Lawrence Berkeley National Laboratory

Title
How to estimate the 3D power spectrum of the Lyman-α forest

Authors
Font-Ribera, Andreu
McDonald, Patrick
Slosar, Anže

How to estimate the 3D power spectrum of the Lyman-α forest

Andreu Font-Ribera [a],[1],[†] Patrick McDonald,[2],[‡] and Anže Slosar[3],[§]

[1] *Department of Physics and Astronomy, University College London, Gower Street, London, United Kingdom*
[2] *Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA*
[3] *Brookhaven National Laboratory, Upton, NY 11973, USA*
(Dated: January 1, 2018)

We derive and numerically implement an algorithm for estimating the 3D power spectrum of the Lyman-α (Lyα) forest flux fluctuations. The algorithm exploits the unique geometry of Lyα forest data to efficiently measure the cross-spectrum between lines of sight as a function of parallel wavenumber, transverse separation and redshift. We start by approximating the global covariance matrix as block-diagonal, where only pixels from the same spectrum are correlated. We then compute the eigenvectors of the derivative of the signal covariance with respect to cross-spectrum parameters, and project the inverse-covariance-weighted spectra onto them. This acts much like a radial Fourier transform over redshift windows. The resulting cross-spectrum inference is then converted into our final product, an approximation of the likelihood for the 3D power spectrum expressed as second order Taylor expansion around a fiducial model. We demonstrate the accuracy and scalability of the algorithm and comment on possible extensions. Our algorithm will allow efficient analysis of the upcoming Dark Energy Spectroscopic Instrument dataset.

I. INTRODUCTION

Inference of the power spectra of cosmic fields is one of the main tools of observational cosmology. Theory behind estimation of power spectra for Cosmic Microwave Background and galaxy surveys (and other point tracers) has been well developed over the past thirty years. It is not difficult to write a likelihood for the power spectrum given measured fields assuming these to be Gaussian. From this one can derive the standard "optimal quadratic estimator"[1] (OQE) [1–3], but it is numerically too expensive to evaluate for large surveys using brute-force linear algebra. Therefore, a number of good approximations have been developed [4–6].

In the field of Lyα forest, the situation is much less developed. [7] used an OQE to measure the Lyα forest one-dimensional (1D) power spectrum, describing the correlations between pixels in a given line of sight, and a similar algorithm was more recently used in [8]. However, as the density of lines of sight increases in present and future Lyα forest surveys, it becomes more important to measure the full three-dimensional (3D) correlations, including correlations between pixels from different lines of sight. Recent analyses of 3D correlations, focused on measuring the large scale feature of Baryon Acoustic Oscillation (BAO), have relied on measuring the correlation function using simple pixel-product algorithms [9–12] (although see [13] for an example of using OQE-like approach to estimate the correlation function). The main goal of this paper is to present a likelihood based algorithm that will allow us to measure the 3D power spectrum of the Lyα forest down to small scales, and including all the information in traditional 1D analyses.

In principle the power spectrum and correlation function are simply linear transforms of one another, potentially containing the same information. In spite of this, in practice they are often thought of as living in disconnected worlds, i.e., estimated using much different algorithms and not compared directly. Each has potential advantages for isolating different kinds of systematics into a limited range of their coordinates (e.g., the power spectrum is good for isolating slowly varying effects to small wavenumbers, while the correlation function is good for isolating effects at specific separations). Different tricks and approximations are generally used to make estimation algorithms tractable for large data sets (including, most importantly, estimation of error bars). The power spectrum has the advantage of generally less correlated errors and diagonalization of the theory in the linear regime limit. The primary point of this paper is to provide a practical algorithm for the 3D Lyα forest power spectrum measurement, in order to exploit its advantages, but an implicit point is that the different statistics do not need to live in isolated worlds. We show in passing how a measurement of the power spectrum can always be converted into a correlation function measurement

---

[a] author list alphabetized
[†] a.font@ucl.ac.uk
[‡] PVMcDonald@lbl.gov
[§] anze@bnl.gov

[1] Strictly speaking, the OQE is only approximately quadratic and optimal, even for Gaussian fields, although in the large data set limit the distinction is not important. The OQE equation as usually written appears at first glance to be quadratic in the data, but it only achieves a maximum likelihood solution after iteration, which makes the final answer non-quadratic in the input data (it is easy to see this by noting that the Fisher matrix, i.e., 2nd derivative of the likelihood, appears in the standard formula, while the maximum likelihood solution can only require the 1st derivative).

and vice versa, so that ultimately one should be able to exploit the advantages of both within the same basic analysis (our conversion of the cross-spectrum to power spectrum is an explicit example of this).

Methods developed for point tracers do not adapt directly to the Lyα forest problem, because of the unique window function of the Lyα forest. We measure the flux fields in individual lines of sight, each of which covers a certain redshift range, which do not overlap perfectly and which generally offer relatively high resolution in the radial direction but only sparse sampling in the transverse direction.

This work is clearly timely given the forthcoming large increases in the density of lines of sight in upcoming Lyα forest surveys such as Dark Energy Spectroscopic Instrument (DESI, [14]). The analysis of these datasets will present numerous challenges beyond power spectrum estimation in terms a plethora of observational and instrumental systematics. In the current work we sweep these under the rug and assume ideal measurements. However, ultimately one of the purposes of our general framework is to allow understanding and propagation of these effects.

We expect that applying Fast Fourier Transform (FFT) algorithm to each spectrum will lead to algorithmic improvements, since each line of sight's correlations are approximately stationary. This also performs "scale sorting", which isolates large-scale radial modes that correlate across large transverse separations from small scale modes which have significant correlations across considerably smaller transverse separations. This line of thought naturally brings cross-spectrum, $P_\times(z, \Delta\theta, k_\parallel)$, between lines of sight as a function of angular separation as a natural intermediate product,[2] from which the 3D power spectrum can be later inferred. In this paper we bring this rough idea to completion, by presenting a detailed algorithm and its numerical implementation. We begin with the likelihood function and then develop the numerically efficient and mathematically rigorous derivation of expressions for cross-spectrum with well understood approximations and then demonstrate how this can be transformed into 3D power spectrum. The final product will be a likelihood for 3D power spectrum as functions of redshift and of radial and transverse wavenumbers.

We start in section II by presenting the basic likelihood analysis formalism we follow in this paper, and the specific case of a clustering analysis of the Lyα forest is discussed in section III. The numerical implementation is detailed in section IV. In section V we revisit 1D analyses in this context, and we show examples of analyses on mock datasets. Measurements of the cross-power spectrum on mock datasets are presented in VI, and in section VII we describe how these can be translated into measurements of the 3D power spectrum. We conclude with a discussion in section VIII.

II.   LIKELIHOOD APPROXIMATION VS. STANDARD POWER SPECTRUM "ESTIMATORS"

While our algorithm should obtain results in the end essentially equivalent to the standard OQE, i.e., it could be understood simply as a computing optimization, we think it is useful to take a step back and shed some of the historical baggage that comes along with the words in OQE. In cosmological applications, the OQE has been introduced first into analysis of the CMB and galaxy survey data [1, 3, 5, 16]. There are two canonical derivations that appear in these papers. In the first one, a general quadratic estimator is written, the weights are then chosen to ensure inverse variance weighting and unbiased estimates and then Cramer-Rao inequality is used to demonstrate the estimator to be optimal (however, strictly speaking this demonstration requires knowing the correct answer in advance to use for the weighting). In the second derivation, the likelihood is written and then the estimator is presented as a Newton-Raphson step towards maximum likelihood with second derivative replaced by Fisher matrix for numerical efficiency. These are both valid approaches, but the OQE process is commonly viewed as simply an estimator that transforms data into summary statistics. In this paper, we want to connect the same fundamental mathematics to a more Bayesian thinking about the likelihood. In particular, we believe that understanding the measurement process explicitly as a likelihood expansion leads to a streamlined understanding of how we can convert between parameter bases, e.g., between cross-spectrum, correlation function, and power spectrum parameters.

Standard Bayesian data analysis says that for any data set $\mathbf{d}$ and parameters $\mathbf{p}$ we should compute the posterior probability for the parameters given the data, $L(\mathbf{p}|\mathbf{d})$. Generally we do this using Bayes theorem, $L(\mathbf{p}|\mathbf{d}) \propto L(\mathbf{d}|\mathbf{p})L(\mathbf{p})$ where $L(\mathbf{p})$ is a prior on the parameters. In multi-dimensional parameter spaces it is frequently not possible to fully tabulate $L(\mathbf{p}|\mathbf{d})$, let alone obtain an analytic form, so we generally need to look for some limited approximate representation of it, sufficient for the use we intend. One such representation is a set of samples of the parameter vector – instances of $\mathbf{p}$ distributed consistent with $L(\mathbf{p}|\mathbf{d})$, e.g., from MCMC. These are convenient if one wants to integrate over $L(\mathbf{p}|\mathbf{d})$, e.g., to produce means or variances or confidence intervals for $\mathbf{p}$.

An alternative is to approximate $\mathcal{L} \equiv \ln L$ by a Taylor series around $\mathbf{p}_0$, i.e., $\mathcal{L}(\mathbf{d}|\mathbf{p}) = \mathcal{L}(\mathbf{d}|\mathbf{p}_0) + \mathcal{L}_{,\alpha}(\mathbf{d}|\mathbf{p}_0)\delta p_\alpha + \frac{1}{2}\mathcal{L}_{,\alpha\beta}(\mathbf{d}|\mathbf{p}_0)\delta p_\alpha \delta p_\beta + ...$, where $\delta\mathbf{p} \equiv \mathbf{p} - \mathbf{p}_0$. If we truncate this series at quadratic order this is equivalent to the

---

[2] In this paper we follow the jargon of [15] in calling the radial Fourier transform of the 3D Lyα forest flux correlation function at non-zero transverse separation the "cross-spectrum." This is not to be confused with the three-dimensional cross-power spectrum between multiple tracers, e.g. the 3D Fourier transform of the cross-correlation function between quasar positions and Lyα forest.

ubiquitous assumption that the errors are Gaussian. This assumption is exact if likelihood has a Gaussian shape and the theory is linear in parameters, but it is in general a good approximation due to central limit theorem or if theory is effectively linear over the range of high-likelihood region. In the power spectrum measurement problem it is wrong in sample-variance dominated measurement when only few modes are present (e.g. the low-ℓ CMB power spectrum). A key point here, especially relevant to band power parameters, is if we have this Taylor series expansion around parameters $\mathbf{p}_0$ close enough to the values where we want to use the likelihood, *we are done*, i.e., there is no fundamental need to estimate "best" values for the parameters, as this is irrelevant to further use of the likelihood function. In fact, for band power parameters where we expect the true power to be smooth, it generally makes more sense to expand around our best *global* estimate of parameter values, e.g., their values in our favorite cosmological model constrained by all available datasets. Moreover, for the same reason it is better not to perform "iterative" expansion around maximum likelihood, because the covariance matrix given by an expansion around the best guess of true cosmological parameter is going to be more accurate than that from an expansion around a model that is the best fit to only one of the datasets (for a concrete example, see [17]). This does not preclude identifying deviations from the best-guess model, as the likelihood expansion still implies some maximum likelihood parameter values – they just are not necessarily $\mathbf{p}_0$. At most, iteration should take place at the most global possible level, e.g., possibly adjusting $\mathbf{p}_0$ if the results of all-data cosmological parameter estimation give a change in the best model.[3]

Given $\mathcal{L}_{,\alpha}(\mathbf{p}_0)$ and $\mathcal{L}_{,\alpha\beta}(\mathbf{p}_0)$, finding the constraints on a different set of parameters, centered on the same fiducial model and assuming a Taylor series for $\mathcal{L}$ is still sufficient, is generally as simple as applying the derivative chain rule, e.g., $\frac{\partial \mathcal{L}}{\partial \mathbf{p}} = \frac{\partial \mathbf{q}}{\partial \mathbf{p}}\frac{\partial \mathcal{L}}{\partial \mathbf{q}}$, where $\mathbf{q}$ is the new set of parameters.[4]

To make this more concrete for our problem: while the goal of a power spectrum measurement is often presented as being to produce a set of estimated values for band parameters and their covariance matrix (e.g., the $P_{1D}(z, k_\parallel)$ estimates of [7] and [8]), we consider our task to be to determine the coefficients $\mathcal{L}_{,\alpha}$ and $\mathcal{L}_{,\alpha\beta}$.

We describe our data as $\mathbf{d} = \mathbf{m} + \mathbf{R}\delta + \epsilon$, where $\mathbf{m} = \langle \mathbf{d} \rangle$ is the mean, $\mathbf{R}$ is the response matrix, δ is the cosmological fluctuation of interest and ϵ is the instrumental noise. We assume a Gaussian likelihood for the data given a model,

$$L(\mathbf{d}|\mathbf{p}) \propto \det(\mathbf{C})^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{d} - \mathbf{m})^t \mathbf{C}^{-1}(\mathbf{d} - \mathbf{m})\right] \ , \tag{1}$$

where $\mathbf{C} = \mathbf{R}\mathbf{S}\mathbf{R}^t + \mathbf{N}$ with $\mathbf{S}$ the covariance of δ (the signal we want to measure) and $\mathbf{N}$ the (diagonal) noise matrix. In the next section we will discuss in more detail what $\mathbf{m}$ and $\mathbf{R}$ are in the context of Lyα forest analyses, but what matters here is that they do not depend on the parameters $\mathbf{p}$ we are trying to estimate.

