent and have covariances  $\langle \mathbf{q}_L \mathbf{q}_L^T \rangle = \mathbf{C}_L \equiv \mathbf{S}_L$  and  $\langle \mathbf{q}_S \mathbf{q}_S^T \rangle = \mathbf{C}_S \equiv \mathbf{S}_S + \mathbf{N}$  respectively, where  $\mathbf{S}_L$  and  $\mathbf{S}_S$  are the long-range (PM) and short-range (PP) signal matrices. Theoretically one should include a  $\mathbf{q}_\Sigma$  with covariance  $\Sigma$  as well, which would be a linear combination of continuum modes with formally infinite variance; but in the limit  $\sigma \rightarrow \infty$ , the covariance of  $\mathbf{C}^{-1} \mathbf{q}_\Sigma$  goes to zero, so it is not necessary to include  $\mathbf{q}_\Sigma$ .

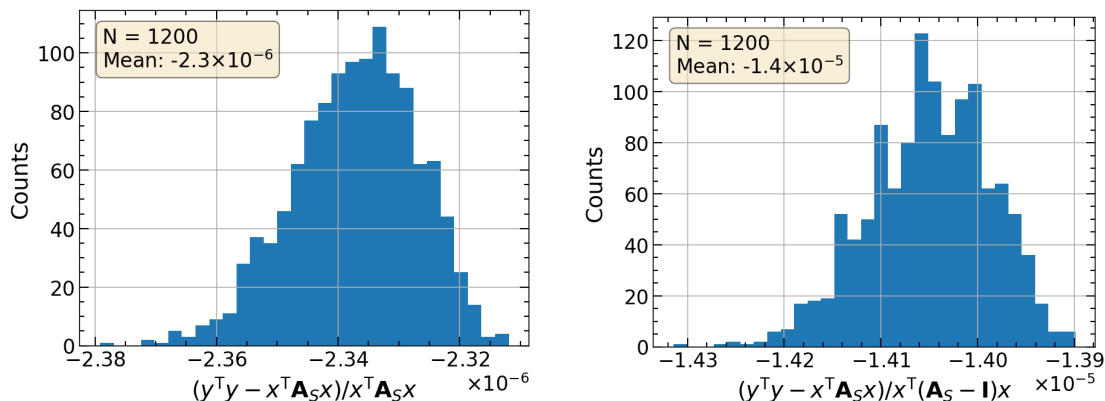
The large-scale component  $\mathbf{q}_L$  is generated via convolution on the mesh by generating white noise in real space and convolving with  $\sqrt{P_L(\mathbf{k})/\Delta V}$  where  $\Delta V$  is the cell volume, and interpolated trilinearly to forest pixel locations.

The large-scale field obtained through convolution on the mesh lacks the small-scale correlations and pipeline noise fluctuations, given by  $\mathbf{A}_S = \mathbf{I} + \mathbf{N}^{-1/2} \mathbf{S}_S \mathbf{N}^{-1/2}$ . To incorporate these fluctuations, we first generate unit Gaussian fields at the forest pixel locations. Denoting this vector as  $\boldsymbol{\eta}$ , we obtain  $\mathbf{q}_S = \mathbf{N}^{1/2} \mathbf{A}_S^{1/2} \boldsymbol{\eta}$ , which has the desired covariance. For numerical reasons,  $\mathbf{N}^{-1/2} \mathbf{q}$  is stored for both cases. This requires us to implement the square root of a matrix ( $\mathbf{A}_S^{1/2}$ ) operating on a vector ( $\boldsymbol{\eta}$ ). There are various iterative algorithms to compute the matrix square root  $\mathbf{Q} = \mathbf{A}_S^{1/2}$ , mainly based on Newton's method and the matrix sign function [72–77]. In appendix A, we review these methods. Among them, the single-iteration Padé approximation of order  $p$  demonstrates the best performance:

$$\mathbf{Q}_p = \frac{\mathbf{A}_S}{p} \sum_{i=1}^p \frac{1}{\xi_i} (\mathbf{A}_S + \alpha_i^2 \mathbf{I})^{-1}, \quad \text{where } \xi_i = \frac{1}{2} \left( 1 + \cos \frac{(2i-1)\pi}{2p} \right) \text{ and } \alpha_i^2 = \frac{1}{\xi_i} - 1, \quad (3.21)$$

where we first shrink the matrix by the largest diagonal,  $\mathbf{A}_S \rightarrow \mathbf{A}_S / \max(A_{S,ii})$ , and later scale up the results by  $\sqrt{\max(A_{S,ii})}$ . While computing  $p$  matrix inverses may initially appear excessively costly, their convergence is accelerated by the addition of  $\alpha_i^2$  to the diagonal. Specifically, for  $p = 4$ , we find all four matrix inverses converge in 42 iterations in total for  $\epsilon = 10^{-5}$ .

To test for the accuracy of this procedure, we draw 1,200 random vectors  $\mathbf{x}$ , calculate the dot product  $\mathbf{y}^T \mathbf{y}$ , where  $\mathbf{y} \equiv \mathbf{Q}_4 \mathbf{x}$ , and compare it to true dot product  $\mathbf{x}^T \mathbf{A}_S \mathbf{x}$ . Figure 1 shows a histogram of the relative error on this dot product. We observe the fourth order to be sufficient with a small mean bias of approximately  $10^{-6}$ . When assessing how well the small-scale signal is captured by removing ones from the diagonal, the mean relative bias slightly



**Figure 1.** (*Left*) A histogram of relative errors in the square root of  $\mathbf{A}_S$  estimate, where  $\mathbf{y} \equiv \mathbf{Q}_4 \mathbf{x}$ . From 1,200 random vectors  $\mathbf{x}$ , the mean relative error indicates a  $-2.3 \times 10^{-6}$  bias. (*Right*) To assess how well the small-scale signal is captured, we subtract the identity matrix from  $\mathbf{A}_S$  and calculate relative errors with respect to  $\mathbf{A}_S - \mathbf{I}$ . The mean relative error bias remains small at  $-1.4 \times 10^{-5}$ .

increases to approximately  $10^{-5}$ . The fifth-order approximation reduces these biases by a factor of 30, but converges in 52 iterations, resulting in a 22% increase in computational cost.

Alternatively, one can define a “noise bias” similar to  $P_{1D}$  formalism  $b_k^N = \text{Tr}(\mathbf{C}^{-1} \mathbf{C}_{,k} \cdot \mathbf{C}^{-1} \mathbf{N})$ , in which case the MC simulations only draw random noise vectors  $\mathbf{n}$  without a need for any signal component, and the quadratic estimator in eq. (3.1) will not have the fiducial  $P_{3D}$  term:

$$\hat{P}_k^{3D} = \sum_{k'} \frac{1}{2} F_{kk'}^{-1} \left[ \mathbf{d}^T \mathbf{C}^{-1} \mathbf{C}_{,k'} \mathbf{C}^{-1} \mathbf{d} - b_{k'}^N \right]. \quad (3.22)$$

However, the Fisher matrix cannot be obtained from these simulations. The same disadvantage applies to other stochastic estimation methods that do not draw from the covariance, such as a vector with random entries of  $\pm 1$  used in ref. [52].

There is a second, but slower method for the Fisher matrix estimation: direct calculation of the trace  $F_{kk'} = \langle \boldsymbol{\eta}^T \mathbf{C}^{-1} \mathbf{C}_{,k} \mathbf{C}^{-1} \mathbf{C}_{,k'} \boldsymbol{\eta} \rangle$  by generating random vectors  $\boldsymbol{\eta}$  of  $\pm 1$ s. This method yields less noisy Fisher matrix estimates even with  $\sim 500$  random vectors, but requires  $N_k$  conjugate gradients and two additional FFTs for every random vector  $\boldsymbol{\eta}$ , which becomes unfeasibly slow unless some time-saving measures are taken. Using the preconditioner (block-diagonal approximation) instead of the conjugate gradient method makes this estimation achievable but fails to capture correlations between multipoles and  $k$  bins.

Lastly, we briefly describe some of our unsuccessful attempts for future reference. One avenue we pursued was based on transforming the equation for the bias to  $b_k = \langle \mathbf{q}'^T \mathbf{C}_{,k} \mathbf{q}' \rangle$ , where  $\mathbf{q}'$  is drawn from  $\mathcal{N}(\mathbf{0}, \mathbf{C}^{-1})$ . There are many algorithms to draw samples from a high-dimensional Gaussian covariance matrix (see ref. [78] for an overview). The conjugate gradient solver can be transformed to a conjugate gradient sampler with minimal change [79]. However, as noted in refs. [78] and [79], this method is only suitable if eigenvalues are well-separated. This is especially problematic for  $\mathbf{C}^{-1}$  in our current formalism since the solver prefers the condition number to be small. The references further note that the performance of the conjugate gradient sampler significantly degrades for vectors drawn from  $\mathbf{C}^{-1}$ . Other

methods might prove more fruitful. We also attempted to draw random vectors from  $\mathbf{C}_k$  to estimate the Fisher matrix directly. This operation requires the square root of the derivative matrix and can be done in Fourier space for the  $(k_\perp, k_\parallel)$  binning scheme. However, the multipole estimation introduces negative values to Fourier modes and cannot be done without supporting imaginary numbers.

### 3.7 Some numerical details

Our initial tests show that shared-memory parallelization through `OpenMP` saturates at 8–16 cores. Most high-performance computing clusters have more cores and have enough memory to run more than one job in a single node. Therefore, we simultaneously run multiple MC simulations by applying MPI parallelization along with `OpenMP`. The synchronization between MPI tasks is not required, so it can be kept at a minimum level. For example, the CCAPP Condo on The Ohio Supercomputer Center (OSC) [80], where we carried out our computations, has 96 cores per compute node. We typically run the estimator using 12 MPI tasks, each utilizing 8 `OpenMP` threads per node.

The integrals and complicated terms such as  $P_L$ ,  $P_{1D}$ , and  $P_{3D}$  are interpolated on a regular grid that eliminates binary search and is preferable over a general grid implementation. In 1D interpolation, a cubic spline offers smoothness and incurs minimal evaluation time. In 2D interpolation for  $\xi_S(r_\perp, r_\parallel)$ , cubic spline becomes costly, but the fastest method, linear interpolation, compromises smoothness. Cubic Hermite spline interpolation achieves smoothness without a hefty computational cost. Specifically, we employ the Bernstein polynomial [81] formulation of cubic Hermite interpolation and evaluate it using de Casteljau’s recursive algorithm [82] (see appendix C and ref. [83] for details).