As usual in power spectrum estimation, we assume that $\mathbf{S}$ is linear in some parameters. The derivatives we need are now quite simple:

$$\mathcal{L}_{,\alpha} = \frac{1}{2}(\mathbf{d} - \mathbf{m})^t \mathbf{C}_0^{-1}\mathbf{C}_{,\alpha}\mathbf{C}_0^{-1}(\mathbf{d} - \mathbf{m}) - \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}_0^{-1}\mathbf{C}_{,\alpha}\right] \ , \tag{2}$$

and

$$\mathcal{L}_{,\alpha\beta} = -(\mathbf{d} - \mathbf{m})^t \mathbf{C}_0^{-1}\mathbf{C}_{,\alpha}\mathbf{C}_0^{-1}\mathbf{C}_{,\beta}\mathbf{C}_0^{-1}(\mathbf{d} - \mathbf{m}) + \frac{1}{2}\mathrm{Tr}\left[\mathbf{C}_0^{-1}\mathbf{C}_{,\alpha}\mathbf{C}_0^{-1}\mathbf{C}_{,\beta}\right] \ , \tag{3}$$

where $\mathbf{C}_0$ is the covariance matrix evaluated using $\mathbf{p}_0$, and $\mathbf{C}_{,\alpha}$ is the derivative of the covariance matrix with respect to the parameter p_α. These Taylor series coefficients imply some mean (maximum likelihood) values for $\mathbf{p}$, and a covariance matrix around that maximum, but those quantities are superfluous – we are generally going to use our power spectrum results to constrain some more fundamental parameters and for that we just need $L(\mathbf{d}|\mathbf{p})$.

The second derivatives have two terms: the first involves the measured data $\mathbf{d}$, and the second is a trace of a matrix product. The expectation value of the first term is proportional to the second term if $\mathbf{C}_0$ is the correct covariance matrix for the data, so we approximate the second derivative by its expected value under this assumption, as is usually done in deriving the OQE:

$$\mathcal{L}_{,\alpha\beta} \approx \langle \mathcal{L}_{,\alpha\beta} \rangle_0 = -\frac{1}{2}\mathrm{Tr}\left[\mathbf{C}_0^{-1}\mathbf{C}_{,\alpha}\mathbf{C}_0^{-1}\mathbf{C}_{,\beta}\right] = -F_{\alpha\beta} \ , \tag{4}$$

where $F_{\alpha\beta}$ is the Fisher matrix. In general we do not need to make this approximation and it may not actually be helpful, but we do not investigate it further in this paper (see [18] for a discussion about this approximation).

---

[3] In contrast, the most standard "estimate the parameter values and covariance matrix" form of data analysis is to find maximum likelihood values of parameters, where $\mathcal{L}_{,\alpha}(\mathbf{p}_0) = 0$, and expand around this point to find the covariance matrix.

[4] As we discuss below, minor subtleties do arise in the definition of $\partial \mathbf{q}/\partial \mathbf{p}$ for band power measurements.

The traditional OQE is derived as a Newton-Raphson (NR) step toward the maximum likelihood value of parameters, including this approximation for the 2nd derivative matrix, i.e., the standard OQE equation is simply a re-shuffling of terms in:

$$\hat{\mathbf{p}} = \mathbf{p}_0 + \mathbf{F}^{-1}\mathcal{L}' \ , \tag{5}$$

where $\mathbf{F}$ is the Fisher matrix and $\mathcal{L}'$ the first derivatives of the log-likelihood.

Our bottom line point of this section is that in this paper we do not consider this NR step to be a fundamental part of our analysis. Our purpose is to estimate the likelihood derivatives, and we only compute things like the implied central value when necessary for diagnostics like plotting.

In this paper we often ignore the priors on the parameters, since they only play two very minor roles: i) we use them to compute implied central values and errorbars when making plots; ii) in section VII we marginalize over some of the band powers, and we need to take into account the priors on the parameters that we marginalize out (see appendix A on marginalization). Whenever they are required, we use Gaussian priors around zero, with a width set to a thousand times the fiducial power evaluated at the center of the band.[5]

III.   CLUSTERING ANALYSES OF THE LYα FOREST

In this section we will introduce basic notation in analyses of the Lyα forest, present our data model and discuss the parameterization of the power spectrum.

A.   Data vector

In a clustering analysis of the Lyα forest the data set consists of a set of optical spectra of high redshift quasars, identified by an angular position θ and a redshift z_q. In the analysis of a real survey we would need to discuss instrumental specifications and details of the data reduction pipeline: sky subtraction, spectral calibration, co-addition of individual exposures, estimates of the noise variance, etc. In this study we will not address these issues in detail, although we will discuss how they would affect our measurements in section VIII.

We will use a simplified data model to describe the measured flux d_i in a pixel of observed wavelength λ_i in the spectrum of quasar q as:

$$d_i = \int d\lambda \ W_i(\lambda) \ C_q(\lambda) \ F(\lambda) + \epsilon_i \ , \tag{6}$$

where $C_q(\lambda)$ is the quasar continuum, $F(\lambda) = e^{-\tau(\lambda)}$ is the transmitted flux fraction, $\tau(\lambda)$ the optical depth and ϵ_i the instrumental noise in the pixel. $W_i(\lambda)$ is the convolution kernel for pixel i, including both spectral resolution and the pixelization (we will discuss this in detail soon).

In cosmological analyses of the Lyα forest we are interested in measuring the statistics of the fluctuations around the mean transmitted flux fraction (often referred as the *mean flux*), $F(\lambda) = \bar{F}(\lambda)\,(1 + \delta_F(\lambda))$, giving, in our simplified model,

$$d_i = m_i \left(1 + \int d\lambda \ W_i(\lambda) \ \delta(\lambda)\right) + \epsilon_i \ , \tag{7}$$

where we have assumed that $C(\lambda)\bar{F}(\lambda)$ can be approximated as constant over the width of the smoothing kernel, and defined $m_i = C(\lambda_i)\bar{F}(\lambda_i)$.

B.   Covariances

The covariance of the observed data vector (defined in Eq. 7) is given by:

$$C_{ij} = \langle d_i \ d_j \rangle - \langle d_i \rangle \langle d_j \rangle = m_i \ m_j \int d\lambda \ W_i(\lambda) \int d\lambda' \ W_j(\lambda') \ \xi_{3D}(\Delta\theta_{ij}, \lambda, \lambda') + N_{ij} \ , \tag{8}$$

---

[5] We have tested that the exact widths of the priors do not qualitatively change the results of this paper.

where $N_{ij} = \sigma_i^2 \delta_{ij}^K$ is the (diagonal) noise matrix and $\xi_{3D}(\Delta\theta_{ij}, \lambda, \lambda')$ is the 3D correlation function of the δ field we are interested in, and $\Delta\theta_{ij}$ is the angular separation between the spectra of pixels i and j. Defining $\Delta v_{ij} = c\,(\ln\lambda_j - \ln\lambda_i)$, we can exactly equivalently describe the 3D correlations as a function of angular separation, velocity separation and redshift, $\xi_{3D}(z_{ij}, \Delta\theta_{ij}, \Delta v_{ij})$, with $1 + z_{ij} = \sqrt{(1+z_i)(1+z_j)} = \sqrt{\lambda_i \lambda_j}/\lambda_\alpha$. Note that this is equivalent to defining $\bar{v}_{ij} \equiv c\,(\ln\lambda_j + \ln\lambda_i)/2$ and then $1 + z_{ij} \equiv \exp(\bar{v}_{ij}/c)/\lambda_\alpha$. These quantities should be understood as just different ways of packaging the two observable wavelengths – we do not rely on them having any more fundamental physical meaning.

In the presence of redshift evolution, the power spectrum is not well-defined as a variance of Fourier modes. We can, however, still model the data as a Gaussian random field, understood to mean a field with Gaussian likelihood function specified by a correlation function, which does not require the field to be stationary. We *define* the 3D power spectrum to be the Fourier transform of this correlation function:

$$P_{3D}(z, k_\perp, k_\parallel) \equiv \int d\boldsymbol{\Delta\theta}\; e^{i\boldsymbol{\Delta\theta}\mathbf{k}_\perp} \int d\Delta v\; e^{i\Delta v k_\parallel}\; \xi_{3D}(z, \Delta\theta, \Delta v)\;, \tag{9}$$

where $k_\perp$ has units of inverse radian, and $k_\parallel$ units of s km^{-1}. We expect that on scales smaller than the scale of the evolution, the power spectrum defined this way will closely follow intuition based on the ideal stationary case, but in any case it is well-defined without any assumption about that. Note that it is not completely trivial here that the Fourier transform is defined to be with respect to observable coordinates, rather than, e.g., an approximation of comoving coordinates using a fiducial model. One should not think, however, that the BAO feature is directly smeared by this definition – a delta function at some separation in the correlation function in comoving coordinates remains a delta function in observable coordinates.[6]

Given a comoving coordinate power spectrum we can always compute this observable coordinate power spectrum by transforming to a correlation function, introducing the observable coordinates, and transforming back. Note, however, that on the BAO scale the effect of ignoring the non-linearity of the coordinate transform is tiny. E.g., in a typical model the radial comoving separation between $z = 2.31$ and $z = 2.39$ is 68.358 Mpc/h, while $(0.08/3.35)cH(z = 2.35) = 68.400$ Mpc/h, a $\sim 0.06\%$ difference. If points at the same two redshifts are separated by 0.02 radians, the true comoving transverse separation is 78.418 Mpc/h, while 0.02 times the angular diameter distance at $z = 2.35$ is 78.425 Mpc/h, $\sim 0.009\%$ different. I.e., at foreseeable levels of precision we can consider the relation between power spectra to be a simple linear change of units, except possibly on much larger scales. The key is that we keep separation calculations symmetric around the central redshift, so that differences are quadratic in the scale changes (strictly speaking we measure distances symmetrically in $\ln(1 + z)$, but it doesn't make any difference to this discussion). We will discuss coordinates further in Section VIII A.

From the same correlation function we can also define the cross-spectrum,

$$P_\times(z, \Delta\theta, k_\parallel) \equiv \int d\Delta v\; e^{i\Delta v k_\parallel}\; \xi_{3D}(z, \Delta\theta, \Delta v) = \int \frac{d\mathbf{k}_\perp}{(2\pi)^2}\; e^{i\boldsymbol{\Delta\theta}\mathbf{k}_\perp}\; P_{3D}(z, k_\perp, k_\parallel)\;, \tag{10}$$

that quantifies the correlation of lines of sight modes as a function of angular separation.

We can now rewrite the elements of the covariance matrix in equation 8 as a function of the cross-spectrum:

$$C_{ij} = m_i\, m_j \int \frac{dk_\parallel}{2\pi}\; e^{-ik_\parallel \Delta v_{ij}}\; P_\times(z_{ij}, \Delta\theta_{ij}, k_\parallel)\; \tilde{W}_i(k_\parallel)\, \tilde{W}_j(k_\parallel) + N_{ij}\;, \tag{11}$$

where $\tilde{W}_i(k_\parallel)$ is the Fourier transform of the convolution kernel for pixel i, i.e.,

$$\tilde{W}_i(k_\parallel) = e^{-k_\parallel^2 G_i^2/2}\; \frac{\sin\left(k_\parallel T_i/2\right)}{k_\parallel T_i/2}\;, \tag{12}$$

where G_i is the Gaussian spectral resolution and T_i is the pixel width, both in velocity units.[7]

C. Parameterization

We parameterize the power spectrum with a grid of band power parameters:

$$P_{3D}(z, k_\perp, k_\parallel) = P_{3D}^{\mathrm{fid}}(z, k_\perp, k_\parallel) + \sum_\alpha w_\alpha^{3D}(z, k_\perp, k_\parallel)\, p_\alpha^{3D}\;, \tag{13}$$

---

[6] The key is to remember that by definition we must consider pairs of pixels at fixed z_{ij} – there is no averaging together here of, e.g., purely transverse pixel pairs at different redshifts where fixed angular separation means different comoving separation.

[7] Physically, the resolution should be modeled as a continuous convolution of the continuous field, only subsequently discretely sampled onto pixels, but currently spectra are usually presented with a single resolution value per pixel, which is generally slowly varying from pixel to pixel, so we will not make this distinction here

where $P_{3D}^{\text{fid}}(z, k_\perp, k_\parallel)$ is the fiducial power spectrum (around which the likelihood is expanded as explained in Section II) and $w_\alpha^{3D}(z, k_\perp, k_\parallel)$ is a trilinear interpolation kernel. Note that the fiducial power spectrum is evaluated at the full resolution, independent of individual band-power bins. This decreases the impact of features smaller than the bin size, which can still be represented perfectly if they are consistent with the fiducial model.

We also consider an alternative parameterization in which we introduce a constant at each redshift that allows us to define a 3D magnitude of the wavevector,

$$\tilde{k} \equiv \sqrt{k_\parallel^2 + f^{-2}(z) k_\perp^2} \tag{14}$$

and cosine of the angle between the wavevector and the line of sight,

$$\tilde{\mu} \equiv \frac{k_\parallel}{\tilde{k}} . \tag{15}$$

$f(z)$ is necessary of course because observable $k_\parallel$ and $k_\perp$ do not have the same units. We can then parameterize P_{3D} by simple interpolation over $\tilde{k}$ and $\tilde{\mu}$ – we will call this $P_{3D'}$

$$P_{3D'}(z, \tilde{k}, \tilde{\mu}) = P_{3D'}^{\text{fid}}(z, \tilde{k}, \tilde{\mu}) + \sum_\alpha w_\alpha^{3D'}(z, \tilde{k}, \tilde{\mu}) \, p_\alpha^{3D'} , \tag{16}$$

We will discuss the relation between these parameterizations further below – the advantage of $\tilde{k}$, $\tilde{\mu}$ parameterization is that if we choose $f(z)$ to be the Alcock-Paczynski factor [19] at each redshift for something close to the true model, then $\tilde{\mu}$ will be the cosine of the comoving angle corresponding to observable $(k_\parallel, k_\perp)$. The BAO feature will be approximately independent of $\tilde{\mu}$ which should allow relatively coarse interpolation in $\tilde{\mu}$, and the linear redshift space boost of power will be approximately independent of $\tilde{k}$, producing aesthetically pleasing plots.