There is one last computational optimization we apply to correlation function interpolations.  $\xi_S(r_\perp, r_\parallel)$  and  $\xi_{1D}(r_\parallel)$  are smoother and slowly changing in  $\log r$ . However, evaluating  $\log r$  for every pixel pair is unexpectedly expensive. Given that double-precision accuracy for  $\log r$  is unnecessary, we employ a bitwise approximation of  $\log_2 r$ ,<sup>7</sup> achieving nearly an order of magnitude improvement in evaluation speed.

## 4 Validation

We test our estimator on flat-sky mocks. Aside from the simplified geometry, we want to work with a mock catalog relevant to present surveys and, therefore, build a mock catalog with background quasar coordinates and forest pixel locations as well as pixel noise properties drawn from a real survey.

We focus on the value-added Ly $\alpha$  forest catalog, which contains transmitted flux fluctuations  $\delta_F$ , from the 16th data release (DR16) of the complete extended Baryonic Oscillation Spectroscopic Survey (eBOSS) [11]. The dataset includes 210,005 quasars in the redshift range  $2.1 \lesssim z_{\text{qso}} \lesssim 4.0$ , with nearly 85% at  $z_{\text{qso}} < 3.0$  in this dataset. The large-scale structure fluctuations  $\sigma_{\text{LSS}}^2$  are included in the weights in this value-added catalog. Our method requires pipeline noise estimates directly, so we subtract the  $\sigma_{\text{LSS}}^2$  contribution from the weights. The Galactic plane divides the data into two regions: the south Galactic cap (SGC) contains 58,592 quasars, while the north Galactic cap (NGC) holds the remaining 151,413.

<sup>7</sup><https://github.com/romeric/fastapprox>

We find that the data in SGC provides enough statistics to study our algorithm, which requires about 30 node-hours with the configuration detailed below. Although our formalism is efficient, it remains expensive, and testing it with the entire sample increases the cost by approximately a factor of three with minimal return. So, our analysis is restricted to the SGC within the redshift range  $2.1 \leq z_{\text{Ly}\alpha} \leq 2.7$ , corresponding to an effective radial distance of 5.8 Gpc and Cartesian box dimensions of (13.2 Gpc, 4.97 Gpc, 755 Mpc). Using a mesh grid of (2048, 768, 384), the transverse grid spacing is 6.5 Mpc. To realize uniform grid spacing across all dimensions, we extend the radial length to 2.48 Gpc by padding both ends along the line-of-sight direction.

Based on these parameters, we set  $R_L = 6$  Mpc and  $f = 8$  as our baseline. Each forest has an average of 10.3 neighboring forests within a 48 Mpc radius. We estimate  $P_\ell(k)$  for  $l = 0 \dots 8$  in 20 linearly-spaced bins at  $k > 0.015 \text{ Mpc}^{-1}$ , with  $\Delta k = 1.2 \times 10^{-2} \text{ Mpc}^{-1}$ , a scale that is several times larger than the fundamental frequency in the line-of-sight direction.

We set the convergence criteria to  $\epsilon = 10^{-6}$  and run 2,880 MC simulations to estimate the Fisher matrix.

#### 4.1 Self-consistency of Monte Carlo simulations

We generate mock Ly $\alpha$  data using our Monte Carlo (MC) simulation procedure to demonstrate that: (1) our method produces correlated sightlines with the correct  $P_{3\text{D}}$ ; and (2) the estimator successfully recovers the input  $P_{3\text{D}}$ .

##### 4.1.1 Simulations with only Ly $\alpha$ forest

We start by only considering the Ly $\alpha$  forest in the signal. The conjugate gradient method converges in 40 iterations on average.

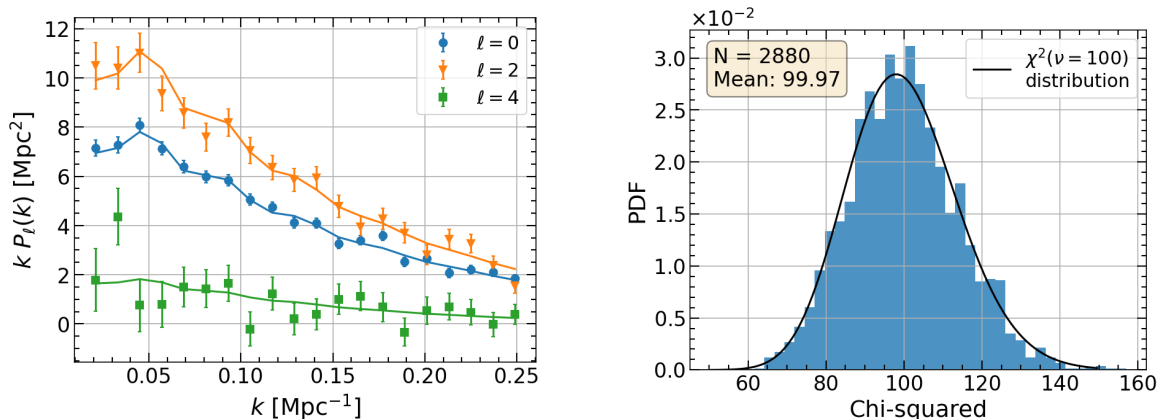
Figure 2 shows multipole estimates from one mock data set on the left panel. All 2,880 MC realizations are independent mock data by themselves. Since these are Gaussian, the calculated  $\chi^2$  follows the theoretically expected  $\chi^2(\nu = 100)$  distribution, which can be seen in the right panel.

Figure 3 shows the correlation matrices for the Fisher matrix on the left and the covariance matrix on the right. The neighboring  $k$  bins are correlated due to linear interpolation, but — as we alluded to in the previous section — all  $k$  bins are correlated, especially bins within the monopole. Correlations between  $k$  bins and multipoles seem weak in the covariance matrix, but the inverse (Fisher) matrix manifests stronger correlations. Since the likelihood calculations are based on the inverse, it is essential to consider these correlations.

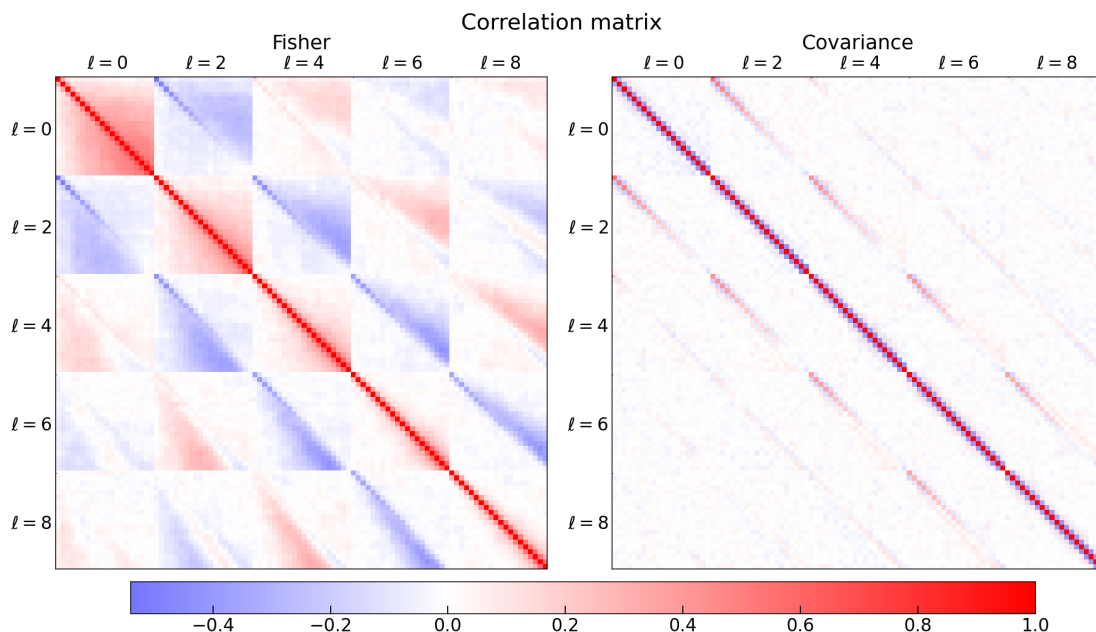
##### 4.1.2 Including HCD and metal systems

We now demonstrate how our prescription for HCDs and ionized silicon systems alters the measurement, with contaminant model parameters noted in section 2.1. The conjugate gradient method converges in 50 iterations on average — a large increase from the previous case.

Figure 4 shows the measured  $P_{3\text{D}}$  multipoles in SGC from one mock data set with contaminants. Due to additional power from HCD and ionized silicon contaminants, the monopole term is now equal to the quadrupole term. As noted in previous sections, the redshift confusion of ionized silicon systems yields a non-negligible  $l = 6$  signal via oscillations in the power spectrum. The SGC quasars alone are not enough to decisively detect  $P_6$ .



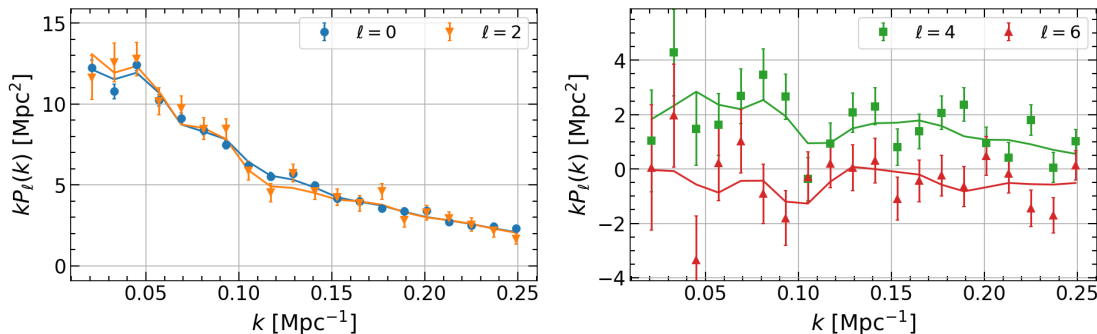
**Figure 2.** (Left)  $P_{3D}$  multipole estimations in SGC from one MC realization without any contaminants. Ly $\alpha$  forest quadrupole is stronger than its monopole.  $l = 6$  multipole is effectively zero and not shown for clarity. (Right) Histogram of  $\chi^2$ s from 2,880 MC realizations. The probability distribution function follows the theoretically expected  $\chi^2(\nu = 100)$  distribution.



**Figure 3.** Fisher correlation matrix on the left and correlation matrix between multipoles on the right. Correlations between  $k$  bins and multipoles seem weak in the covariance matrix, but the Fisher matrix manifests stronger correlations. These should be taken into account in likelihood calculations for cosmological inference.