As discussed in Section IV, we will start our analysis by estimating the likelihood as a function of cross-spectrum parameters, defined as:

$$P_\times(z, \Delta\theta, k_\parallel) = P_\times^{\text{fid}}(z, \Delta\theta, k_\parallel) + \sum_\alpha \Theta_\alpha(\Delta\theta) \, w_\alpha^\times(z, k_\parallel) \, p_\alpha^\times , \tag{17}$$

where $w_\alpha^\times(z, k_\parallel)$ is a bilinear interpolation kernel and $\Theta_\alpha(\Delta\theta)$ is a top-hat kernel in the transverse direction. We chose a nearest-grid-point (NGP) interpolation in the transverse direction because, in combination with the numerical approximation described in IV B, will reduce considerably the number of elements of the Fisher matrix that we need to compute. Using NGP in $w_\alpha^\times(z, k_\parallel)$, on the other hand, would not reduce it further. This will be discussed again in IV B.

We will discuss after explaining our detailed numerical algorithm why we do not from the start define bands in approximate comoving coordinates using a fiducial cosmology. Suffice it to say at this point that this is not necessary as long as we plan to make measurements in sufficiently narrow redshift bins. As long as we use a good interpolation scheme, i.e., one that converges to an exact representation of the function in the limit of fine point-spacing, using different coordinates to represent the same function can only improve the efficiency of the representation, i.e., allow us to represent the function accurately with fewer interpolation points. We will always test explicitly that results of interest have converged with respect to band spacing.

The covariance matrix is given by $\mathbf{C} = \mathbf{C}^{\text{fid}} + \mathbf{C}_{,\alpha} \, p_\alpha^\times + \mathbf{N}$, with the derivative with respect to one of these band power parameters being:

$$\mathbf{C}_{ij,\alpha} = m_i \, m_j \, \Theta_\alpha(\Delta\theta_{ij}) \int \frac{dk_\parallel}{2\pi} \, e^{ik_\parallel \Delta v_{ij}} \, \tilde{W}_i(k_\parallel) \, \tilde{W}_j(k_\parallel) \, w_\alpha(z_{ij}, k_\parallel) , \tag{18}$$

and we have all the pieces required to compute the derivatives of the (log-) likelihood described in section II.

Note that each matrix $\mathbf{C}_{,\alpha}$, corresponding to a cross-spectrum parameter $p_\alpha^\times$ with angular separation $\Delta\theta_\alpha$, will have a very sparse structure, since only the sub-matrices corresponding to pairs of spectra separated by $\Delta\theta_{ij} \in \Delta\theta_\alpha$ will be non-zero. This would not be the case if we wanted to directly estimate the three-dimensional power, where a given pair of pixels would contribute to all band power parameters.

1. Inverting functional parameterizations

We have introduced several parameterization of continuous functions of the form $P(k) = P^{\text{fid}}(k) + \sum_\alpha w_\alpha(k) p_\alpha$, where we use k here to represent all continuous variables. As usual, we describe the likelihood to be a function of data

given a set of parameters $\mathbf{p}$, $L(\mathbf{d}|\mathbf{p})$. Theories generally predict $P(k)$, so to use this kind of likelihood function we need to know how to compute $\mathbf{p}[P(k)]$. We intend to work mostly close to the limit of high resolution parameterization, i.e., k spacing between p_α points smaller than any structure in $P(k)$, where it should be fairly obvious that we can simply take $p_\alpha = P(k_\alpha) - P^{\mathrm{fid}}(k_\alpha)$. However, to approach this limit robustly it is useful to think a little more carefully about this problem. A formal solution is $p_\alpha = \int dk\, w_\alpha^{-1}(k) \left[P(k) - P^{\mathrm{fid}}(k) \right]$, viewing $w_\alpha(k)$ as an $\infty \times N_p$ matrix, however, this matrix is not generally invertible. A natural thing to do is use the pseudo-inverse of $w_\alpha(k)$, equivalent to finding $\mathbf{p}$ to give the least square deviation from the target $P(k)$. Assuming equal weight for all k, i.e., using uniform discretization to make $w_\alpha(k)$ into a finite dimensional matrix, the solution is:

$$p_\alpha = \sum_\beta \left(I^{-1} \right)_{\alpha\beta} \int dk\, \left[P(k) - P^{\mathrm{fid}}(k) \right] w_\beta(k) \qquad (19)$$

with $I_{\alpha\beta} = \int dk\, w_\alpha(k) w_\beta(k)$. We see that $I^{-1}w$ is the response of the parameter to a delta function impulse in $P(k)$ away from $P^{\mathrm{fid}}(k)$. In the simple case of top-hat bands which do not overlap, this reduces to simply averaging $P(k)$ over the band, but for other cases like linear interpolation it is not quite so simple. E.g., for two parameters at $k = 0$ and $k = 1$ determining the $0 < k < 1$ interval by linear interpolation, the effective weight on $P(k)$ for the first $(k = 0)$ parameter is $4 - 6k$, i.e., it peaks at $k = 0$ and goes linearly through zero at $k = 2/3$ to become negative at higher k. Intuitively, if the power is high at $k \sim 1$, the 2nd parameter will be larger, requiring a lower value of the first parameter to describe the intermediate range. When more bins are considered, I is a band matrix with just one additional diagonal, but I^{-1} is not, coupling elements that are more than one bin apart. In less trivial setups we can always evaluate the formula numerically, but we see also that in the limit of narrow $w_\alpha(k)$, relative to the structure in $P(k)$, we are justified using $p_\alpha = P(k_\alpha) - P^{\mathrm{fid}}(k_\alpha)$.

Another approach to this problem is to compute the expected response of estimated parameter values to changes in the true model in much narrower bands [20, 21], which amounts to going back to the data to re-compute the Fisher matrix with one $\partial \mathbf{S}/\partial \mathbf{p}$ replaced with derivatives with respect to the much finer bands. While it is possible that this approach could save some computation over simply making the original measurement bands themselves narrower, in this paper our plan is to always make sure that our end results have converged with respect to the width of the measurement bins, i.e., we will make our measurement bands narrow enough that the detailed structure within them is not important. A natural question at this point is: what is the relationship between this data-derived method and the mathematical pseudo-inverse above? The key is that the discretization required for the pseudo-inverse, assumed to be uniform in k in Eq. (19), is ambiguous as a matter of pure math. Another way to look at the pseudo-inverse is that if P was measured using many finely and uniformly spaced points with equal errorbars, then Eq. (19) would give the best fit for the coarsely interpolated parameterization. The data-derived method determines a more correctly representative, in general non-uniform, weighting within the bins.

D. Connection to the traditional 1D power spectrum

In the limit of zero angular separation, the cross-spectrum becomes the traditional 1D power spectrum:

$$P_{1D}(z, k_\parallel) \equiv P_\times(z, \Delta\theta = 0, k_\parallel) = \int \frac{d\mathbf{k}_\perp}{(2\pi)^2}\, P_{3D}(z, k_\perp, k_\parallel) \,. \qquad (20)$$

It is very common in 1D analyses to ignore the correlation between the fluctuations in neighboring spectra, resulting in a block-diagonal covariance matrix and in a likelihood that is factorizable. The derivatives of the log-likelihood are then additive, and we recover the method to estimate the 1D power spectrum presented in [7] and later used in [8].[8]

IV. NUMERICAL IMPLEMENTATION

A naive implementation of the algorithm implied by sections II and III is computationally intractable, since it involves thousands of linear algebra operations with products of extremely large matrices, of size set by the number of pixels in the survey ($\approx 10^8$ for relevant surveys).

----

[8] Note, however, that the redshift interpolation of the power is better defined in our analysis. For instance, [8] split each spectrum into two or three segments, and assumed that each segment contributed to a single redshift bin.

In this section we will discuss some approximations and numerical implementations that make the problem tractable, and that allow us to analyze the BOSS data set typically using a few thousand CPU hours depending on the number of bins used.

We will start in IV A by choosing a parameterization of the likelihood in which a quasar pair contributes only to a small fraction of parameters. In IV B we present a particular fiducial clustering model that allows us to have a block-diagonal covariance matrix, reducing drastically the size of the linear algebra operations required. In IV C we show that the information in the response matrices can be described by a very small number of eigenmodes. Finally, in IV D we further speed up the algorithm by internally using Fast Fourier Transforms (FFT) to compute some of the matrix multiplications.

A. Cross-spectrum

We aim to estimate first the cross-spectrum, because using it each pair of quasars contributes only to a single parameterized band in the transverse coordinate (angular separation), instead of contributing to all parameters if we went straight to a 3D power spectrum. Estimating $P_\times$ from the raw data is the computationally demanding step. After that, the data set is massively compressed and we have a lot of options to transform to other parameterizations, which we will discuss below. To be clear: it is the sparsity of the quasars on the sky in real data sets, relative to the scales of interest, that motivates approaching them pair-wise, and therefore using the cross-spectrum. This is analogous to the efficiency of estimating a 3D correlation function pair-wise for a sparse set of point objects (which could similarly be converted to a power spectrum by the methods below). If we had much denser sampling (or wanted to efficiently push the measurement to much larger scales) we would be motivated to use a method that, e.g., compressed spectra in an angular pixel into some aggregate measure of the flux in that pixel, transformed in the transverse directions, and then measured the 3D power spectrum directly. To cover both regimes we could imagine something akin to the P^3M approach to N-body simulations (e.g., [22–24] – generally, any of the ways similar problems have been solved for N-body simulations could be useful).

B. Zero cross-correlation fiducial model gives block-diagonal $\mathbf{C}_0^{-1}$

In order to measure 3D correlations we need to consider the full likelihood, and we need to choose a fiducial power $P_\times^{\mathrm{fid}}(z, \Delta\theta, k_\parallel)$ in order to specify the global weighting matrix $\mathbf{C}_0$. The ideal choice would be to use our global best guess of the <u>true</u> power spectrum, a ΛCDM model with linear bias and redshift space distortions and appropriate non-linear correction on smaller scales [12, 25, 26]. If we choose a different model, our Taylor series generally will not be as accurate near the true model as discussed in Section II.

In this work we take the fiducial cross-spectrum to be

$$P_\times^{\mathrm{fid}}(z, \Delta\theta, k_\parallel) = P_{1D}^{\mathrm{fid}}(z, k_\parallel)\ \delta^D(\Delta\theta), \tag{21}$$

i.e. it is vanishing everywhere except at zero angular separation. This makes the weighting matrix block-diagonal giving a huge speed up. Because variance of the field is dominated by very small scale modes, or equivalently, because the $P_\times$ is very rapidly falling with $\Delta\theta$ (as shown in Figure 6 below), this is a reasonably good approximation to the truth. This is further supported by the study presented in Section VII below, and in particular in Figure 8, where we show that any possible underestimation of the measured uncertainties caused by this approximation is smaller than 6-8%.

We use subscripts I and J to identify blocks corresponding to a particular quasar. We can now describe the (log-) likelihood derivatives as sum over blocks:

$$\mathcal{L}_{,\alpha} = \frac{1}{2}\sum_I \sum_J \mathbf{y}_I^t\ \mathbf{C}_{IJ,\alpha}\ \mathbf{y}_J - \frac{1}{2}\sum_I \mathrm{Tr}\left[\mathbf{C}_{0\ I}^{-1}\ \mathbf{C}_{II,\alpha}\right]\ , \tag{22}$$

and

$$\mathcal{L}_{,\alpha\beta} \approx -F_{\alpha\beta} = -\frac{1}{2}\sum_I \sum_J \mathrm{Tr}\left[\mathbf{C}_{0\ I}^{-1}\ \mathbf{C}_{IJ,\alpha}\ \mathbf{C}_{0\ J}^{-1}\ \mathbf{C}_{JI,\beta}\right]\ , \tag{23}$$

where we have also defined $\mathbf{y} = \mathbf{C}_0^{-1}(\mathbf{d} - \mathbf{m})$ and $\mathbf{C}_{IJ,\alpha}$ is the sub-matrix of $\mathbf{C}_{,\alpha}$ corresponding to the cross-correlation of pixels in quasars I and J. From equation 22 above it is clear that the second term of the first derivative, the one with a trace, will only be non-zero for derivatives with respect to P_{1D} parameters. With our top-hat parameterization in the $\Delta\theta$ direction, the Fisher matrix is non-zero only for parameters corresponding to the same value of $\Delta\theta$.

C. Eigenmode decomposition of $\mathbf{C}_{,\alpha}$ matrices

We started with a problem that was computationally impossible, and thanks to the choice of parameters ($\mathbf{p}^{\times}$ instead of $\mathbf{p}^{3D}$) and the choice of weighting matrices (block diagonal $\mathbf{C}_0$) we have now a problem that is just extremely computationally intensive.

Let us consider a BOSS like survey, with an area of 10 000 square degrees and a density of lines of sight of roughly 15 per square degree. If we are interested in measuring 3D correlation up to 3 degrees (roughly 200 h^{-1}Mpc), we will need to correlate each spectrum with an average of 400 neighbors, for a total of roughly 30 million pairs of spectra (not counting the same pair twice). If we are measuring a total of 1000 cross-power spectrum parameters (as discussed below we typically use several thousand parameters), the Fisher matrix (or second derivative) will have a million elements $F_{\alpha\beta}$, and even if we consider only the most relevant covariances we still need to compute several tens of thousand elements. As shown in equation 23, for each element $F_{\alpha\beta}$ we need to compute the trace of 30 million products of four matrices, of size equal to the number of pixels in a spectrum, roughly 500.

The algorithm described above would work on a small data set, with limited number of parameters, but it quickly becomes too slow when applied to any relevant survey. In this section we will introduce numerical implementations that will speed up dramatically the algorithm without adding further approximations.