### 4.1.3 Continuum marginalization

A pillar of optimal estimators is the marginalization of contaminated modes — in our case, quasar continuum fitting errors. We account for typical errors introduced by the eBOSS and DESI pipelines, modeled as amplitude and slope terms for each forest [11, 84]. However, continuum error templates can be adapted for novel approaches, including machine learning and neural network-based methods [85, 86].



**Figure 4.** Measured  $P_{3D}$  multipoles in SGC from one MC realization with HCD and ionized silicon contaminants. The monopole strengthens relative to the quadrupole due to contributions from contaminants. The oscillations induced by Si II and Si III generate a non-negligible  $l = 6$  signal, though the constraining power of SGC forests is insufficient for decisive detection.

Our algorithm requires eigenvalue decomposition for each forest, along with storing and retrieving a projection matrix for every quasar from the disk — making implementation the primary challenge rather than effectiveness. Here, we focus on the implementation and its overall impact, while the efficacy of the method is discussed in section 4.2.2.

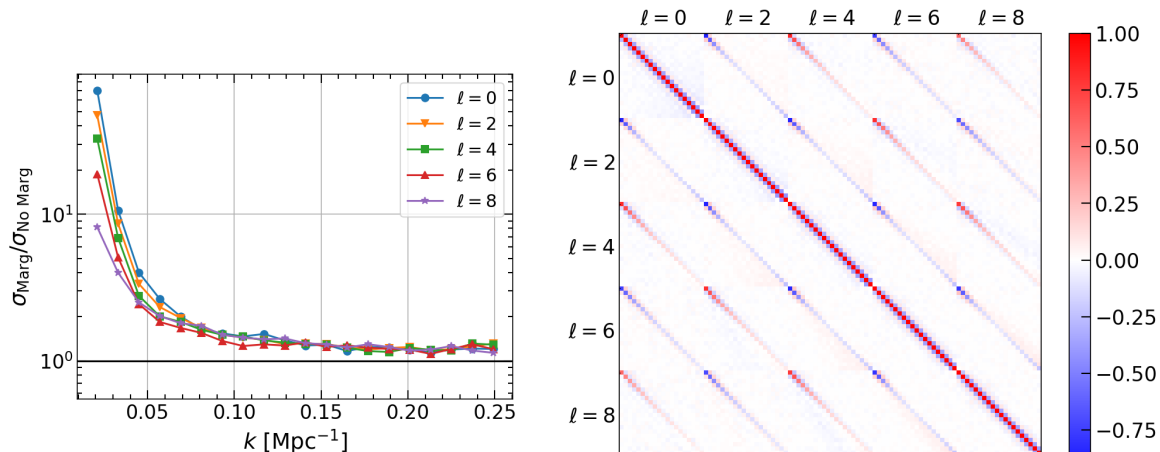
We assess the computational feasibility and convergence behavior of the conjugate gradient method using the same setup as before, but without any contamination or added continuum errors. Each multiplication with the matrix  $\mathbf{A}$  requires two disk accesses per forest. Despite this significant requirement, reading the matrix  $\mathbf{\Pi}$  and multiplying it with vectors constitute only 3.3% of the total computational time for the  $\mathbf{Az}$  operation. This efficiency is thanks to OSC’s high-performance disk space per compute node and our use of raw binary files, which can be read serially by each `OpenMP` thread.

The effect of continuum marginalization on  $P_{3D}$  mirrors its effect on  $P_{1D}$ : it inflates errors in low- $k$  modes. Specifically, this negative impact would have been localized to low  $k_{\parallel}$  modes in the  $(k_{\perp}, k_{\parallel})$  binning scheme [87]. In the multipole expansion, the effect of marginalization on these modes spreads over a range of  $k$  bins. This can be seen on the left panel of figure 5. For the lowest bin at  $k = 2.1 \times 10^{-2} \text{ Mpc}^{-1}$ , the errors increase by a factor of nearly 35 across all multipoles. For the next bin at  $k = 3.3 \times 10^{-2} \text{ Mpc}^{-1}$ , the errors increase by a factor of 7 compared to the case without continuum marginalization. For modes with  $k \gtrsim 5.0 \times 10^{-2} \text{ Mpc}^{-1}$ , the errors are 30% larger on average.

Additionally, the correlations between multipoles are amplified due to the marginalized modes. The right panel of figure 5 illustrates the correlation coefficient derived from the covariance matrix. In contrast to the unmarginalized case, the positive correlations between  $\delta\ell = 2$  modes transition to anti-correlations, and correlations generally extend to  $\delta\ell = 8$  and, most likely, even larger  $\delta\ell$ s.

## 4.2 Gaussian simulations

In this section, we evaluate the validity of our formalism using an externally generated data vector. Preliminary analysis based on eBOSS data and mocks revealed that projections from spherical to Cartesian geometry introduce significant distortions in the measurements.



**Figure 5.** (Left) Increase in error after continuum marginalization. Errors in low- $k$  modes are significantly inflated. (Right) Correlation matrix between multipoles in the continuum marginalization case. In contrast to the unmarginalized case, the positive correlations between  $\delta\ell = 2$  modes transform to anti-correlations, and correlations extend to higher  $\delta\ell$ .

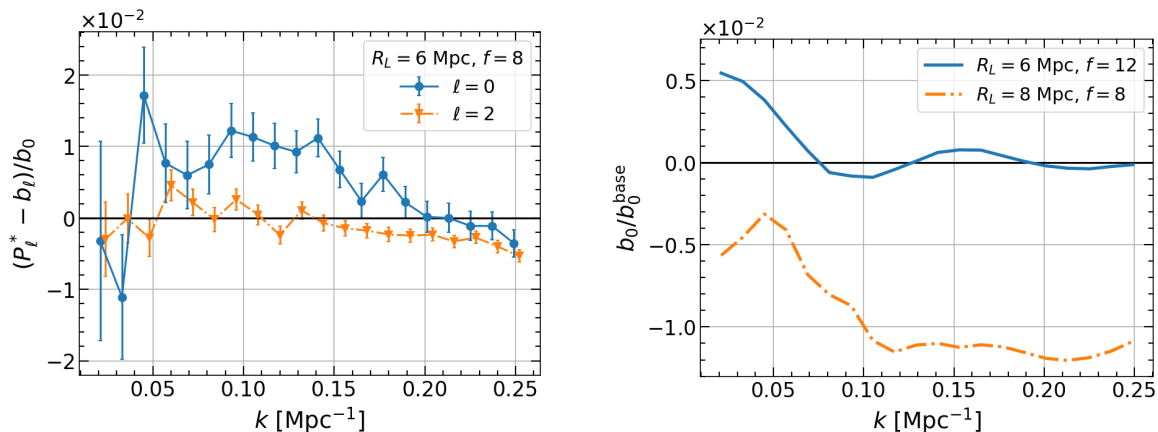
Therefore, we restrict our external data set to cases without spherical geometry and defer a comprehensive treatment of spherical geometry to future work.

We generate a random Gaussian field on a Cartesian grid with three times finer spacing, convolved with the total Ly $\alpha$  power spectrum, and interpolate this field to the eBOSS Ly $\alpha$  forest data locations. We then add Gaussian noise consistent with the pipeline to the synthetic data. This approach captures the eBOSS survey window function without introducing distortions.

As discussed in section 2.1,  $P_{3D}$  extends to smaller angular scales compared to radial scales, as they are constrained only by the pressure-smoothing scale. Without appropriate suppression, the angular modes alias in a grid of 2.2 Mpc spacing. To prevent this problem, we adopted an additional isotropic suppression scale of  $k_A = 0.4 \text{ Mpc}^{-1}$ .

Across all tests, we find the estimator to be biased at the percent level, leading to high chi-squared values. This bias is illustrated for the baseline setting in the left panel of figure 6 as the error in the unnormalized multipole  $P_\ell^*(k) = \mathbf{d}^T \mathbf{C}^{-1} \mathbf{C}_{,\ell k} \mathbf{C}^{-1} \mathbf{d}$  and total bias  $b_\ell(k) = \text{Tr}(\mathbf{C}^{-1} \mathbf{C}_{,\ell k})$  estimates relative to  $b_0$  (these are the terms inside parentheses in eq. (3.1)). According to the Fisher matrix, these points are highly correlated yet show a clear bias in the estimator.

There are two possible sources for this bias. (1) A consistent hard boundary between large and small scales cannot be imposed in a P<sup>3</sup>M code. Mathematically, all quasar pairs have to be included for the decomposition to be exact ( $f \rightarrow \infty$ ). The missing quasar pairs due to a finite choice of  $f$  could bias the estimator. We investigate this below. (2) The density field has exactly the correct  $P_{3D}$  on the mesh, regularly sampled points, which are interpolated to data locations. The unsmoothness of trilinear interpolation introduces an anisotropic suppression of power and ringing to the interpolated density field [88]. These dominate at the Nyquist frequency  $k_{\text{Nyq}} \approx 3 \text{ Mpc}^{-1}$ ; however, they could leave percent-level artifacts on the scales we consider here. The investigation of higher order interpolation kernels is left to future work.



**Figure 6.** (Left) Error in unnormalized multipole  $P_\ell^*$  and total bias  $b_\ell$  estimates (as defined in the text) relative to  $b_0$  for the baseline setting  $R_L = 6$  Mpc and  $f = 8$ . Note that these points are highly correlated according to the Fisher matrix. Nevertheless, the estimator seems to be biased at the percent level. (Right) Estimated total monopole bias  $b_0$  for different large and small scale settings relative to the baseline monopole bias  $b_0^{\text{base}}$ . Increasing the neighboring radius factor from  $f = 8$  to  $f = 12$  changes the total bias estimate at a sub-percent level on large scales. Increasing the smoothing scale  $R_L$  to 8 Mpc decreases  $b_0$  on all scales by 1% on average.

In real-data applications, tuning the estimator’s parameters will be necessary. The error can be corrected using simulated data or mitigated by adjusting the estimator configuration.

#### 4.2.1 Varying large- and small-scale split factors

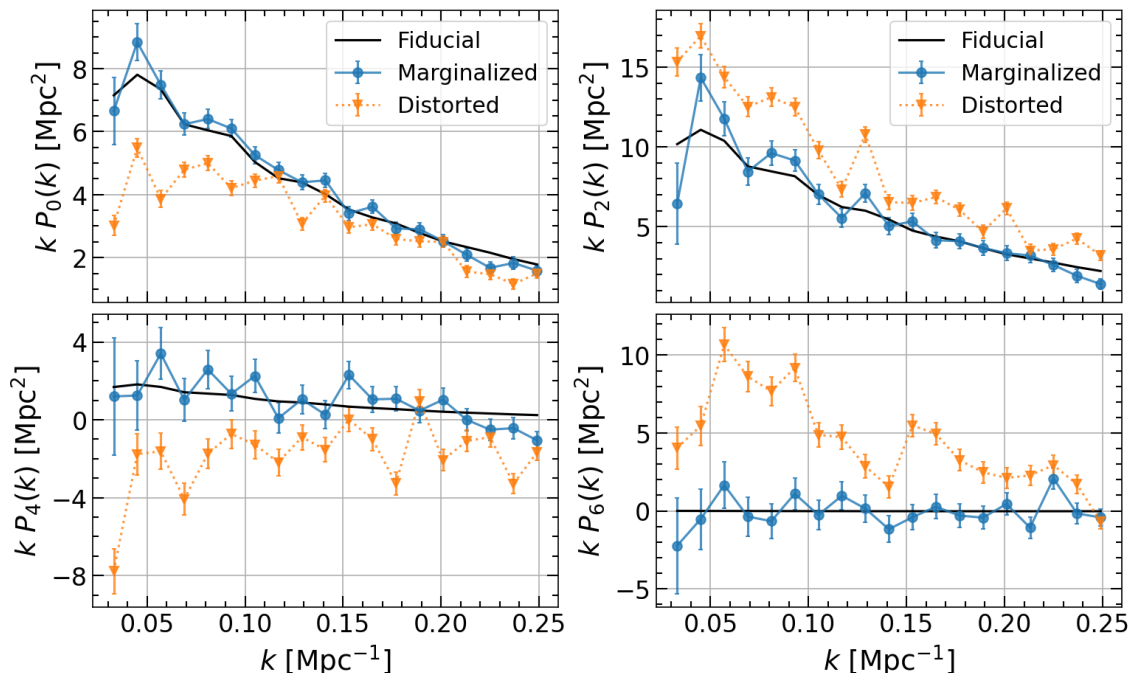
We conduct two variations on the baseline setting to evaluate the impact of hyperparameters: the large-scale radius  $R_L$  and the neighboring radius factor  $f$ . The impact of these variations on the total monopole bias  $b_0$  is shown in the right panel of figure 6.

Keeping the large-scale radius fixed while increasing the neighboring radius factor to  $f = 12$  raises the average number of neighbors to 23 and changes  $b_0$  at the sub-percent level on large scales. This change on large scales is expected since correlations from farther forests contribute to the bias calculation. However, the increased number of neighbors substantially impacts the computation time, increasing it by a factor of 3.5. Though the chi-squared value slightly decreases to 185 from 192 per 100 degrees of freedom, it remains high.

Increasing the large-scale radius to  $R_L = 8$  Mpc while keeping the neighboring radius factor fixed raises the average number of neighbors to 18 and decreases  $b_0$  by approximately 1% across all scales. The chi-squared value becomes even larger. This indicates that the PM component captures non-negligible small-scale information.

#### 4.2.2 Adding continuum errors and assessing its impact

We evaluate the efficacy of continuum marginalization using Gaussian simulations by emulating the impact of continuum fitting on the forest spectrum with a simple projection. Continuum marginalization has proven effective for the 1D power spectrum [51, 89]; however, the continuum fitting procedure itself introduces higher-order correlations, which we do not



**Figure 7.** The distorted  $P_{3D}$  (orange dotted lines with triangles) after projecting out the mean and the slope of  $\delta_F$ . The distortions suppress the monopole on large scales, boost the quadrupole on all scales, and extend power to higher-order multipoles ( $\ell \geq 6$ ). The continuum marginalized  $P_{3D}$  (blue solid lines with circles) agrees with the input fiducial model (black solid lines).

address here. Simulating realistic quasar continua and reproducing the eBOSS quasar continuum fitting pipeline are beyond the scope of this work.

The eBOSS continuum fitting is performed in the forest region following methodologies developed in refs. [8, 9], which has also been applied to 1D studies [89–91]. In this method, the definition of the quasar continuum includes the mean transmission  $\bar{F}(z)$  of the IGM, and each quasar “continuum”  $\bar{F}C_q(\lambda_{RF})$  is modeled by a first-order polynomial scaling of a global continuum:  $\bar{F}C_q(\lambda_{RF}) = \bar{C}(\lambda_{RF})(a_q + b_q\Lambda)$ , where  $\lambda_{RF}$  is the wavelength in the quasar’s rest frame and  $\Lambda \sim \ln \lambda_{RF}$ . Since the fitting is done in the forest region, the resulting forest has, to a good approximation, a zero mean  $\langle \delta_F \rangle = 0$  and a zero slope  $\langle \Lambda \delta_F \rangle = 0$ .

We project out these modes from  $\delta_F$  in every forest to imitate continuum fitting errors. This projection is coupled to the underlying large-scale density mode and introduces distortions to 3D correlations. Figure 7 illustrates the distorted multipoles in orange dotted lines. This parallels the findings of ref. [33]: the monopole is suppressed on large scales, and the quadrupole is boosted on all scales. Additionally, distortions spread spurious signals to higher-order multipoles, which is undesired as it makes theoretical modeling difficult. As an example, we plot  $P_{\ell=6}(k)$  in the bottom right panel of figure 7.

The continuum marginalization, described in section 3.4, removes these distortions from the measurement, shown in blue solid lines in figure 7. The first three main multipoles now agree with the input model, and the large distorted signal of  $\ell = 6$  is completely removed. However, as established previously with MC simulations, this marginalization diminishes the precision of measured  $P_{3D}$  and correlates higher-order multipoles.

## 5 Discussion

We have developed an optimal estimator for the 3D power spectrum ( $P_{3D}$ ) of the Ly $\alpha$  forest, designed to overcome the unique challenges posed by its highly anisotropic, uneven sampling and continuum fitting errors that distort the two-point functions. Our approach builds on optimal estimator techniques of cosmic microwave background (CMB) data. It is statistically optimal (for Gaussian fields; in the presence of non-Gaussianities, it remains unbiased but is not, strictly speaking, optimal), deconvolves the survey window function, and marginalizes contaminated modes. The current implementation works in the flat-sky approximation.

To make the optimal estimator computationally feasible for large datasets, we implemented a particle-particle/particle-mesh (P<sup>3</sup>M) decomposition of the covariance matrix. This allows us to efficiently handle large-scale correlations using fast Fourier-space convolution while computing small-scale correlations via direct summation. We validated our implementation using Monte Carlo mocks and Gaussian simulations, and demonstrated that our estimator is computationally viable, albeit slightly sensitive to hyperparameter choices such as the large-scale split scale  $R_L$  and neighboring radius factor  $f$ . We found the estimator to be biased at the percent level, which must be corrected using simulated data or mitigated by tuning the estimator configuration in real data applications. Higher order interpolation kernels could reduce this bias [88], which we leave to future work.

The current estimator successfully handles several complexities of the Ly $\alpha$  forest. We addressed redshift evolution by computing the signal matrix at a pivot redshift and applying a separable redshift growth function. By propagating this to derivative matrices, we were able to estimate  $P_{3D}$  at the pivot redshift, removing the mixing of different time periods. Our redshift evolution treatment handles only the amplitude, which is valid for linear scales, but invalid for non-linear scales. However, these non-linear scales are better probed by densely sampled surveys and by using a higher-resolution mesh with a shorter radial distance, in which case the error of this simplified treatment will be alleviated. One could overcome this shortcoming by using multiple grids at different redshifts and interpolating them. Additionally, our single-grid approach assumes metals and DLAs grow similarly to the Ly $\alpha$  forest, which may require refinement in future studies, such as employing separate grids for these systems.

We also implemented marginalization over quasar continuum-fitting degeneracies in the weighting of the estimator rather than on the data vector itself. This means that we do not have to include the continuum-fitting distortion of the theory; instead, the loss of information manifests as increased errors on large scales and increased correlations between multipoles. Specifically, the positive correlations between  $\delta\ell = 2$  modes change to anti-correlations, and correlations extend to  $\delta\ell \geq 8$  after mode marginalization.

Our estimator is most effective at recovering large-scale information when applied to sparsely sampled large-scale structure surveys. Accurate access to these large-scale modes enables robust interpretations of Alcock-Paczyński distortions, ionizing background fluctuations, and signatures of primordial non-Gaussianity. In addition, BAO can be measured in power spectrum space using a peak/smooth template decomposition, complementing Ly $\alpha$  forest correlation function analyses [63, 92]. The BAO feature is delocalized in Fourier space and appears as oscillations that must be densely sampled in  $k$  bins to avoid smearing. A key advantage of an optimal power spectrum measurement is its robustness against survey

window complications, which pose an important challenge in galaxy power spectrum BAO analyses [93].

Our next step toward application to real data will be to implement spherical geometry. Initial analyses of eBOSS data and mocks that are generated on a spherical geometry revealed that the estimator is strongly biased by geometric distortions. These effects, coupled with continuum error modes, contaminate the measurement beyond recovery. Though tuning the estimator settings, such as estimating  $P_{3D}$  in smaller sky patches and narrower redshift bins, can mitigate this effect, we deem reformulating the estimator in spherical geometry as the most robust solution.

The generalization to a spherical sky will require significant additions to the PM part of the code (and more modest changes to the  $P^2$  algorithm). Spherical geometry requires the replacement of the 3D FFTs with spherical harmonics for the angular directions and spherical Bessel functions for the radial directions. The 3D Fourier transform has cost  $\mathcal{O}(N \log N)$ , where  $N = N_x N_y N_z$  is the number of mesh points. The spherical convolution on a mesh requires several steps. First is an FFT in the longitude direction to replace right ascension  $\phi$  with angular momentum quantum number  $m$ , with cost  $\mathcal{O}(N_r N_\theta N_\phi \log N_\phi)$ , where  $N_r$ ,  $N_\theta$ , and  $N_\phi$  are the number of mesh points in the radial, declination, and right ascension directions, respectively. Second is the replacement of declination  $\frac{\pi}{2} - \theta$  with angular momentum quantum number  $\ell$ , with cost  $\mathcal{O}(N_r N_\theta^2 N_\phi)$  if done using standard recursion relations (as done in, e.g., the spherical PM implementation in SDSS analyses [49]). Then comes the transformation of radial distance  $r$  to wavenumber  $k$ : this has cost  $\mathcal{O}(N_r^2 N_\theta N_\phi)$  if done by direct multiplication by the spherical Bessel functions  $j_\ell(kr)$ . In principle, one could use a logarithmic mesh in  $r$  and do an FFT-based convolution in the radial direction [68], but given the narrow range in  $\log r$  probed by the Ly $\alpha$  forest, it is not clear whether we want to pursue this approach. The inclusion of redshift-space distortions also requires multiplication by matrices involving the derivatives of  $j_\ell(kr)$  [94] or equivalently the adjacent  $j_{\ell\pm 2}(kr)$  [95], which fortunately can be computed from the same recursion relations as the  $j_\ell$ 's. If  $N_r$ ,  $N_\theta$ , and  $N_\phi$  are of the same order of magnitude, then, we expect the spherical implementation to have cost  $\mathcal{O}(N^{4/3})$ , which is slower than the FFT but by the modest and (hopefully) manageable factor  $\sim N^{1/3}/\log N$ .