To motivate the math below, let us consider first the simplified example of a measurement on a very large periodic box, with parallel lines of sight and a static universe, and ignoring pixelization or resolution effects. In this case we could define the following modes for each line of sight:

$$\tilde{\delta}(\theta, k_{\parallel}) = \int_{-\infty}^{\infty} dv \; e^{ik_{\parallel}v} \; \delta(\theta, v) \;, \tag{24}$$

and it is easy to show that these modes would be uncorrelated for different values of $k_{\parallel}$, and that their correlations would be given by the cross-power spectrum:

$$\left\langle \tilde{\delta}(\theta, k_{\parallel}) \; \tilde{\delta}(\theta', k_{\parallel}') \right\rangle = (2\pi) \; \delta^D(k_{\parallel} + k_{\parallel}') \; P_{\times}(\Delta\theta, k_{\parallel}) \;, \tag{25}$$

where $\delta^D(k_{\parallel})$ is the Dirac delta function. In this space, both the covariance matrix $\mathbf{C}_{0\,I}$ and its derivatives $\mathbf{C}_{IJ,\alpha}$ would be diagonal. Redshift evolution and the finite length of the lines of sight in a realistic survey complicate this picture, but as we show below most of the information in the matrices $\mathbf{C}_{IJ,\alpha}$ is still captured in a small number of modes.

So we would like to project each spectrum onto some kind of Fourier-like modes, but while the parameter response matrices $\mathbf{S}_{,\alpha}$ are localized in redshift, Fourier modes are not. For this reason we could hope to do better by using Fourier transforms over a set of ad hoc redshift windows. What we do will be qualitatively similar to this, but with the modes and envelopes more carefully derived. Our first thought to derive modes might be to use Karhunen-Loeve (KL) eigenmodes [27], which are generally eigenmodes of $\mathbf{L}^{-1}\mathbf{C}_{,\alpha}\mathbf{L}^{-t}$, where $\mathbf{C} = \mathbf{L}\mathbf{L}^t$ is a Cholesky decomposition. As we have set things up, each quasar pair would have its own set of $\mathbf{L}^{-1}\mathbf{C}_{IJ,\alpha}\mathbf{L}^{-t}$ matrices, covering a different set of wavelengths, and with different pixelization and resolution. We would need to compute the eigenmode decomposition of all of them, which is easier than computing KL modes for the fully coupled data set, but would still make the algorithm intractable.

For this reason, we will first extend a little bit our data model. Let us define a common pixel grid covering the whole wavelength range of the survey.[9] We can map all our spectra into that grid, giving effectively infinite noise to cells that are not covered by a given spectrum (or masked by the pipeline). If we discretize the integral over wavelength in equation 7 we can rewrite the data model as:

$$\mathbf{d} = \mathbf{R}\delta + \mathbf{m} + \epsilon \;, \tag{26}$$

with $R_{ij} = m_i \, W_i(\lambda_j)$, and the covariance matrix is:

$$\mathbf{C} = \left\langle (\mathbf{d} - \mathbf{m}) \, (\mathbf{d} - \mathbf{m})^t \right\rangle = \mathbf{R}\mathbf{S}\mathbf{R}^t + \mathbf{N} \tag{27}$$

where $\mathbf{S}$ is now the signal covariance without pixel or resolution effects.[10]

---

[9] In particular we use a grid of 2450 cells, of same width as in BOSS coadded spectra $dv = 69.03$km s^{-1}, covering the redshift range $1.74 < z < 3.81$.

[10] To prevent numerical artifacts when evaluating Equation 29 we compute $\mathbf{S}$ with a small Gaussian smoothing of $G_0 = 20$km s^{-1}, and correct for it by using $\sqrt{G_i^2 - G_0^2}$ as the pixel resolution when computing the resolution matrices $\mathbf{R}$. We have checked that the results do not depend on the exact value of G_0 used.

The derivatives of the covariance matrix for a pair of spectra are now:

$$\mathbf{C}_{IJ,\alpha} = \mathbf{R}_I \mathbf{S}_{IJ,\alpha} \mathbf{R}_J^t \ , \tag{28}$$

and similar to equation 18 we can write them as:

$$S_{IJij,\alpha} = \Theta_\alpha(\Delta\theta_{IJ}) \int \frac{dk_\parallel}{2\pi} \ e^{-ik_\parallel \Delta v_{ij}} \ w_\alpha(z_{ij}, k_\parallel) \equiv \Theta_\alpha(\Delta\theta_{IJ}) \ \tilde{S}_{ij,\alpha} \ . \tag{29}$$

Equation 29 implies that: i) the derivative matrix $\mathbf{S}_{IJ,\alpha}$ will be zero unless $\Delta\theta_{IJ} \in \Delta\theta_\alpha$ and ii) if non-zero, the matrix $\tilde{\mathbf{S}}_{,\alpha}$ will only depend on (z_α, k_α) and will be the same for all quasar pairs. Therefore, if we choose to project onto eigenmodes of $\tilde{\mathbf{S}}_{,\alpha}$, we will only need to decompose as many matrices $\tilde{\mathbf{S}}_{,\alpha}$ as parameters in our analysis. In Figure 1 we show a couple of $\tilde{\mathbf{S}}_{,\alpha}$ matrices that might appear in a typical analysis.
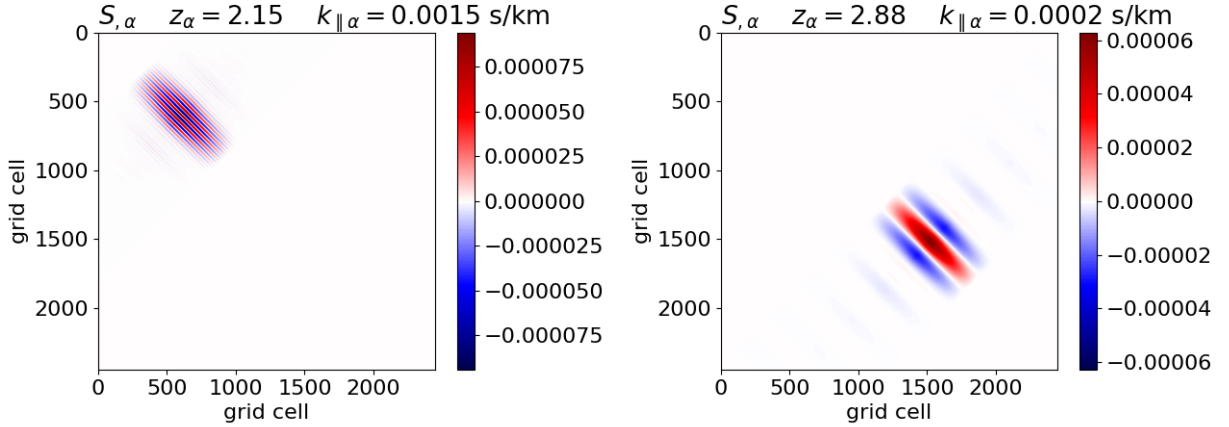

FIG. 1. $\tilde{\mathbf{S}}_{,\alpha}$ matrices for two different parameters. Left: eighth k-bin ($k_\parallel = 0.0015$s km^{-1}) of the second z-bin ($z = 2.15$); Right: first non-zero k-bin ($k_\parallel = 0.0002$s km^{-1}) of the fifth z-bin ($z = 2.88$). The redshift bin (z-bin) of the parameter sets the diagonal band that is non-zero, and the wavenumber bin (k-bin) sets the oscillation frequency.

Now we simply *exactly diagonalize* $\tilde{\mathbf{S}}_{,\alpha}$, i.e., find $\tilde{\mathbf{S}}_{,\alpha} = \mathbf{T}_\alpha \mathbf{D}_\alpha \mathbf{T}_\alpha^t$ where $\mathbf{D}_\alpha$ is diagonal. This is motivated by the fact that in the limit of infinitely narrow k-bins and infinitely large redshift bin, these signal matrices would have exactly two non-zero eigenvalues.[11] A matrix with finite redshift and k-bins is thus expected to have a few dominant eigenvectors.

In Figure 2 we present the distribution of eigenvalues for the $\tilde{\mathbf{S}}_{,\alpha}$ matrices presented in figure 1, sorted by absolute value. In both cases the number of relevant eigenvalues is very small (6 or 7 out of 2450). The dotted lines show the $\pm e_{\min} = 0.02$ of the maximum eigenvalue, and we generally neglect those eigenmode with an eigenvalue smaller (in absolute value) than that. This means that not only the $\mathbf{D}_\alpha$ matrices presented above are diagonal, but only a handful of elements are considered non-zero.

---

[11] To see this, note that a single Fourier mode of any phase will have correlation function $\xi(\Delta v) \propto \langle \sin(\phi) \sin(\phi + k\Delta v) \rangle \propto \cos(k\Delta v)$ and that $C_{ab} \propto \cos(k(a-b)\ dv) = \cos(ka\ dv)\cos(kb\ dv) + \sin(ka\ dv)\sin(kb\ dv)$, where dv is the cell width of the grid. In this case the pixel covariance can be written as $C(r) \propto [\cos(kv)]^T[\cos(kv)] + [\sin(kv)]^T[\sin(kv)]$.
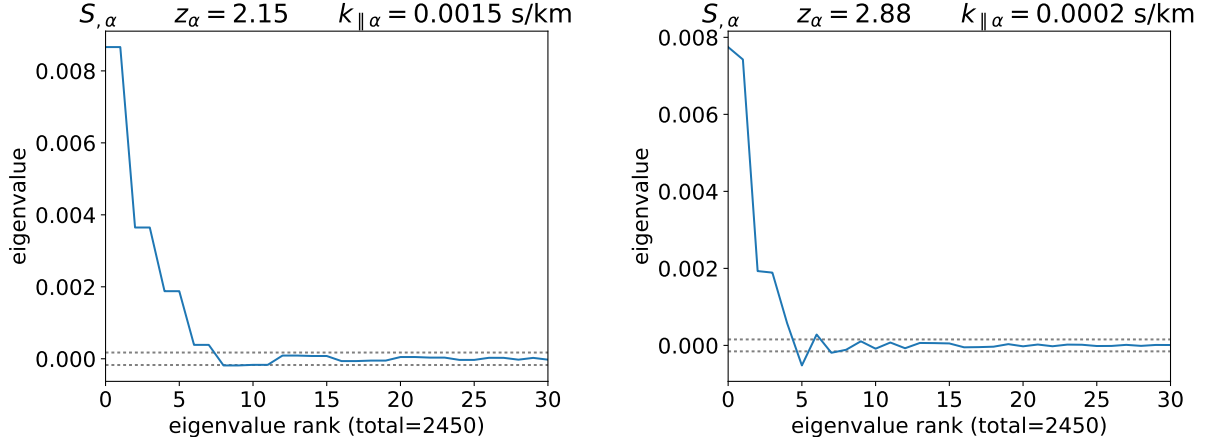
FIG. 2. Eigenvalues for the two $\tilde{\mathbf{S}}_{,\alpha}$ matrices shown in Figure 1, sorted by absolute value (the total number of eigenvalues is 2450). The dotted lines show the $\pm 2\%$ of the maximum eigenvalue, and we ignore eigenmodes with an eigenvalue smaller than that (in absolute value). Note that negative eigenvalues are legitimate, reflecting the fact that $\tilde{\mathbf{S}}_{,\alpha}$ is not itself a legitimate (positive definite) covariance matrix.

In Figure 3 we show the first eigenvector for the same matrices $\tilde{\mathbf{S}}_{,\alpha}$. The left panel corresponds to a parameter with wavenumber 7.5 times larger than the wavenumber in the right panel. They are both localized in a region of the grid corresponding to the redshift of the parameter in the derivative, as shown by the vertical dotted lines.
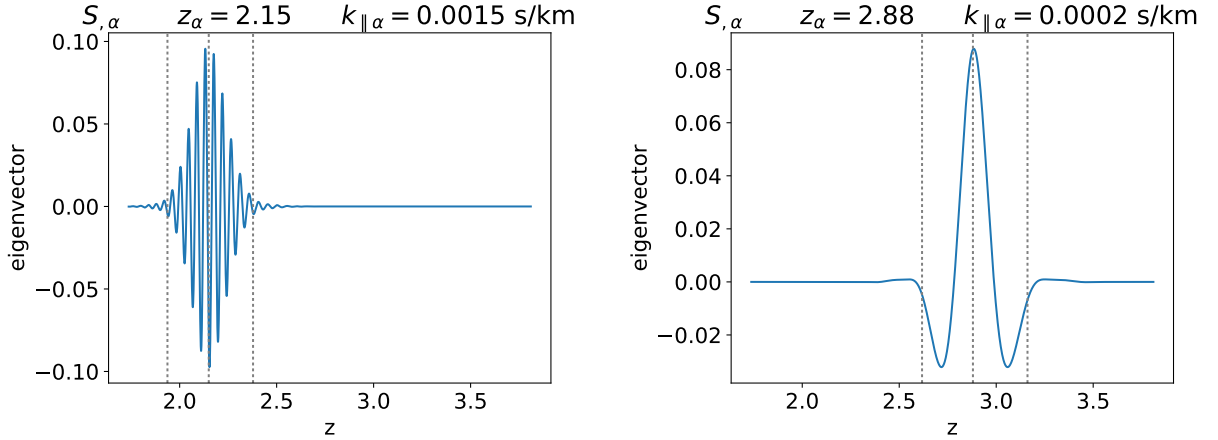

FIG. 3. First eigenvector for the $\tilde{\mathbf{S}}_{,\alpha}$ matrices in figure 1. The vertical dotted lines show the center and the edges of the redshift bin z_α for the parameter. The wavenumber $k_{\parallel\alpha}$ of the parameter in the left panel is 7.5 times larger than the one in the right panel.