The estimator's sensitivity to the large-scale split scale  $R_L$  and the search radius for neighbors  $fR_L$  necessitates careful calibration through simulations tailored to the specific dataset. Specifically, the parameter  $f$  describes the overlap region between the large- and small-scale treatments. There must be some transition region in any  $P^3M$  code since the large scales are handled in Fourier space, the small scales in configuration space, and one cannot consistently impose a hard boundary in both spaces. However, increasing  $f$  significantly raises computational costs, making it impractical beyond certain limits. On the other hand, performing the PP component on GPUs could substantially improve the efficiency of the small-scale matrix-vector multiplications and allow more distant neighbors to be included. Exploring GPU implementations to optimize performance is a promising avenue for future work.

One possible drawback of our optimal estimator is that it auto-correlates spectra by construction in contrast to the estimators of refs. [31, 33]. The auto-correlations lead to added spectral noise power and are susceptible to pipeline calibration errors and sky modeling errors [96]. Furthermore, residuals in the sky background subtraction model correlate pixels between neighboring quasars observed in the same spectrograph [96]. The cosmology inference

must include templates to marginalize errors in noise calibration and sky modeling. We show in appendix B how the estimator can be altered to calculate only cross-forest contributions, and discard some combinations of forest pairs, albeit sub-optimally.

This work represents a significant step forward in the precise measurement of  $P_{3D}$ , paving the way for more accurate cosmological constraints from upcoming large-scale structure surveys like DESI and Spec-S5 [97].

## A Matrix square root $\mathbf{Q} = \mathbf{A}^{1/2}$

Ref. [76] is a great resource on matrix square root algorithms, and we have heavily relied on it below.

### A.1 Newton iteration

The ordinary Newton iteration is as follows:  $\mathbf{Q}_{i+1} = (\mathbf{Q}_i + \mathbf{A}\mathbf{Q}_i^{-1})/2$ . Starting the iterations with the block-diagonal matrix,  $\mathbf{Q}_0 = \mathbf{A}_{\text{BD}}^{1/2}$ , we arrive at the following expression for the second iteration:

$$\mathbf{Q}_2 = \frac{1}{4}\mathbf{Q}_0 + \mathbf{H}\mathbf{Q}_0 \left( \frac{1}{4}\mathbf{I} + (\mathbf{I} + \mathbf{H})^{-1} \right), \quad (\text{A.1})$$

defining  $\mathbf{H} \equiv \mathbf{A}\mathbf{A}_{\text{BD}}^{-1}$ . We find that the matrix  $\mathbf{H}$  has a small condition number. The inverse matrix  $(\mathbf{I} + \mathbf{H})^{-1}$  can be solved with the conjugate gradient method, which has a faster convergence property than the full covariance matrix, thanks to the additional identity matrix. Note that  $\mathbf{H}$  is not symmetric. Nevertheless, the conjugate gradient method rapidly converges. Unfortunately, subsequent iterations require nested matrix inverses and are prohibitively expensive. Furthermore, this formulation of the Newton iteration is actually unstable [76]. Both of these factors render this method impractical and unworthy of further pursuit.

### A.2 Newton-Schulz iteration

The Newton-Schulz (NS) iteration is a commonly used alternative that is more stable but requires two coupled matrices [72, 76, 98, 99]. Starting from  $\mathbf{Q}_0 = \mathbf{A}$  and  $\mathbf{Z}_0 = \mathbf{I}$ , the coupled iteration is as follows:

$$\mathbf{Q}_{i+1} = \frac{1}{2}\mathbf{Q}_i(3\mathbf{I} - \mathbf{Z}_i\mathbf{Q}_i), \quad \mathbf{Z}_{i+1} = \frac{1}{2}(3\mathbf{I} - \mathbf{Z}_i\mathbf{Q}_i)\mathbf{Z}_i. \quad (\text{A.2})$$

Provided proper normalization (e.g.,  $\mathbf{A} \rightarrow \mathbf{A}/\lambda_{\text{max}}$ ,  $\mathbf{A} \rightarrow \mathbf{A}/\max(A_{ii})$  or  $\mathbf{A} \rightarrow \mathbf{A}/\|\mathbf{A}\|_F$ ), the NS iteration has quadratic convergence of  $\mathbf{Q}_i$  to  $\mathbf{A}^{1/2}$  and  $\mathbf{Z}_i$  to  $\mathbf{A}^{-1/2}$ . Unlike the ordinary Newton iteration, the initial values cannot be altered in the NS method. Additionally, the NS method does not involve inverting matrices. But this advantage disappears as we cannot store either  $\mathbf{Q}_i$  or  $\mathbf{Z}_i$  matrices. We implemented the NS iteration as recursively nested multiplications, storing the intermediate resulting vectors. For  $N_i$  iterations, this requires  $2 \times N_i$  additional temporary vectors. For low values of  $N_i$ , the additional memory requirement remains manageable. However, the number of matrix multiplications increases rapidly with each iteration, reaching 365 multiplications for  $N_i = 6$ .

### A.3 Padé iteration

The Padé iteration is also derived from the matrix sign function like NS iteration [73–76]. For a Padé approximation order  $p$ , the coupled iteration for the matrix square root is the following:

$$\mathbf{Q}_{i+1} = \frac{1}{p} \mathbf{Q}_i \sum_{i=1}^p \frac{1}{\xi_i} (\mathbf{Z}_i \mathbf{Q}_i + \alpha_i^2 \mathbf{I})^{-1}, \quad \mathbf{Z}_{i+1} = \frac{1}{p} \mathbf{Z}_i \sum_{i=1}^p \frac{1}{\xi_i} (\mathbf{Q}_i \mathbf{Z}_i + \alpha_i^2 \mathbf{I})^{-1}, \quad (\text{A.3})$$

where  $\xi_i = (\cos((2i-1)\pi/2p) + 1)/2$  and  $\alpha_i^2 = 1/\xi_i - 1$ . The iterations start from  $\mathbf{Q}_0 = \mathbf{A}$  and  $\mathbf{Z}_0 = \mathbf{I}$  and cannot be altered like the NS method. The nested matrix inverses are still prohibitively expensive; however, unlike ordinary Newton iteration, the accuracy of the Padé approximation can be increased arbitrarily by  $p$ . Additionally, increasing  $p$  does not require additional memory. While computing  $p$  matrix inverses may initially appear excessively costly, their convergence is accelerated by the addition of  $\alpha_i^2$  to the diagonal. We find that for  $p = 4$ , all four matrix inverses converge in fewer than 40 iterations in total for  $\epsilon = 10^{-5}$ . Ref. [76] demonstrates that  $p = 4$  converges within a few iterations across all scenarios, further supporting our single-iteration formulation.

## B Quasi-optimal cross-forest estimator

QMLE formalism can be extended to construct an unbiased cross-forest estimator that partially preserves inverse covariance weighting. However, this estimator does not inherit the theoretical minimum variance property of the full optimal estimator as discussed also in the context of CMB anisotropies [100].

First, to prevent different forests from mixing, the inverse covariance weighting must be isolated to individual forests through block-diagonal approximation. This is a common approximation in previous  $\text{Ly}\alpha$  analyses [4, 101] but leads to sub-optimality. Second, only different sightlines must be correlated in the derivative matrix operation. Putting these together, we can start formulating the estimator as follows:

$$y_k^* = \frac{1}{2} \mathbf{d}^T \mathbf{C}_{\text{BD}}^{-1} \mathbf{C}_{,k}^X \mathbf{C}_{\text{BD}}^{-1} \mathbf{d} - b_k^X, \quad (\text{B.1})$$

where  $\mathbf{C}_{,k'}^X$  is non-zero only for off-block-diagonal terms and symmetric. This is unbiased if  $\langle y_k^* \rangle = 0$ , which means

$$b_k^X = \frac{1}{2} \text{Tr} \left( \mathbf{C}_{\text{BD}}^{-1} \mathbf{C}_{,k}^X \mathbf{C}_{\text{BD}}^{-1} \mathbf{C} \right). \quad (\text{B.2})$$

Then, we can write the quasi-optimal cross-forest estimator analogous to eq. (3.1):

$$\hat{P}_k^{3\text{D}} = P_k^{3\text{D},\text{fid}} + \sum_{k'} \frac{1}{2} F_{kk'}^{-1} \left[ \mathbf{d}^T \mathbf{C}_{\text{BD}}^{-1} \mathbf{C}_{,k'}^X \mathbf{C}_{\text{BD}}^{-1} \mathbf{d} - b_{k'}^X \right], \quad (\text{B.3})$$

and

$$F_{kk'} = \frac{1}{2} \text{Tr}(\mathbf{C}_{\text{BD}}^{-1} \mathbf{C}_{,k}^X \mathbf{C}_{\text{BD}}^{-1} \mathbf{C}_{,k'}^X). \quad (\text{B.4})$$

However, the covariance matrix of the estimated  $\hat{P}_k^{3\text{D}}$  is not simply the inverse Fisher matrix. Denoting the covariance of  $y_k^*$  as  $V_{kk'}^{(y)} \equiv \langle y_k^* y_{k'}^* \rangle$ , the covariance of estimated  $\hat{P}_k^{3\text{D}}$  is

given by  $\mathbf{V}^{(P)} = \mathbf{F}^{-1}\mathbf{V}^{(y)}\mathbf{F}^{-1}$ . The stochastic trace estimation of  $b_k^X$  will give  $\mathbf{V}^{(y)}$  for free as before. The random vectors with correlated sightlines can be drawn with our prescription in section 3.6. In case direct estimation is desired,  $\mathbf{V}^{(y)}$  is analytically given by

$$V_{kk'}^{(y)} = 2 \text{Tr}(\mathbf{C}\mathbf{E}_k\mathbf{C}\mathbf{E}_{k'}) \quad (\text{B.5})$$

under the Gaussianity assumption, where we defined  $\mathbf{E}_k \equiv (1/2)\mathbf{C}_{\text{BD}}^{-1}\mathbf{C}_{,k}^X\mathbf{C}_{\text{BD}}^{-1}$  for clarity.

This estimator does not require a conjugate gradient method solution, which is a significant speed improvement over the optimal estimator. However, the derivative matrix operation can be unfeasible if every forest pair needs to be multiplied. This operation can be performed efficiently by utilizing  $\mathbf{C}_{,k}^X = \mathbf{C}_{,k}^{\text{Total}} - \mathbf{C}_{,k}^{\text{BD}}$ , such that the operation  $\mathbf{C}_{,k}^{\text{Total}}$  can be performed using the PM formalism from which the block-diagonal contributions can be subtracted. This formalism can be extended to explicitly remove some combinations of forest pairs, such as neighboring quasars of the same spectrograph, from  $\mathbf{C}_{,k}^X$ .

## C Hermite interpolation details

The most publicly accessible reference for cubic Hermite interpolation can be found online in Wikipedia, the free encyclopedia [102]. Here we lay out our formulation and cite the historical references that contribute to the efficient calculation of this interpolation scheme. A central reference we recommend is a book by Farin [83]. We will assume the input points  $x_k$  are regularly spaced with  $\Delta x$  spacing. The interpolating cubic Hermite polynomial between points  $x_k$  and  $x_{k+1}$  is given by

$$y(x) = y_k B_0^3(t) + (y_k + y'_k \Delta x/3) B_1^3(t) + (y_{k+1} - y'_{k+1} \Delta x/3) B_2^3(t) + y_{k+1} B_3^3(t), \quad (\text{C.1})$$

where  $t \equiv (x - x_k)/\Delta x$  and  $B_i^n(t) = \binom{n}{i} t^i (1-t)^{n-i}$  are the Bernstein polynomials [81]. As we will soon highlight, these polynomials are not explicitly evaluated. We use the finite-difference method for the derivatives, which becomes a Catmull-Rom spline [103] for regular grids:  $y'_k = (y_{k+1} - y_{k-1})/2\Delta x$ . Putting this into the equation above, we see that the four nearest points of the input are used in evaluation.

This polynomial can be efficiently evaluated using de Casteljau's recursive algorithm [82] as follows. First, we write  $y(x) = \sum_{i=0}^3 \beta_i B_i^3(t)$ , where  $\beta_i$  are the coefficients of the interpolating polynomial. The algorithm recursively updates these coefficients. For iteration  $r$ , the new coefficients are  $\beta_i^r = \beta_i^{r-1}(1-t) + \beta_{i+1}^{r-1}t$  where  $i = 0, \dots, 3-r$  and  $r = 1, 2, 3$ . The interpolation result is  $y(x) = \beta_0^3$ . This efficient algorithm performs six operations, each with two multiplications and one addition, and does not evaluate Bernstein polynomials.

## Acknowledgments

We thank Nikhil Padmanabhan for the fruitful discussion.

N.G.K. acknowledges support from the United States Department of Energy, Office of High Energy Physics under Award Number DE-SC0011726.

During the preparation of this work, C.H. was supported by the David & Lucile Packard Foundation; and by the National Aeronautics and Space Administration (grant 22-ROM-AN11-0011).

**Software.** Our estimator is written in `c++`.<sup>8</sup> It depends on CBLAS and LAPACKE routines for matrix/vector operations, `GSL`<sup>9</sup> for certain scientific calculations [104], `FFTW3`<sup>10</sup> for fast Fourier transforms [105]; and uses the Message Passing Interface (MPI) standard<sup>11</sup> to parallelize tasks. We use the following commonly-used software in `python` analysis: `astropy`<sup>12</sup> a community-developed core `python` package for Astronomy [106–108], `numpy`<sup>13</sup> an open source project aiming to enable numerical computing with `python` [109], `scipy`<sup>14</sup> an open-source project with algorithms for scientific computing. Finally, we make plots using `matplotlib`<sup>15</sup> a comprehensive library for creating static, animated, and interactive visualizations in `python` [110].

## References

- [1] M. Schmidt, *Large Redshifts of Five Quasi-Stellar Sources*, *Astrophys. J.* **141** (1965) 1295.
- [2] J.E. Gunn and B.A. Peterson, *On the Density of Neutral Hydrogen in Intergalactic Space*, *Astrophys. J.* **142** (1965) 1633 [INSPIRE].
- [3] A. Slosar et al., *The Lyman- $\alpha$  forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, *JCAP* **09** (2011) 001 [arXiv:1104.5244] [INSPIRE].
- [4] A. Slosar et al., *Measurement of baryon acoustic oscillations in the Lyman- $\alpha$  forest fluctuations in BOSS data release 9*, *JCAP* **04** (2013) 026 [arXiv:1301.3459] [INSPIRE].
- [5] N.G. Busca et al., *Baryon Acoustic Oscillations in the Ly- $\alpha$  forest of BOSS quasars*, *Astron. Astrophys.* **552** (2013) A96 [arXiv:1211.2616] [INSPIRE].
- [6] A. Font-Ribera et al., *Quasar-Lyman  $\alpha$  Forest Cross-Correlation from BOSS DR11: Baryon Acoustic Oscillations*, *JCAP* **05** (2014) 027 [arXiv:1311.1767] [INSPIRE].
- [7] T. Delubac et al., *Baryon acoustic oscillations in the Ly $\alpha$  forest of BOSS DR11 quasars*, *Astron. Astrophys.* **574** (2015) A59 [arXiv:1404.1801] [INSPIRE].
- [8] J.E. Bautista et al., *Measurement of baryon acoustic oscillation correlations at  $z = 2.3$  with SDSS DR12 Ly $\alpha$ -Forests*, *Astron. Astrophys.* **603** (2017) A12 [arXiv:1702.00176] [INSPIRE].
- [9] M. Blomqvist et al., *Baryon acoustic oscillations from the cross-correlation of Ly $\alpha$  absorption and quasars in eBOSS DR14*, *Astron. Astrophys.* **629** (2019) A86 [arXiv:1904.03430] [INSPIRE].
- [10] V. de Sainte Agathe et al., *Baryon acoustic oscillations at  $z = 2.34$  from the correlations of Ly $\alpha$  absorption in eBOSS DR14*, *Astron. Astrophys.* **629** (2019) A85 [arXiv:1904.03400] [INSPIRE].
- [11] H. du Mas des Bourboux et al., *The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Ly $\alpha$  Forests*, *Astrophys. J.* **901** (2020) 153 [arXiv:2007.08995] [INSPIRE].

---

<sup>8</sup><https://github.com/p-slash/lyspeq>

<sup>9</sup><https://www.gnu.org/software/gsl>

<sup>10</sup><https://fftw.org>

<sup>11</sup><https://www.mpich.org>

<sup>12</sup><https://www.astropy.org>

<sup>13</sup><https://numpy.org>

<sup>14</sup><https://scipy.org>

<sup>15</sup><https://matplotlib.org>

- [12] A. Cuceu, A. Font-Ribera, S. Nadathur, B. Joachimi and P. Martini, *Constraints on the Cosmic Expansion Rate at Redshift 2.3 from the Lyman- $\alpha$  Forest*, *Phys. Rev. Lett.* **130** (2023) 191003 [[arXiv:2209.13942](#)] [[INSPIRE](#)].
- [13] E. Boera, G.D. Becker, J.S. Bolton and F. Nasir, *Revealing Reionization with the Thermal History of the Intergalactic Medium: New Constraints from the Ly $\alpha$  Flux Power Spectrum*, *Astrophys. J.* **872** (2019) 101 [[arXiv:1809.06980](#)] [[INSPIRE](#)].
- [14] M. Walther, J. Oñorbe, J.F. Hennawi and Z. Lukić, *New Constraints on IGM Thermal Evolution from the Ly $\alpha$  Forest Power Spectrum*, *Astrophys. J.* **872** (2019) 13 [[arXiv:1808.04367](#)] [[INSPIRE](#)].
- [15] B. Villaseñor, B. Robertson, P. Madau and E. Schneider, *Inferring the Thermal History of the Intergalactic Medium from the Properties of the Hydrogen and Helium Ly $\alpha$  Forest*, *Astrophys. J.* **933** (2022) 59 [[arXiv:2111.00019](#)] [[INSPIRE](#)].
- [16] R.A.C. Croft, W. Hu and R. Davé, *Cosmological Limits on the Neutrino Mass from the Ly $\alpha$  Forest*, *Phys. Rev. Lett.* **83** (1999) 1092 [[astro-ph/9903335](#)] [[INSPIRE](#)].
- [17] N. Palanque-Delabrouille et al., *Constraint on neutrino masses from SDSS-III/BOSS Ly $\alpha$  forest and other cosmological probes*, *JCAP* **02** (2015) 045 [[arXiv:1410.7244](#)] [[INSPIRE](#)].
- [18] C. Yèche, N. Palanque-Delabrouille, J. Baur and H. du Mas des Bourboux, *Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100*, *JCAP* **06** (2017) 047 [[arXiv:1702.03314](#)] [[INSPIRE](#)].
- [19] V.K. Narayanan, D.N. Spergel, R. Davé and C.-P. Ma, *Constraints on the mass of warm dark matter particles and the shape of the linear power spectrum from the Ly $\alpha$  forest*, *Astrophys. J. Lett.* **543** (2000) L103 [[astro-ph/0005095](#)] [[INSPIRE](#)].
- [20] U. Seljak, A. Makarov, P. McDonald and H. Trac, *Can sterile neutrinos be the dark matter?*, *Phys. Rev. Lett.* **97** (2006) 191303 [[astro-ph/0602430](#)] [[INSPIRE](#)].
- [21] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy and M. Viel, *Lyman- $\alpha$  constraints on warm and on warm-plus-cold dark matter models*, *JCAP* **05** (2009) 012 [[arXiv:0812.0010](#)] [[INSPIRE](#)].
- [22] M.-Y. Wang, R.A.C. Croft, A.H.G. Peter, A.R. Zentner and C.W. Purcell, *Lyman- $\alpha$  forest constraints on decaying dark matter*, *Phys. Rev. D* **88** (2013) 123515 [[arXiv:1309.7354](#)] [[INSPIRE](#)].
- [23] M. Viel, G.D. Becker, J.S. Bolton and M.G. Haehnelt, *Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman- $\alpha$  forest data*, *Phys. Rev. D* **88** (2013) 043502 [[arXiv:1306.2314](#)] [[INSPIRE](#)].
- [24] V. Iršič, M. Viel, M.G. Haehnelt, J.S. Bolton and G.D. Becker, *First constraints on fuzzy dark matter from Lyman- $\alpha$  forest data and hydrodynamical simulations*, *Phys. Rev. Lett.* **119** (2017) 031302 [[arXiv:1703.04683](#)] [[INSPIRE](#)].
- [25] V. Iršič et al., *New Constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman- $\alpha$  forest data*, *Phys. Rev. D* **96** (2017) 023522 [[arXiv:1702.01764](#)] [[INSPIRE](#)].
- [26] B. Villaseñor, B. Robertson, P. Madau and E. Schneider, *New constraints on warm dark matter from the Lyman- $\alpha$  forest power spectrum*, *Phys. Rev. D* **108** (2023) 023502 [[arXiv:2209.14220](#)] [[INSPIRE](#)].
- [27] M. Levi et al., *The DESI Experiment, a whitepaper for Snowmass 2013*, [arXiv:1308.0847](#) [[INSPIRE](#)].