We will compute the eigendecomposition only once per parameter, and use the decomposition for all quasars pairs. With this diagonalization we can now look again at the derivatives. The first term in the first derivatives, i.e., the one involving the data vector, will now be:

$$
\begin{aligned}
\mathbf{y}_I^t \ \mathbf{C}_{IJ,\alpha} \ \mathbf{y}_J &= \mathbf{y}_I^t \ \mathbf{R}_I \ \tilde{\mathbf{S}}_{,\alpha} \ \mathbf{R}_J^t \ \mathbf{y}_J \\
&= \mathbf{y}_I^t \ \mathbf{R}_I \ \mathbf{T}_\alpha \mathbf{D}_\alpha \mathbf{T}_\alpha^t \ \mathbf{R}_J^t \ \mathbf{y}_J \\
&= \mathbf{y}_{I\alpha}^t \ \mathbf{D}_\alpha \ \mathbf{y}_{J\alpha} \ ,
\end{aligned}
\tag{30}
$$

where $\tilde{\mathbf{S}}_{,\alpha}$ is the same for all quasars pairs that contribute to this particular angular separation bin, $\mathbf{y}_I = \mathbf{C}_{0\ I}^{-1}(\mathbf{d}_I - \mathbf{m}_I)$ is now formally defined over all wavelengths but it is approximately limited to the range actually covered by the

spectrum, and we have defined $\mathbf{y}_{I\alpha} = \mathbf{T}_\alpha^t \, \mathbf{R}_I^t \, \mathbf{y}_I$. For each parameter, $\mathbf{y}_{I\alpha}$ is akin to the collection of Fourier modes in the k range corresponding to the parameter, for a transform windowed to the z range covered by the parameter.

Similarly we can expand the second term of the first derivative:

$$
\begin{aligned}
\mathrm{Tr}\left[\mathbf{C}_{0\ I}^{-1} \, \mathbf{C}_{II,\alpha}\right] &= \mathrm{Tr}\left[\mathbf{C}_{0\ I}^{-1} \, \mathbf{R}_I \, \tilde{\mathbf{S}}_{,\alpha} \, \mathbf{R}_I^t\right] \\
&= \mathrm{Tr}\left[\mathbf{C}_{R\ I}^{-1} \, \mathbf{T}_\alpha \mathbf{D}_\alpha \mathbf{T}_\alpha^t\right] && (31) \\
&= \mathrm{Tr}\left[\mathbf{M}_{I\alpha\alpha}\mathbf{D}_\alpha\right] && (32)
\end{aligned}
$$

and the second derivative (or Fisher matrix):

$$
\begin{aligned}
\mathrm{Tr}\left[\mathbf{C}_{0\ I}^{-1} \, \mathbf{C}_{IJ,\alpha} \, \mathbf{C}_{0\ J}^{-1} \, \mathbf{C}_{JI,\beta}\right] &= \mathrm{Tr}\left[\mathbf{C}_{0\ I}^{-1} \, \mathbf{R}_I \, \tilde{\mathbf{S}}_{,\alpha} \, \mathbf{R}_J^t \, \mathbf{C}_{0\ J}^{-1} \, \mathbf{R}_J \, \tilde{\mathbf{S}}_{,\beta} \, \mathbf{R}_I^t\right] \\
&= \mathrm{Tr}\left[\mathbf{C}_{R\ I}^{-1} \, \mathbf{T}_\alpha \mathbf{D}_\alpha \mathbf{T}_\alpha^t \, \mathbf{C}_{R\ J}^{-1} \, \mathbf{T}_\beta \mathbf{D}_\beta \mathbf{T}_\beta^t\right] \\
&= \mathrm{Tr}\left[\mathbf{M}_{I\beta\alpha} \, \mathbf{D}_\alpha \, \mathbf{M}_{J\alpha\beta} \, \mathbf{D}_\beta\right] \ , && (33)
\end{aligned}
$$

where we have defined $\mathbf{C}_{R\ I}^{-1} = \mathbf{R}_I^t \mathbf{C}_{0\ I}^{-1} \, \mathbf{R}_I$ and $\mathbf{M}_{I\alpha\beta} = \mathbf{T}_\alpha^t \, \mathbf{C}_{R\ I}^{-1} \, \mathbf{T}_\beta$. We will have to compute the objects $\mathbf{y}_{I\alpha}$ and $\mathbf{M}_{I\alpha\beta}$ only once for each spectrum, and once we have them the cross-correlations will be very fast.

Note that since we only consider $N \sim 7$ eigenmodes (see Figure 2), the diagonal matrices $\mathbf{D}_\alpha$ have only N non-zero elements. This means that we only need to compute a very small block $N \times N$ of elements of the $\mathbf{M}_{I\alpha\beta}$ matrices.

D. FFT of eigenvectors

The eigenmode decomposition described above moved the computational challenge from the spectra cross-correlations to the computation of *per-spectrum* objects: $\mathbf{y}_{I\alpha}$ and $\mathbf{M}_{I\alpha\beta}$. In particular the slowest part is to compute the rotated inverse covariances:

$$
\mathbf{M}_{I\alpha\beta} = \mathbf{T}_\alpha^t \, \mathbf{C}_{R\ I}^{-1} \, \mathbf{T}_\beta \ , \tag{34}
$$

for which naively we need to compute a product of three matrices of size equal to the number of cells in our radial grid, 2450. For each quasar I, naively we need to compute this product for each pair of matrices $(\mathbf{T}_\alpha, \mathbf{T}_\beta)$, specified by their values of $(z, k_\parallel)$.[12] In the analysis presented in Section VI, this would imply over 10^4 matrices per quasar. On the other hand, we know that in the idealized high symmetry case this rotation amounts to FFTs along each row and column which can be done orders of magnitude faster.

Fortunately, matrices $\mathbf{M}_{I\alpha\beta}$ are only used to compute the elements of the Fisher matrix $F_{\alpha\beta}$, as shown in Equation 33. Parameters that have different values of either z or $k_\parallel$ will be very weakly correlated, and therefore we decide to compute only a subset of the elements of the Fisher matrix. In particular, we ignore the correlation of parameter pairs with either $|\Delta k_\parallel| > F_k$ or $|\Delta \ln(1+z)| > F_z + 2\pi/(c\,\bar{k}_\parallel)$, where $\bar{k}_\parallel$ is the geometric mean and c is the speed of light, with the exception of neighboring parameters that are always included.[13] In our default analysis, this reduces the number of $\mathbf{M}_{I\alpha\beta}$ matrices we need to compute by an order of magnitude.

The eigenvectors shown in figure 3 have clear oscillatory features, what suggests that they would be very sparse in Fourier space. For this reason we introduce discrete Fourier transform and inverse Fourier transform matrices pairs $\mathbf{F}\,\mathbf{F}^{-1}$ in the computation of the per-quasar objects:

$$
\begin{aligned}
\mathbf{M}_{I\alpha\beta} &= \mathbf{T}_\alpha^t \, \mathbf{C}_{R\ I}^{-1} \, \mathbf{T}_\beta \\
&= \mathbf{T}_\alpha^t \, \mathbf{F}^t \, \mathbf{F}^{-t} \, \mathbf{C}_{R\ I}^{-1} \, \mathbf{F}^{-1} \, \mathbf{F} \, \mathbf{T}_\beta \\
&= \tilde{\mathbf{T}}_\alpha^t \, \tilde{\mathbf{C}}_{R\ I}^{-1} \, \tilde{\mathbf{T}}_\beta \ , && (35)
\end{aligned}
$$

where we have defined $\tilde{\mathbf{T}}_\alpha = \mathbf{F}\,\mathbf{T}_\alpha$ and $\tilde{\mathbf{C}}_{R\ I}^{-1} = \mathbf{F}^{-t} \, \mathbf{C}_{R\ I}^{-1} \, \mathbf{F}^{-1}$. Note that while $\tilde{\mathbf{C}}_{R\ I}^{-1}$ is a 2D (inverse) Fourier transform, $\tilde{\mathbf{T}}_\alpha$ is a 1D Fourier transform, with each column of the matrix containing the 1D Fourier transform of each eigenvector.

----

[12] Remember that in Equation 29 we introduced $\tilde{\mathbf{S}}_\alpha$ as matrices that do not depend on $\Delta\theta_\alpha$.

[13] The last term was added to allow correlations of the lowest values of $k_\parallel$ over all redshift ranges.
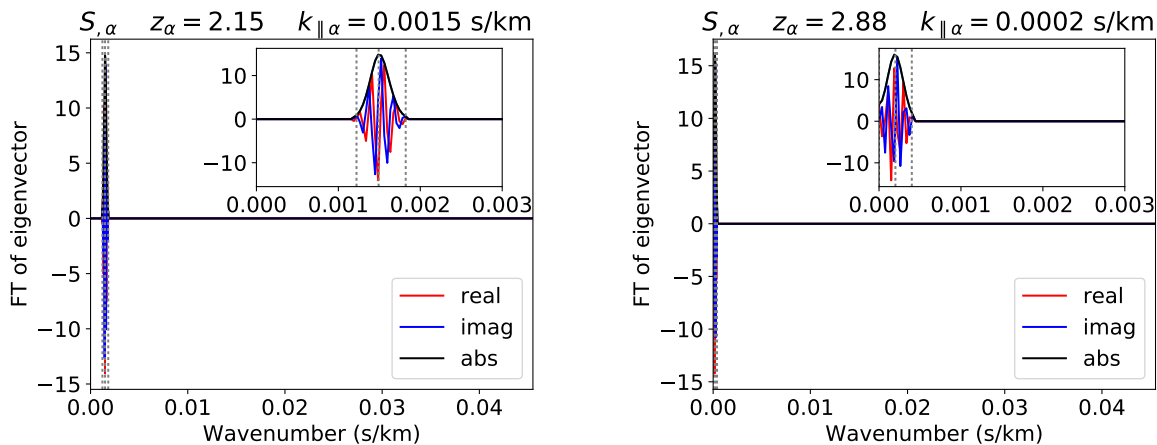
FIG. 4. Fourier transform of the eigenvectors shown in Figure 3, from the decomposition of the $\tilde{\mathbf{S}}_{,\alpha}$ matrices shown in Figure 1. The vertical dotted lines show the center and the edges of the wavenumber bin $k_{\parallel\alpha}$ for the parameter. The inner sub-plot zooms in to a shorter wavenumber range to better show the very compact eigenvector in Fourier space.

In Figure 4 we show the Fourier transform of the eigenvectors shown in figure 3. It is the sparsity of these modes what results in another big speed up in the computation of the different $\mathbf{M}_{I\alpha\beta}$ objects.

The modes for some eigenvectors with negative eigenvalues are not as localized as those shown in figure 4, and for this reason we apply a cut in the tails of the Fourier modes, and we only keep $f_{\mathrm{norm}} = 0.999$ of its norm.

In order to speed up the computation of equation 35, we use another approximation. In the limit of infinitely long windows, the matrices $\tilde{\mathbf{C}}_{R\,I}^{-1}$ would be diagonal. We checked that including off-diagonal correlations between modes separated by less than $\Delta k_{\mathrm{off}} = 5.57 \times 10^{-4}\mathrm{s\ km}^{-1}$ (corresponding to 15 off-diagonal elements in our default setting) was enough to recover unbiased results in the analysis on mock data sets presented in Section V.

E. Summary of numerical optimizations

To summarize, we are trying to evaluate equations 2 and 3. The brute force evaluation would require us to deal with matrices of size $N_q N_p$ where N_q is the number of quasars and N_p is the number of pixels in a spectrum. We do the following approximations.

- Expand around the model with zero cross-power, in effect reducing the covariance matrix to N_q dense matrices of size N_p. All relevant terms can now be calculated by manipulating (millions of) $N_p \times N_p$ matrices.

- Diagonalizing the $N_p \times N_p$ matrices of derivatives of covariance matrix with respect to clustering parameters. This allows us to project relevant eigenvectors separately for each spectrum, speeding up consequent operations.

- Computing the projections of the weighted data vector and more importantly inverse covariance matrix ($\mathbf{y}_{I\alpha}$ and $\mathbf{M}_{I\alpha\beta}$) by rotating to Fourier space where the signal eigenvectors become very sparse, and $\tilde{\mathbf{C}}_{R\,I}^{-1}$ fairly diagonal, then rotating back.

There are several parameters that control the accuracy of the latter two approximations, which we summarize in Table I.

Parameter	Default value	Meaning
e_{min}	0.02	Keep eigenmodes whose absolute eigenvalue is larger that e_{min} of maximum eigenvalue.
f_{norm}	0.999	Cut tails of eigenvector FFTs so that it contains f_{norm} of its norm.
Δk_{off}	5.57×10^{-4}	Ignore correlations in $\tilde{\mathbf{C}}_{R\,I}^{-1}$ between modes separated by more than Δk_{off} (in s km^{-1}).
F_z	0.175	Ignore elements of Fisher matrix ($F_{\alpha\beta}$) with $|\Delta \ln(1+z)| > F_z$.
F_k	0.0005	Ignore elements of Fisher matrix ($F_{\alpha\beta}$) with $|\Delta k_{\parallel}| > F_k$ (in s km^{-1}).

TABLE I. Parameters controlling the accuracy of our approximations.

V. ANALYSIS OF THE 1D POWER SPECTRUM ON MOCK DATA

We will begin our numerical exploration by using mock datasets to compare measurements of the 1D power spectrum $P_{1D}(z, k_\parallel)$ implemented using brute force real space matrix multiplication (as done in, e.g., [7]) with our numerical implementation presented in IV. The main motivation here is to establish basic sanity of our numerical work and to test the validity of approximations that we describe in Section IV. Measurements of 1D power spectrum alone are sufficiently fast that we can afford to switch these approximations off and see the impact that they have.

A. Mock dataset for $P_{1D}(z, k_\parallel)$ measurements

We simulate data mimicking the quasar spectra from the twelfth data release (DR12, [28]) of the Sloan Digital Sky Survey (SDSS-III, [29]), containing mostly spectra from the Baryon Oscillation Spectroscopic Survey (BOSS, [30]). We generate mock spectra for 181,506 quasar spectra with $z > 2.1$, using quasar continua matching the amplitude of each observed quasar, and adding Gaussian noise to the spectra using the noise variance estimated by the SDSS pipeline.