- [28] DESI collaboration, *The DESI Experiment Part I: Science, Targeting, and Survey Design*, [arXiv:1611.00036](#) [INSPIRE].
- [29] DESI collaboration, *The DESI Experiment Part II: Instrument Design*, [arXiv:1611.00037](#) [INSPIRE].
- [30] K.S. Dawson et al., *The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *Astron. J.* **151** (2016) 44 [[arXiv:1508.04473](#)] [INSPIRE].
- [31] A. Font-Ribera, P. McDonald and A. Slosar, *How to estimate the 3D power spectrum of the Lyman- $\alpha$  forest*, *JCAP* **01** (2018) 003 [[arXiv:1710.11036](#)] [INSPIRE].
- [32] M.L. Abdul Karim, E. Armengaud, G. Mention, S. Chabanier, C. Ravoux and Z. Lukić, *Measurement of the small-scale 3D Lyman- $\alpha$  forest power spectrum*, *JCAP* **05** (2024) 088 [[arXiv:2310.09116](#)] [INSPIRE].
- [33] R. de Belsunce, O.H.E. Philcox, V. Iršič, P. McDonald, J. Guy and N. Palanque-Delabrouille, *The 3D Lyman- $\alpha$  forest power spectrum from eBOSS DR16*, *Mon. Not. Roy. Astron. Soc.* **533** (2024) 3756 [[arXiv:2403.08241](#)] [INSPIRE].
- [34] B. Horowitz, R. de Belsunce and Z. Lukić, *Maximum a posteriori Ly $\alpha$  estimator (MAPLE): band power and covariance estimation of the 3D Ly $\alpha$  forest power spectrum*, *Mon. Not. Roy. Astron. Soc.* **536** (2025) 845 [[arXiv:2403.17294](#)] [INSPIRE].
- [35] A. Pontzen, *Scale-dependent bias in the baryonic-acoustic-oscillation-scale intergalactic neutral hydrogen*, *Phys. Rev. D* **89** (2014) 083010 [[arXiv:1402.0506](#)] [INSPIRE].
- [36] S. Gontcho A Gontcho, J. Miralda-Escudé and N.G. Busca, *On the effect of the ionizing background on the Ly $\alpha$  forest autocorrelation function*, *Mon. Not. Roy. Astron. Soc.* **442** (2014) 187 [[arXiv:1404.7425](#)] [INSPIRE].
- [37] A. Meiksin and M. McQuinn, *Time-dependent fluctuations in the metagalactic photoionization background*, *Mon. Not. Roy. Astron. Soc.* **482** (2019) 4777 [[arXiv:1809.08645](#)] [INSPIRE].
- [38] H. Long and C.M. Hirata, *Probing large-scale ionizing background fluctuation with Lyman  $\alpha$  forest and galaxy cross-correlation at  $z = 2.4$* , *Mon. Not. Roy. Astron. Soc.* **520** (2023) 948 [[arXiv:2209.07019](#)] [INSPIRE].
- [39] P. Montero-Camacho, C.M. Hirata, P. Martini and K. Honscheid, *Impact of inhomogeneous reionization on the Lyman- $\alpha$  forest*, *Mon. Not. Roy. Astron. Soc.* **487** (2019) 1047 [[arXiv:1902.02892](#)] [INSPIRE].
- [40] P. Montero-Camacho and Y. Mao, *Ly $\alpha$  forest power spectrum as an emerging window into the epoch of reionization and cosmic dawn*, *Mon. Not. Roy. Astron. Soc.* **499** (2020) 1640 [[arXiv:2003.10077](#)] [INSPIRE].
- [41] M. McQuinn, L. Hernquist, A. Lidz and M. Zaldarriaga, *The signatures of large-scale temperature and intensity fluctuations in the Lyman  $\alpha$  forest*, *Mon. Not. Roy. Astron. Soc.* **415** (2011) 977 [[arXiv:1010.5250](#)] [INSPIRE].
- [42] P. Montero-Camacho, Y. Zhang and Y. Mao, *The long-lasting effect of X-ray pre-heating in the post-reionization intergalactic medium*, *Mon. Not. Roy. Astron. Soc.* **529** (2024) 3666 [[arXiv:2307.10598](#)] [INSPIRE].
- [43] N. Dalal, O. Doré, D. Huterer and A. Shirokov, *The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects*, *Phys. Rev. D* **77** (2008) 123514 [[arXiv:0710.4560](#)] [INSPIRE].

- [44] U. Seljak, *Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Ly- $\alpha$  forest*, *JCAP* **03** (2012) 004 [[arXiv:1201.0594](#)] [[INSPIRE](#)].
- [45] H.A. Feldman, N. Kaiser and J.A. Peacock, *Power spectrum analysis of three-dimensional redshift surveys*, *Astrophys. J.* **426** (1994) 23 [[astro-ph/9304022](#)] [[INSPIRE](#)].
- [46] O.H.E. Philcox and D.J. Eisenstein, *Computing the Small-Scale Galaxy Power Spectrum and Bispectrum in Configuration-Space*, *Mon. Not. Roy. Astron. Soc.* **492** (2020) 1214 [[arXiv:1912.01010](#)] [[INSPIRE](#)].
- [47] A.J.S. Hamilton, *Towards optimal measurement of power spectra I: minimum variance pair weighting and the fisher matrix*, *Mon. Not. Roy. Astron. Soc.* **289** (1997) 285 [[astro-ph/9701008](#)] [[INSPIRE](#)].
- [48] S.P. Oh, D.N. Spergel and G. Hinshaw, *An Efficient technique to determine the power spectrum from cosmic microwave background sky maps*, *Astrophys. J.* **510** (1999) 551 [[astro-ph/9805339](#)] [[INSPIRE](#)].
- [49] C.M. Hirata, N. Padmanabhan, U. Seljak, D. Schlegel and J. Brinkmann, *Cross-correlation of CMB with large-scale structure: Weak gravitational lensing*, *Phys. Rev. D* **70** (2004) 103501 [[astro-ph/0406004](#)] [[INSPIRE](#)].
- [50] P. McDonald et al., *The Ly $\alpha$  Forest Power Spectrum from the Sloan Digital Sky Survey*, *Astrophys. J. Suppl.* **163** (2006) 80 [[astro-ph/0405013](#)] [[INSPIRE](#)].
- [51] N.G. Karaçaylı, A. Font-Ribera and N. Padmanabhan, *Optimal 1D Ly  $\alpha$  forest power spectrum estimation — I. DESI-lite spectra*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) 4742 [[arXiv:2008.06421](#)] [[INSPIRE](#)].
- [52] N. Padmanabhan, U. Seljak and U.L. Pen, *Mining weak lensing surveys*, *New Astron.* **8** (2003) 581 [[astro-ph/0210478](#)] [[INSPIRE](#)].
- [53] A.B. Newman et al., *LATIS: The Ly $\alpha$  Tomography IMACS Survey*, *Astrophys. J.* **891** (2020) 147 [[arXiv:2002.10676](#)] [[INSPIRE](#)].
- [54] N.Y. Gnedin and L. Hui, *Probing the Universe with the Ly $\alpha$  forest — I. Hydrodynamics of the low-density intergalactic medium*, *Mon. Not. Roy. Astron. Soc.* **296** (1998) 44 [[astro-ph/9706219](#)] [[INSPIRE](#)].
- [55] A. Arinyo-i-Prats, J. Miralda-Escudé, M. Viel and R. Cen, *The non-linear power spectrum of the Lyman alpha forest*, *JCAP* **12** (2015) 017 [[arXiv:1506.04519](#)] [[INSPIRE](#)].
- [56] P. McDonald, *Toward a Measurement of the Cosmological Geometry at  $z \sim 2$ : Predicting Ly $\alpha$  Forest Correlation in Three Dimensions and the Potential of Future Data Sets*, *Astrophys. J.* **585** (2003) 34 [[astro-ph/0108064](#)] [[INSPIRE](#)].
- [57] S. Chabanier et al., *The ACCEL2 project: simulating Lyman- $\alpha$  forest in large-volume hydrodynamical simulations*, *Mon. Not. Roy. Astron. Soc.* **534** (2024) 2674 [[arXiv:2407.04473](#)] [[INSPIRE](#)].
- [58] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [[astro-ph/9911177](#)] [[INSPIRE](#)].
- [59] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](#)] [[INSPIRE](#)].
- [60] A. Font-Ribera and J. Miralda-Escudé, *The effect of high column density systems on the measurement of the Lyman- $\alpha$  forest correlation function*, *JCAP* **07** (2012) 028 [[arXiv:1205.2018](#)] [[INSPIRE](#)].