The mock spectra have a mean transmitted flux fraction $\bar{F}(z)$ that approximately matches the one measured in the data, and Gaussian fluctuations around this mean are generated using an analytical expression for $P_{1D}(z, k_\parallel)$ from [8], corrected to be flat at very low-k. A continuum for each spectrum is generated using a mean restframe shape for all quasars, redshifted and normalized accordingly to roughly match the observed magnitude. Finally, the spectra are convolved using the mean resolution within each forest, pixelized using the pixel width of the SDSS coadded spectra, $T = 69.03 \mathrm{km\ s^{-1}}$, and Gaussian noise is added using the variance provided by the SDSS pipeline.

B. Band power measurement

We parameterize $P_{1D}(z, k_\parallel)$ following equation 17, for the special case of $\Delta\theta = 0$, and we use the same $P_{1D}^{\mathrm{fid}}(z, k_\parallel)$ that was used to generate the mock spectra. [14] We will use a grid of band power parameters with $N_z = 7$ *z-bins* and $N_k = 21$ *k-bins*, for a total of 147 parameters. The first z-bin is centered at $z = 1.936$, and the other bins are separated by $\Delta\ln(1 + z) = 0.07$. The last z-bin is centered at $z = 3.468$. The first k-bin is centered at $k_\parallel = 0$, and the following 5 bins are linearly spaced with a separation of $\Delta k_\parallel = 0.0002 \mathrm{\ s\ km^{-1}}$. The last 15 k-bins increase logarithmically with $\Delta\ln k_\parallel = 0.2$, with the last bin centered at $k_\parallel = 0.0201 \mathrm{\ s\ km^{-1}}$.

We present the measured power in figure 5, where we plot the measured power divided by the input power used to generate the mock spectra. The definition of "measured power" in all of our plots is the maximum likelihood implied by our Taylor expansion of the likelihood, equivalent to the result of a single NR step given by Equation 5. The errorbars are the square root of the diagonal elements of the inverse of the Fisher matrix. The dotted lines show the measurement using brute force matrix multiplication, while the dashed lines shows the measurement when using the $P_\times$-oriented implementation described in section IV. We see that both methods agree well with each other and the expectation for these mocks, with the biggest deviations maybe not surprisingly appearing at the extremes of k and z.

---

[14] We have tested that we recover the right power spectrum when starting with a fiducial not far from the truth.
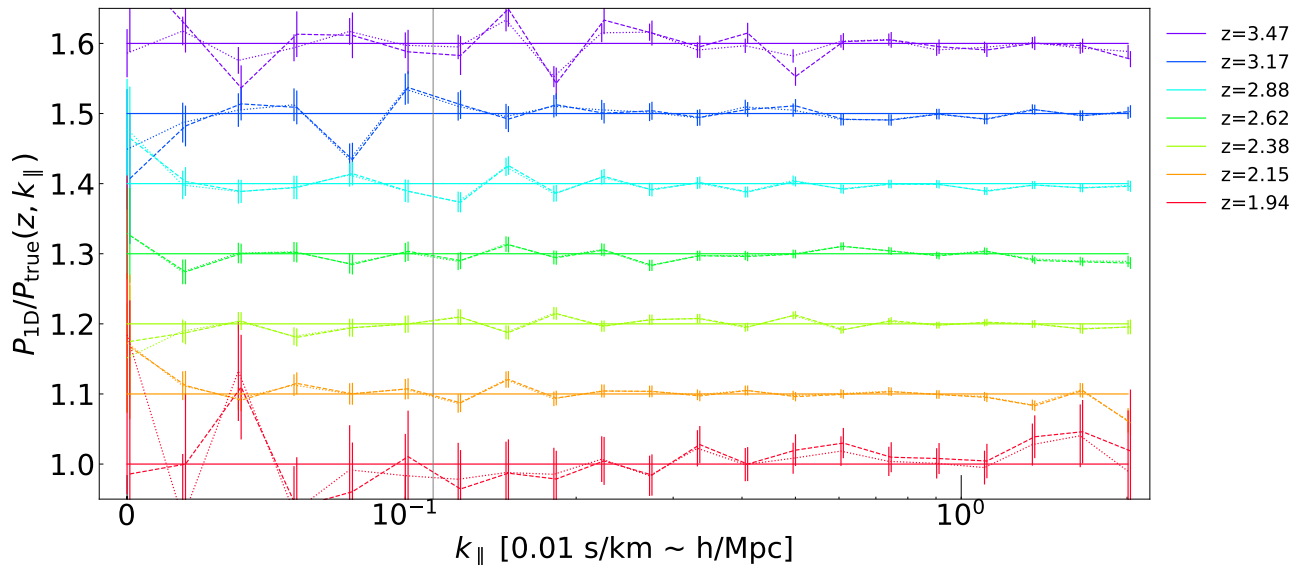
FIG. 5. Ratio of P1D measured over input power used to generate the mocks, for the different redshift bins as measured using the standard likelihood estimator (dotted lines) and the $P_\times$-oriented implementation presented in section IV (dashed lines, with linear points shifted by 0.001 and logarithm points multiplied by 1.01 for better visualization). The vertical line divides the linearly spaced bins (to the left) from the logarithmically spaced bins (to the right). The wavenumbers in the x-axis have been multiplied by 100 so they can be approximately compared to wavenumbers in units of $h\mathrm{Mpc}^{-1}$.

C.   Goodness of fit - χ^2

In the $P_{1D}(z, k_\parallel)$ analysis above we computed the Taylor expansion of the log-likelihood around the true power spectrum $P_{1D}^{\mathrm{fid}}(z, k_\parallel)$. In this case, the expected value of the first derivatives is zero, and their covariance equals the Fisher matrix, $\langle \mathcal{L}_{,\alpha} \mathcal{L}_{,\beta} \rangle = F_{\alpha\beta}$. The standard χ^2 can then be computed directly from the derivatives using $\chi^2 = \mathcal{L}_{,\alpha} F_{\alpha\beta}^{-1} \mathcal{L}_{,\beta}$ and its expected value is equal to the number of band power parameters.

In table II we show the performance of the different algorithms, and the sensitivity to the different approximations discussed in Section IV. When using all the (147) bins in the measurement, we measure a value of $\chi^2 = 155.1$ for the exact matrix multiplication analysis (probability of 30.7%), and a value of $\chi^2 = 176.9$ for our new default implementation (probability of 4.7%). $k_\parallel = 0.02\mathrm{s}\ \mathrm{km}^{-1}$ is the limit typically used for data of this resolution, since on large scales the suppression of power caused by the spectral resolution is very large and difficult to correct for accurately. If we marginalize over the last $k_\parallel$ bin at $k = 0.0201\mathrm{s}\ \mathrm{km}^{-1}$, the agreement with the new implementation is considerably improved to $\chi^2 = 156.0$ for 140 degrees of freedom (d.o.f), and the probability is of 16.9%. In Appendix A we describe how we marginalize the likelihood over unwanted parameters.

In table II we also show the effect of the different numerical settings on the values of χ^2 as a function of $k_{\parallel\ \mathrm{max}}$.

Analysis	χ^2, prob		
$k_{\parallel\ \mathrm{max}}[\mathrm{s\ km}^{-1}]$ (d.o.f.)	0.0201 (147)	0.0164 (140)	0.0135 (133)
Default settings	176.9 0.047	156.0 0.169	139.9 0.323
$F_z = 0.1$	176.9 0.047	156.0 0.169	139.9 0.323
$F_k = 0.00045$	176.9 0.047	156.0 0.168	140.0 0.322
$e_{\mathrm{min}} = 0.03$	177.4 0.044	156.4 0.162	140.3 0.316
$f_{\mathrm{norm}} = 0.995$	184.4 0.020	163.5 0.085	144.2 0.239
$\Delta k_{\mathrm{off}} = 7.43 \times 10^{-4}\mathrm{s\ km}^{-1}$	171.5 0.081	152.0 0.230	136.9 0.390
exact matrix ops	155.1 0.307	139.0 0.508	125.3 0.670

TABLE II. For a single mock BOSS data set, χ^2 and probability for the different P_{1D} analyses, when using different maximum values for $k_{\parallel}$. The default analysis uses $F_z = 0.175$, $F_k = 0.0005$, $e_{\mathrm{min}} = 0.02$, $f_{\mathrm{norm}} = 0.999$ and $\Delta k_{\mathrm{off}} = 5.57 \times 10^{-4}\mathrm{s\ km}^{-1}$. Note that the only purpose of this analysis is to help understand the 3D algorithm – we would not use this algorithm if we were only interested in 1D power.

One can see that we are not very sensitive to the details of the numerical implementation, although Δk_{off} does seem to make a non-negligible difference. Our algorithm is not optimized for P_{1D}, so we take these results as good enough and move on to our real goal of measuring 3D cross-correlations.

VI. ANALYSIS OF THE CROSS-POWER ON MOCK DATA

After having established validity of the method on 1D power spectrum measurement, we proceed to cross-power spectra.

A. Mock dataset for $P_\times(z, \Delta\theta, k_{\parallel})$ measurements

We will present the measurement on 40 realizations of the BOSS DR11 mocks, that have been widely used in past BOSS Lyα BAO analyses [11, 12]. These mocks were generated for the eleventh data release (DR11) of BOSS, for a total of 134 386 quasars. Since BOSS did not generate mock spectra for the final data release (DR12), and since the difference in area is roughly 10%, we will use these mocks as realistic simulations of the BOSS survey.

The algorithm to generate the mock Lyα forest skewers is described in detail in [31]. In short, a correlated and anisotropic Gaussian field is computed along all lines of sight, with an input power spectrum such that after applying a lognormal transformation to the Gaussian field to obtain an optical depth, the resulting field has the desired Lyα forest power spectrum from [25]. The important thing to note is that while these mocks have been extensively used in the past BOSS paper, they are known to contain imperfections, especially on small scales. Some of these imperfections are related to redshift interpolation applied when making these mocks and would have been missed for analysis that use a single bin across the survey.

B. Band power measurements

We will use the same grid of redshift ($N_z = 7$) and line of sight wavenumbers ($N_k = 21$) bins that were used in measuring $P_{1D}(z, k_{\parallel})$ in section V, and we will use a total of $N_\theta = 27$ angular separation bins: the first bin will be at zero separation (containing 1D information), followed by five bins linearly spaced by an angular separation of 0.0002 radians, or roughly 1.2 transverse Mpc; the last 21 bins are distributed logarithmically with a fractional increase of 0.2, with a last bin centered at 0.0606 radians, or roughly 360 Mpc.

In Figure 6 we show the cross-power spectrum $P_\times(z, \Delta\theta, k_{\parallel})$ measured from 40 mock realizations of the BOSS survey. The measured power is as usual plotted at the maximum likelihood point implied by our quadratic expansion of $\ln L$, i.e., the result of a single NR step away from $\mathbf{p}_0$ (Equation 5). The errorbars are the square root of the diagonal elements of the inverse of the Fisher matrix. We show a representative sample of angular separations, and the four redshift bins that are better measured in the analysis. The solid lines show the expected measurement given the input theory that went into generating the mock data, described in more detail in section VII B. While we plot results for diagnostic purposes, we do not intend $P_\times$ to be an end-stage result of our algorithm – it is an intermediate data compression step in the process of producing P_{3D}.
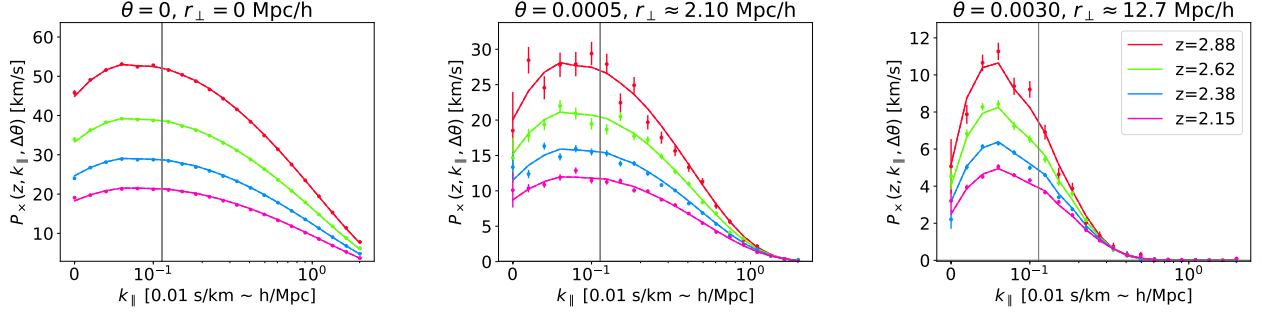
FIG. 6. $P_\times(z, \Delta\theta, k_\parallel)$ measured from 40 mock realizations of the BOSS survey, as a function of wavenumber $k_\parallel$, with different colors identifying different redshift bins, and each panel showing a different angular separation bin. We only show the four redshift bins that are better measured, and a representative sample of angular separations. The wavenumbers in the x-axis have been multiplied by 100 so they can be approximately compared to wavenumbers in units of $h\mathrm{Mpc}^{-1}$. In the title of each plot we have approximated the comoving transverse separation as $r_\perp \approx 4200\ \theta\ h^{-1}\mathrm{Mpc}$. The vertical line divides the linearly spaced bins (to the left) from the logarithmically spaced bins (to the right).

VII.   FROM $P_\times(z, \Delta\theta, k_\parallel)$ TO $P_{3D}(z, k_\perp, k_\parallel)$

In the previous sections we have presented an algorithm to obtain a second order Taylor expansion of the log-likelihood function around a fiducial model spectrum, describing the cross-spectrum of the Lyα forest $P_\times(z, \Delta\theta, k_\parallel)$.