- [61] K.K. Rogers, S. Bird, H.V. Peiris, A. Pontzen, A. Font-Ribera and B. Leistedt, *Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers*, *Mon. Not. Roy. Astron. Soc.* **476** (2018) 3716 [[arXiv:1711.06275](#)] [[INSPIRE](#)].
- [62] J.J. Givans et al., *Non-linearities in the Lyman- $\alpha$  forest and in its cross-correlation with dark matter halos*, *JCAP* **09** (2022) 070 [[arXiv:2205.00962](#)] [[INSPIRE](#)].
- [63] DESI collaboration, *DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest*, *JCAP* **01** (2025) 124 [[arXiv:2404.03001](#)] [[INSPIRE](#)].
- [64] M. Tegmark, A.N. Taylor and A.F. Heavens, *Karhunen-Loève Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?*, *Astrophys. J.* **480** (1997) 22 [[astro-ph/9603021](#)] [[INSPIRE](#)].
- [65] M. Tegmark, A.J.S. Hamilton, M.A. Strauss, M.S. Vogeley and A.S. Szalay, *Measuring the galaxy power spectrum with future redshift surveys*, *Astrophys. J.* **499** (1998) 555 [[astro-ph/9708020](#)] [[INSPIRE](#)].
- [66] U. Seljak, *Weak lensing reconstruction and power spectrum estimation: minimum variance methods*, *Astrophys. J.* **506** (1998) 64 [[astro-ph/9711124](#)] [[INSPIRE](#)].
- [67] J.D. Talman, *Numerical Fourier and Bessel transforms in logarithmic variables*, *J. Comput. Phys.* **29** (1978) 35.
- [68] A.J.S. Hamilton, *Uncorrelated modes of the nonlinear power spectrum*, *Mon. Not. Roy. Astron. Soc.* **312** (2000) 257 [[astro-ph/9905191](#)] [[INSPIRE](#)].
- [69] A.S. Bolton and D.J. Schlegel, *Spectro-Perfectionism: An Algorithmic Framework for Photon Noise-Limited Extraction of Optical Fiber Spectroscopy*, *Publ. Astron. Soc. Pac.* **122** (2010) 248 [[arXiv:0911.2689](#)] [[INSPIRE](#)].
- [70] J. Guy et al., *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, *Astron. J.* **165** (2023) 144 [[arXiv:2209.14482](#)] [[INSPIRE](#)].
- [71] D. Kirkby et al., *Fitting methods for baryon acoustic oscillations in the Lyman- $\alpha$  forest fluctuations in BOSS data release 9*, *JCAP* **03** (2013) 024 [[arXiv:1301.3456](#)] [[INSPIRE](#)].
- [72] G. Schulz, *Iterative Berechnung der reziproken Matrix*, *J. Appl. Math. Mech.* **13** (1933) 57.
- [73] C. Kenney and A.J. Laub, *Rational Iterative Methods for the Matrix Sign Function*, *SIAM J. Matrix Anal. Appl.* **12** (1991) 273.
- [74] P. Pandey, C. Kenney and A.J. Laub, *A parallel algorithm for the matrix sign function*, *Int. J. High Speed Comput.* **02** (1990) 181.
- [75] C. Kenney and A. Laub, *A hyperbolic tangent identity and the geometry of Padé sign function iterations*, *Numer. Algorithm.* **7** (1994) 111.
- [76] N. Higham, *Stable iterations for the matrix square root*, *Numer. Algorithm.* **15** (1997) 227.
- [77] M.S. Petković and S. Lakić, *On the Matrix Square Root*, *J. Appl. Math. Mech.* **78** (1998) 173.
- [78] M. Vono, N. Dobigeon and P. Chainais, *High-Dimensional Gaussian Sampling: A Review and a Unifying Approach Based on a Stochastic Proximal Point Algorithm*, *SIAM Rev.* **64** (2022) 3 [[arXiv:2010.01510](#)].
- [79] A. Parker and C. Fox, *Sampling Gaussian Distributions in Krylov Spaces with Conjugate Gradients*, *SIAM J. Sci. Comput.* **34** (2012) B312.
- [80] Ohio Supercomputer Center, *Ohio supercomputer center*, (1987), <http://osc.edu/ark:/19495/f5s1ph73>.

- [81] S. Bernstein, *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*, *Commun. Kharkov Math. Soc.* **13** (1912) 1.
- [82] P. de Casteljau, *Outillages méthodes calcul*, André Citroën Automobiles SA, Paris, France (1959).
- [83] G. Farin, *Curves and Surfaces for CAGD: A Practical Guide*, in *The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling*, fifth edition, Morgan Kaufmann (2002) [[DOI:10.1016/b978-1-55860-737-8.x5000-5](https://doi.org/10.1016/b978-1-55860-737-8.x5000-5)].
- [84] C. Ramírez-Pérez et al., *The Lyman- $\alpha$  forest catalog from the Dark Energy Spectroscopic Instrument Early Data Release*, *Mon. Not. Roy. Astron. Soc.* **528** (2024) 6666 [[arXiv:2306.06312](https://arxiv.org/abs/2306.06312)] [[INSPIRE](#)].
- [85] Z. Sun, Y.-S. Ting and Z. Cai, *Quasar Factor Analysis — An Unsupervised and Probabilistic Quasar Continuum Prediction Algorithm with Latent Factor Analysis*, *Astrophys. J. Suppl.* **269** (2023) 4 [[arXiv:2211.11784](https://arxiv.org/abs/2211.11784)] [[INSPIRE](#)].
- [86] W. Turner et al., *New measurements of the Lyman- $\alpha$  forest continuum and effective optical depth with LyCAN and DESI Y1 data*, *Astrophys. J.* **976** (2024) 143 [[arXiv:2405.06743](https://arxiv.org/abs/2405.06743)] [[INSPIRE](#)].
- [87] M. Blomqvist et al., *Broadband distortion modeling in Lyman- $\alpha$  forest BAO fitting*, *JCAP* **11** (2015) 034 [[arXiv:1504.06656](https://arxiv.org/abs/1504.06656)] [[INSPIRE](#)].
- [88] G.M. Bernstein and D. Gruen, *Resampling images in Fourier domain*, *Publ. Astron. Soc. Pac.* **126** (2014) 287 [[arXiv:1401.2636](https://arxiv.org/abs/1401.2636)] [[INSPIRE](#)].
- [89] N.G. Karaçaylı et al., *Optimal 1D Ly $\alpha$  Forest Power Spectrum Estimation — III. DESI early data*, *Mon. Not. Roy. Astron. Soc.* **528** (2024) 3941 [[arXiv:2306.06316](https://arxiv.org/abs/2306.06316)] [[INSPIRE](#)].
- [90] S. Chabanier et al., *The one-dimensional power spectrum from the SDSS DR14 Ly $\alpha$  forests*, *JCAP* **07** (2019) 017 [[arXiv:1812.03554](https://arxiv.org/abs/1812.03554)] [[INSPIRE](#)].
- [91] C. Ravoux et al., *The Dark Energy Spectroscopic Instrument: one-dimensional power spectrum from first Ly  $\alpha$  forest samples with Fast Fourier Transform*, *Mon. Not. Roy. Astron. Soc.* **526** (2023) 5118 [[arXiv:2306.06311](https://arxiv.org/abs/2306.06311)] [[INSPIRE](#)].
- [92] DESI collaboration, *DESI DR2 results. I. Baryon acoustic oscillations from the Lyman alpha forest*, accepted for publication in *Phys. Rev. D* (2025), [arXiv:2503.14739](https://arxiv.org/abs/2503.14739) [[DOI:10.1103/2wmw-xjm5](https://doi.org/10.1103/2wmw-xjm5)] [[INSPIRE](#)].
- [93] F. Beutler et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Anisotropic galaxy clustering in Fourier-space*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 2242 [[arXiv:1607.03150](https://arxiv.org/abs/1607.03150)] [[INSPIRE](#)].
- [94] K.B. Fisher, C.A. Scharf and O. Lahav, *A spherical harmonic approach to redshift distortion and a measurement of  $\Omega_0$  from the 1.2 Jy IRAS Redshift Survey*, *Mon. Not. Roy. Astron. Soc.* **266** (1994) 219 [[astro-ph/9309027](https://arxiv.org/abs/astro-ph/9309027)] [[INSPIRE](#)].
- [95] N. Padmanabhan et al., *The Clustering of Luminous Red Galaxies in the Sloan Digital Sky Survey Imaging Data*, *Mon. Not. Roy. Astron. Soc.* **378** (2007) 852 [[astro-ph/0605302](https://arxiv.org/abs/astro-ph/0605302)] [[INSPIRE](#)].
- [96] J. Guy et al., *Characterization of contaminants in the Lyman-alpha forest auto-correlation with DESI*, *JCAP* **01** (2025) 140 [[arXiv:2404.03003](https://arxiv.org/abs/2404.03003)] [[INSPIRE](#)].
- [97] R. Besuner et al., *The Spectroscopic Stage-5 Experiment*, [arXiv:2503.07923](https://arxiv.org/abs/2503.07923) [[INSPIRE](#)].

- [98] M. Challacombe, T. Haut and N. Bock, *A N-Body Solver for Square Root Iteration*, [arXiv:1508.05856](#).
- [99] Y. Song, N. Sebe and W. Wang, *Fast Differentiable Matrix Square Root*, [arXiv:2201.08663](#).
- [100] PLANCK collaboration, *Planck intermediate results. XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth*, *Astron. Astrophys.* **596** (2016) A107 [[arXiv:1605.02985](#)] [[INSPIRE](#)].
- [101] R.B. Metcalf, N. Tessore and R.A.C. Croft, *Reconstructing the gravitational lensing potential from the Lyman- $\alpha$  forest*, *Astron. Astrophys.* **642** (2020) A122 [[arXiv:2005.04109](#)] [[INSPIRE](#)].
- [102] Wikipedia contributors, *Cubic hermite spline*, Wikipedia, The Free Encyclopedia (2025), [https://en.wikipedia.org/w/index.php?title=Cubic\\_Hermite\\_spline&oldid=1281273976](https://en.wikipedia.org/w/index.php?title=Cubic_Hermite_spline&oldid=1281273976).
- [103] E. Catmull and R. Rom, *A class of local interpolating splines*, in *Computer Aided Geometric Design*, R.E. Barnhill and R.F. Riesenfeld eds., Academic Press (1974), pp. 317–326 [[DOI:10.1016/b978-0-12-079050-0.50020-5](#)].
- [104] M. Galassi et al., *GNU scientific library: Reference manual (release 2.7)*, (2021), <https://www.gnu.org/software/gsl/>.
- [105] M. Frigo and S.G. Johnson, *The Design and Implementation of FFTW3*, *Proc. IEEE* **93** (2005) 216 [[INSPIRE](#)].
- [106] ASTROPY collaboration, *Astropy: A Community Python Package for Astronomy*, *Astron. Astrophys.* **558** (2013) A33 [[arXiv:1307.6212](#)] [[INSPIRE](#)].
- [107] ASTROPY collaboration, *The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package*, *Astron. J.* **156** (2018) 123 [[arXiv:1801.02634](#)] [[INSPIRE](#)].
- [108] ASTROPY collaboration, *The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package*, *Astrophys. J.* **935** (2022) 167 [[arXiv:2206.14220](#)] [[INSPIRE](#)].
- [109] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [[arXiv:2006.10256](#)] [[INSPIRE](#)].
- [110] J.D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comput. Sci. Eng.* **9** (2007) 90 [[INSPIRE](#)].