For a sufficiently fine binning, this presents a complete summary statistics describing 2-pt correlations, and we could decide to stop here. We could compute theoretical predictions for the cross-spectrum, and do parameter estimation (BAO scale, linear power...) directly in this space. However, since the models are usually defined in Fourier space, it is useful to translate the measurement into a truly three-dimensional power spectrum $P_{3D}(z, k_\perp, k_\parallel)$. In principle we could convert the $P_\times$ measurement into any other parameterization, including, for example, the correlation function.

A.   Determining $P_{3D}(z, k_\perp, k_\parallel)$

We would like to compute the Taylor expansion of the (log-) likelihood with respect to a new set of parameters p_γ^{3D}. Using the derivative chain rule we can compute the new derivatives $(\mathcal{L}_{,\gamma}^{(3D)}, \mathcal{L}_{,\gamma\delta}^{(3D)})$ as a function of the old ones $(\mathcal{L}_{,\alpha}^{(\times)}, \mathcal{L}_{,\alpha\beta}^{(\times)})$:

$$\mathcal{L}_{,\gamma}^{(3D)} = \mathcal{L}_{,\alpha}^{(\times)}\ \frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}} + \mathcal{L}_{,\alpha\beta}^{(\times)}\ \frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}}\ \Delta P_\times(z_\beta, \Delta\theta_\beta, k_{\parallel_\beta}) \tag{36}$$

$$\mathcal{L}_{,\gamma\delta}^{(3D)} = \mathcal{L}_{,\alpha\beta}^{(\times)}\ \frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}}\ \frac{\partial p_\beta^\times}{\partial p_\delta^{3D}}\ , \tag{37}$$

where

$$\Delta P_\times(z, \Delta\theta, k_\parallel) = \int \frac{d\mathbf{k}_\perp}{(2\pi)^2}\ e^{i\mathbf{\Delta\theta}\mathbf{k}_\perp}\ P_{3D}^{\mathrm{fid}}(z, k_\perp, k_\parallel) - P_\times^{\mathrm{fid}}(z, \Delta\theta, k_\parallel)\ . \tag{38}$$

The second term in the first derivative takes into account that the Taylor expansions generally are not computed around the same fiducial spectrum. The reason for this complication is that for 3D power spectrum we want to expand around the true power, while for the cross-spectrum we are forced to set cross-power at non-zero separations to zero for numerical expediency (as discussed in Section IV). The second derivative is not affected since its value is constant when assuming a Gaussian likelihood, but would have received similar contribution had we expanded the initial likelihood to third order. Since the two descriptions of power spectra are linearly related, terms involving

second derivatives $\partial^2 p_\alpha^\times / \partial p_\gamma^{3D} \partial p_\delta^{3D}$ vanish. The definition of $\frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}}$ is clear given Eq. (19):

$$\frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}} = \sum_\beta \left(I^{-1}\right)_{\alpha\beta} \int dz \, d\mathbf{\Delta\theta} \, \frac{dk_\parallel}{2\pi} \, w_\beta(z, \Delta\theta, k_\parallel) \int \frac{d\mathbf{k}_\perp}{(2\pi)^2} \, e^{i\mathbf{\Delta\theta}_\alpha \mathbf{k}_\perp} w_\gamma(z, \mathbf{k}_\perp, k_\parallel) \tag{39}$$

$$\simeq \int \frac{d\mathbf{k}_\perp}{(2\pi)^2} \, e^{i\mathbf{\Delta\theta}_\alpha \mathbf{k}_\perp} w_\gamma(z_\alpha, \mathbf{k}_\perp, k_\parallel^\alpha)$$

where $I_{\alpha\beta} = \int dz \, d\mathbf{\Delta\theta} \, \frac{dk_\parallel}{2\pi} w_\alpha(z, \Delta\theta, k_\parallel) w_\beta(z, \Delta\theta, k_\parallel)$. In Appendix B we discuss in more detail these derivatives, and give an explicit window function linking 3D power spectrum and 3D power spectrum parameters.

We will use the same grid of $(z, k_\parallel)$ parameters that was used to measure $P_\times(z, \Delta\theta, k_\parallel)$, i.e., a total of $N_z = 7$ and $N_{k_\parallel} = 21$ bins. We will define the following $k_\perp$ grid: the first bin will be centered at $k_\perp = 0$, and it will be followed by 5 linear spaced bins of width $\Delta k_\perp = 60$ (in units of inverse radians), followed by a set of 15 bins logarithmically spaced $\Delta k_\perp / k_\perp = 0.2$, for a total of $N_{k_\perp} = 21$ bins.
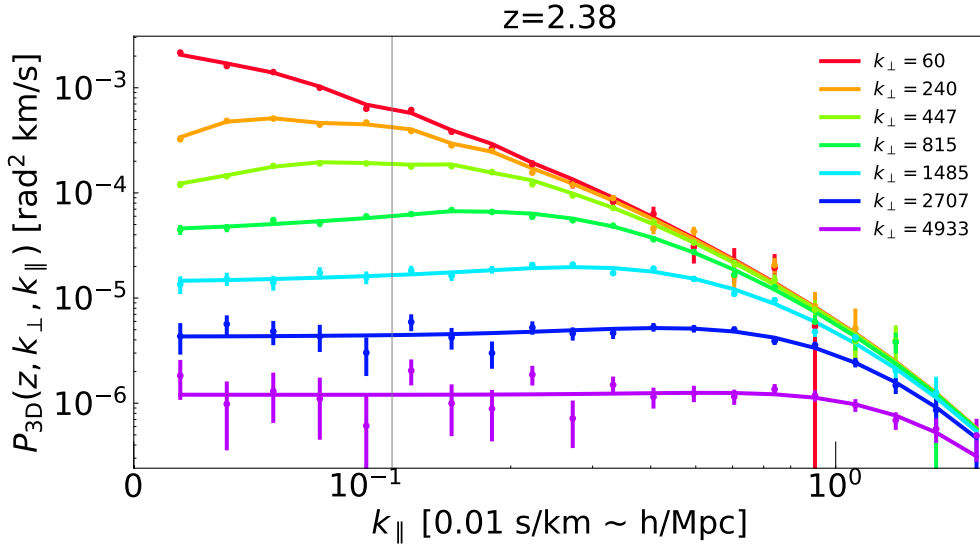

FIG. 7. $P_{3D}(z, k_\perp, k_\parallel)$ measured as a function of line of sight wavenumber $k_\parallel$, for several different transverse wavenumbers $k_\perp$, at $z = 2.38$. Measurement from 40 synthetic realizations of the BOSS survey. The solid lines show the input theory used to generate the mocks. The radial wavenumbers in the x-axis have been multiplied by 100 so they can be approximately compared to wavenumbers in units of $h\mathrm{Mpc}^{-1}$. The vertical line divides the linearly spaced bins (to the left) from the logarithmically spaced bins (to the right).

After computing the Taylor expansion of the log-likelihood with band powers $\mathbf{p}^{3D}$ we can visualize the results by computing the implied maximum of the resulting Gaussian likelihood approximation (i.e., the result of a single Newton-Raphson step away from $\mathbf{p}_0$). Figure 7 shows some of the estimated band power values, and the true power in the mocks. Note that in figure 7 we have marginalized the likelihood over band powers corresponding to bins with $k_\parallel = 0$ or $k_\perp = 0$, following the recipe described in Appendix A. We expect these modes to be very difficult to measure in a real data analysis, since $k_\parallel = 0$ modes are very sensitive to distortions caused in continuum fitting and $k_\perp = 0$ are very sensitive to systematics that vary slowly in the sky.

In the bottom panels of Figure 8 we measure our success quantitatively by computing χ^2 for the estimated power relative to our expectation for these mocks.
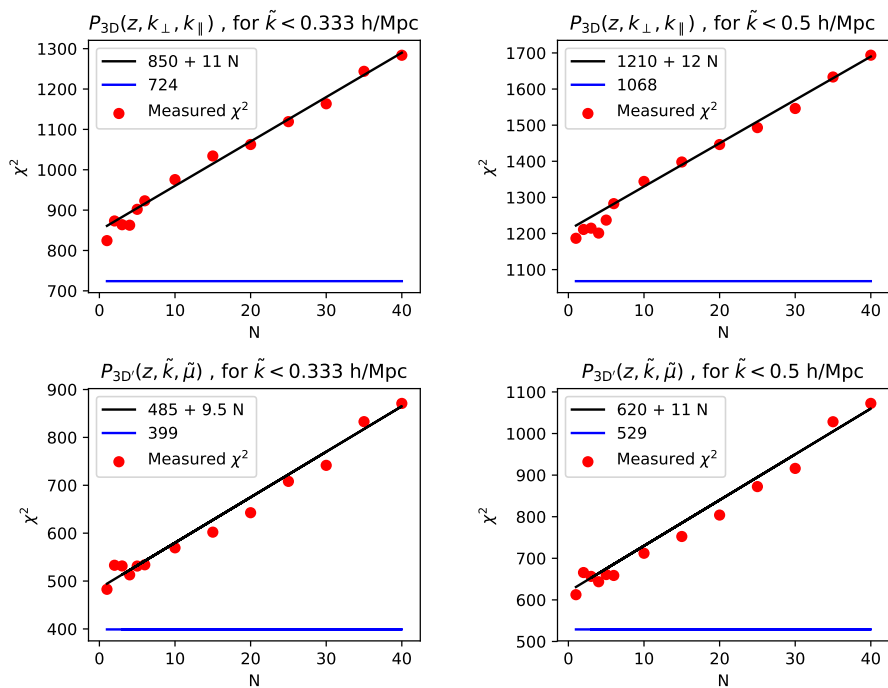
FIG. 8. Value of χ^2 as a function of number of realizations combined, N, when comparing the measured band powers to the input model, after marginalizing over bands with either $k_\perp = 0$ or $k_\parallel = 0$. The top panels show the results for $P_{3D}(z, k_\perp, k_\parallel)$, while the bottom panels show the results for $P_{3D'}(z, \tilde{k}, \tilde{\mu})$. In both cases we also marginalize over modes with $\tilde{k} > 0.00333\text{s km}^{-1} \approx 0.333\ h\text{Mpc}^{-1}$ (left) or $\tilde{k} > 0.005\text{s km}^{-1} \approx 0.5\ h\text{Mpc}^{-1}$ (right). The blue line shows the number of degrees of freedom, and the black line is an approximate fit to the linear dependence of χ^2 with the number of realizations (see text).

We see that χ^2 per degree of freedom is near 1 but slightly higher. The value increases with the number of realizations averaged together, suggesting some systematic errors that are becoming larger relative to the statistical errors. The $N = 0$ intercept should give an indication of how much we have underestimated our statistical errors, i.e., $\chi^2/\nu = 1.13 - 1.17$ indicating 6-8% underestimation of rms errors (depending on the maximum k used). In general the underestimation could come from neglected non-Gaussianity in the field, expanding around zero cross-correlation, or errors in the various other approximations in our algorithm. Qualitatively, these results seem pretty good, but the more important question is how these errors propagate into measurement of more relevant parameters (like the BAO scale or the amplitude of the linear power spectrum).

It is not always trivial to compute a value of χ^2 from the Taylor expansion of the likelihood, since often the Fisher matrix is not positive definite. For instance, this is clearly the case whenever we have more p_γ^{3D} than $p_\alpha^\times$ parameters.[15] In general, in order to prevent poorly measured modes to contribute to the value of χ^2, we compute an eigenvalue decomposition of the Fisher matrix and consider only the contribution from those eigenmodes with an eigenvalue that is at least 10^{-10} times the largest eigenvalue.[16] This explains why the number of degrees of freedom in this section are usually smaller than the number of band powers measured.

In the bottom panels of Figure 8 we show a similar χ^2 comparison after converting the measured power into $(\tilde{k}, \tilde{\mu})$ bands, as described in Section VII C below. We find a similar conclusion that rms errors are probably underestimated by 8-10%.

---

[15] To be clear, a singular Fisher matrix is not a problem for our main science goal of propagating the information in the data into final cosmological parameter constraints. In fact, if not for computational limitations we would ideally use an essentially infinite number of parameters, i.e., a continuous parameterization of our function. The parameters would not be individually constrained in this limit, but would provide the best possible representation of the information we have.

[16] We have tested that the exact cut used does not qualitatively affect the main results of this study.

B. Fitting bias parameters on measured $P_{3D}(z, k_\perp, k_\parallel)$

The model describing the Lyα forest correlations in the mocks is better described as a function of comoving wavenumber **q**, related to our observed wavenumbers by:

$$q_\parallel = \frac{H(z)}{1+z}\, k_\parallel\,, \qquad q_\perp = \frac{1}{d_A(z)(1+z)}\, k_\perp\,, \tag{40}$$

where $H(z)$ is the Hubble parameter at the redshift where the power spectrum is measured, and $d_A(z)$ is the angular diameter distance to that redshift.

Using these coordinates, the power measured in the mocks can be modeled by:

$$P_{3D}(z, \mathbf{q}) = b^2(z)\,\left[1 + \beta(z)\mu_q^2\right]^2\, P_L(z, q)\, D_{NL}(q, \mu_q)\,, \tag{41}$$

where $b(z)$ is the linear density bias, β is the linear redshift space distortion parameter, q is the modulus of the comoving wavenumber, μ_q is the cosine of its angle with respect to the line of sight, $P_L(z, q)$ is the linear power spectrum and $D_{NL}(q, \mu_q)$ is an analytical expression to account for non-linearities in the flux power spectrum derived in [25]. The mocks where generated with a constant value of $\beta = 1.4$, and with an evolving bias of $b = b_0\,[(1+z)/(1+z_0)]^{2.9}$, with $z_0 = 2.25$ and $b_0 = -0.14$.

If we were to fit these values from a measurement on real data (where the true underlying model is unknown) we would need to fit at the same time the relation between observed and comoving coordinates, i.e., $H(z)$ and $d_A(z)$. However, in this section we will fit b_0 and β on the measured likelihood expansion while keeping fixed the rest of the model, and using the conversion factors from the cosmological model used in generating the mocks.

For each pair of values for (b_0, β) we compute the prediction for all band power parameters p_{3D} by simply evaluating the model power at the central **k** and z for the parameter (if we wanted to be more careful we could use the functional inversion formula, Eq. 19, again to define the conversion). We find the best fit parameters that maximize the likelihood (the standard procedure for this final small number of non-linear parameters, rather than continuing to use the chain rule which assumes the resulting likelihood is sufficiently Gaussian). We get uncertainties in these parameters by looking at the second derivative of the log-likelihood evaluated at the best fit parameters. In table III we show the best fit values of b_0 and β for two different analyses: including $P_{1D}(z, k_\parallel)$ information, and marginalizing over it.

Include P_{1D}	$100\,\tilde{k}_{\max}[\text{s km}^{-1}]$	$b_0\ (=-0.14)$	$\beta\ (=1.4)$	χ^2 (d.o.f.)
Y	2.00	-0.1392 ± 0.00004	1.412 ± 0.0011	6361.3 (2473)
Y	1.00	-0.1398 ± 0.00018	1.404 ± 0.0039	2497.0 (1800)
Y	0.50	-0.1398 ± 0.00020	1.406 ± 0.0043	1693.8 (1068)
N	2.00	-0.1398 ± 0.00019	1.403 ± 0.0040	3604.8 (2631)
N	1.00	-0.1399 ± 0.00019	1.402 ± 0.0041	2670.4 (1943)
N	0.50	-0.1398 ± 0.00020	1.406 ± 0.0044	1820.0 (1203)
Y	2.00	-0.1392 ± 0.00004	1.411 ± 0.0011	4817.3 (959)
Y	1.00	-0.1398 ± 0.00018	1.404 ± 0.0039	1416.1 (761)
Y	0.50	-0.1397 ± 0.00021	1.408 ± 0.0046	1072.5 (529)
N	2.00	-0.1397 ± 0.00020	1.405 ± 0.0043	1994.5 (1048)
N	1.00	-0.1398 ± 0.00020	1.404 ± 0.0044	1538.3 (846)
N	0.50	-0.1397 ± 0.00022	1.408 ± 0.0048	1177.4 (603)

TABLE III. Best fit values of bias (b_0) and redshift space distortion parameter (β) from the measured $P_{3D}(z, k_\perp, k_\parallel)$ (top six rows) and $P_{3D'}(z, \tilde{k}, \tilde{\mu})$ (bottom six rows), combining 40 realizations of the BOSS survey. Some rows included $P_{1D}(z, k_\parallel)$ information, while other do not. The second column specifies the maximum wavenumber included in the fits, in units of s km^{-1} multiplied by 100 so that the numbers can be approximately compared to hMpc^{-1}. The last column shows the value χ^2 with respect to the expected theory in the mocks (not the fit). Note that the value of χ^2 in the third (ninth) row correspond to the last point in the upper (lower) right panel of Figure 8.

In all cases, we marginalize over $k_\parallel = 0$ and $k_\perp = 0$. For each type of analysis we show the results as a function of maximum wavenumber used in the fits $\tilde{k}_{\max}$, where $\tilde{k}$ is defined by Eq. 14, with $f(z) = H(z)d_A(z)$ for the fiducial model, i.e., even though our bands are labeled by $(k_\parallel,\ k_\perp)$ here, we make the cutoff uniform in magnitude of $\tilde{k}$. The bottom line is that, even for statistical errors corresponding to ~ 40 times BOSS, there is no detectable bias in our measurement of the bias parameters (except in the unrealistic case where we fit to high k using 1D power). Apparently the systematic errors that produce poor χ^2 in this limit have a peculiar form that does not project onto these parameters (it is hard to identify these errors by eye given so many precisely measured points). For this large

number of mocks, the $\sim 8\%$ underestimate of the statistical errors makes a sub-dominant contribution to the excess χ^2.

We can see in table III that when the $P_{1D}(z, k_\parallel)$ information is not included, the uncertainties in the bias parameters are fairly independent of our choice of $\tilde{k}_{\mathrm{max}}$. This means that in a BOSS-like survey, the cross-correlation between different lines of sight mostly constrain large scales modes with $\tilde{k} \lesssim 0.005\mathrm{s\ km}^{-1}$ ($q \lesssim 0.5\ h\mathrm{Mpc}^{-1}$), although radial modes should be well-measured to higher k than transverse modes. This is analogous to the way that point object clustering measurements become shot-noise limited at small scales. On the other hand, when we include $P_{1D}(z, k_\parallel)$ the uncertainties are driven by $\tilde{k}_{\mathrm{max}}$, since we are able to measure correlations down to very small scales. Similarly, by comparing the results using $\tilde{k}_{\mathrm{max}} = 0.005\mathrm{s\ km}^{-1}$ we can see that the contribution of $P_{1D}(z, k_\parallel)$ is quite small, the measurement is dominated by cross-correlations from different lines of sight.

Note that when analyzing real data we would not be able to fit the biasing model down to these very small scales, since deviations from linear theory are already important at $q \approx 0.2\ h\mathrm{Mpc}^{-1}$ or so (see Figure 2 of [26]). However, the mocks we are using are well described by the model presented in equation 41 down to a fairly small scales, as shown by correlation function analyses by the BOSS collaboration ([11, 12, 32]).

C. $P_{3\mathrm{D}'}(z, \tilde{k}, \tilde{\mu})$

Another example of likelihood conversion is to change to a set of parameters labeled by transformed coordinates $\tilde{k}$ and $\tilde{\mu}$ as defined in Equations 14 and 15. Again, we use the chain rule to convert between different representations of the power spectrum. It is important to understand that this is really just a different interpolation scheme to describe the same measured power spectrum that is fundamentally a function of $(k_\parallel, k_\perp)$. E.g., every band in $P_{3\mathrm{D}'}(z, \tilde{k}, \tilde{\mu})$ can still be labeled by well-defined $(k_\parallel, k_\perp)$, it is just that we assume interpolation is linear in $\tilde{k}$ and $\tilde{\mu}$ instead of $k_\parallel$ and $k_\perp$. With $f(z) \equiv H(z) d_A(z)$ for a fiducial cosmology, $\tilde{\mu}$ is approximately a proper physical angle, and $\tilde{k}$ is approximately the magnitude of a proper physical wavevector.[17] Note that in general we could go straight from $P^\times$ to $P_{3\mathrm{D}'}(z, \tilde{k}, \tilde{\mu})$, but going through $P_{3\mathrm{D}}(z, k_\parallel, k_\perp)$ allows things like marginalization over $k_\perp \sim 0$ which we expect may be corrupted by systematic errors. We compute $P_{3\mathrm{D}}(z, k_\parallel, k_\perp)$ parameters given $P_{3\mathrm{D}'}(z, \tilde{k}, \tilde{\mu})$ by simply evaluating at the central values of $(k_\parallel, k_\perp)$, although we could use another copy of the functional inversion Equation, (19), if desired.

In Figure 9 we show the measured $P_{3D'}(z, \tilde{k}, \tilde{\mu})$, as a function of angular direction $\tilde{\mu}$. The plot shows the measurement from the combination of 40 realizations of the BOSS survey. Whenever we change the set of parameters we can decide to use a different set of redshift bins, and in order to compress all the information in a single plot in this section we plot $P_{3D'}(z, \tilde{k}, \tilde{\mu})$ measured at a single redshift bin.

In section VII B we estimated the best fit values of the bias parameters (b_0, β) from measurements of $P_{3D}(z, k_\perp, k_\parallel)$. We can similarly fit for the bias parameters from measurements of $P_{3D'}(z, \tilde{k}, \tilde{\mu})$, as shown in the second part of Table III. We can see that the estimates of the bias parameters, and their uncertainties, are very similar to those obtained directly from $P_{3D}(z, k_\perp, k_\parallel)$.

---

[17] The measured power is still technically defined by the Fourier transform of the correlation function in observable coordinates, as in Eq. 9, so simply multiplying $\tilde{k}$ by $H(z)$ will not give exactly the same result one would have obtained by doing the transforms in comoving coordinates from the start, but this distinction is small.
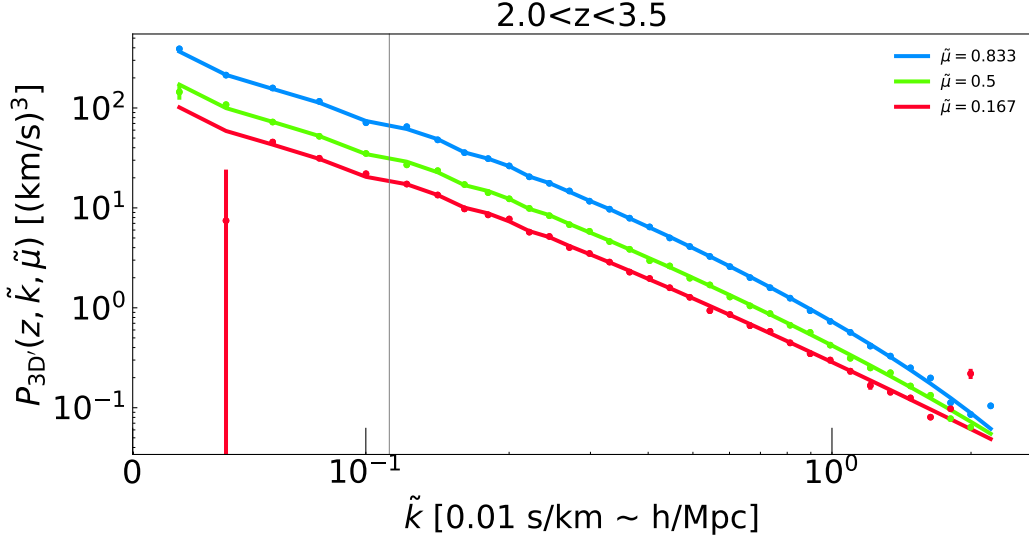
FIG. 9. Measured $P_{3D'}(z, \tilde{k}, \tilde{\mu})$ as a function of wavenumber $\tilde{k}$, for different angular directions $\tilde{\mu}$. Measurement from 40 synthetic realizations of the BOSS survey. The solid lines show the input theory used to generate the mocks. The wavenumbers in the x-axis have been multiplied by 100 so they can be approximately compared to wavenumbers in units of $h\mathrm{Mpc}^{-1}$. The vertical line divides the linearly spaced bins (to the left) from the logarithmically spaced bins (to the right). All the information has been compressed into a single redshift bin, and we have used the model described in equation 41 as our fiducial power. These mocks were designed to have the correct large scale clustering for BAO analysis of the Lyα forest, and it is not surprising that there are disagreements on the smallest scales where they had not been tested before.

In order to highlight the baryonic features on large scales, in Figure 10 we plot both the measured and the expected power spectra divided by a smooth version of the power without the oscillations.
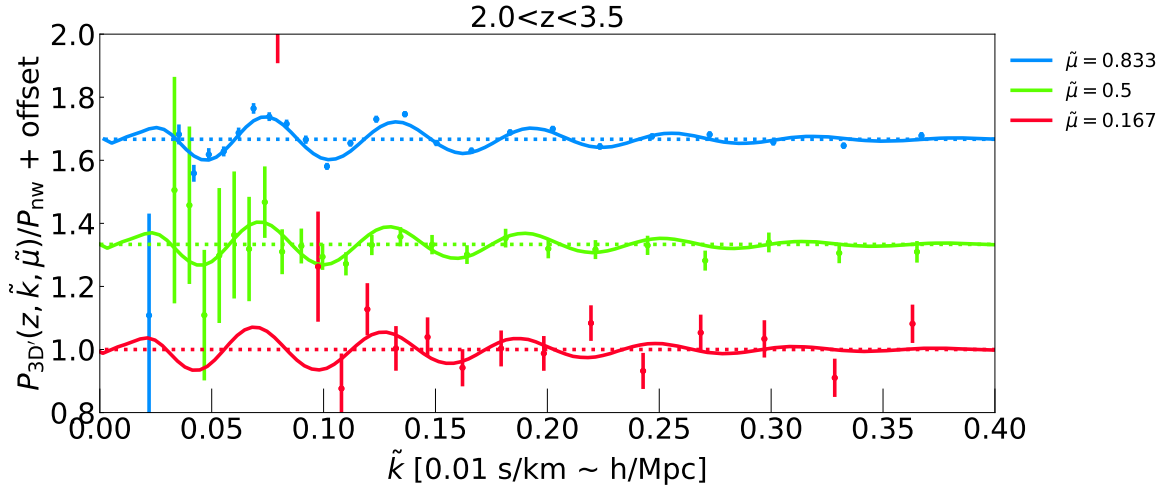

FIG. 10. Same $P_{3D'}(z, \tilde{k}, \tilde{\mu})$ as figure 9, but divided by a smoothed version of the power without the oscillations, and zooming on larger scales relevant for BAO. Note that an arbitrary offset has been added for better visualization. Measurement on 40 mock realizations of the BOSS survey. In order to resolve the oscillations this analysis has used a finer binning in $k_{\parallel}$ and $k_{\perp}$ and it has ignored some of the high-k range covered in figure 9.

In a BAO analysis it is important to resolve the oscillatory features, and for this reason we have used a finer binning in $k_{\parallel}$ and $k_{\perp}$ in the analysis presented in Figure 10. We see, as expected, that the BAO oscillations are well-measured in the radial direction, and much less well measured in the transverse direction. At this high precision we do see what

looks like some small systematic deviations from the curve. Presumably by the time this level of deviation matters we can tighten up our numerics enough to eliminate it.

VIII.   DISCUSSION

In this paper we have presented a new method to estimate the 3D power spectrum of the Lyα forest power spectrum. The central tenet of our method is Taylor series expansion of the likelihood around a fiducial model. We use cross-spectrum as a natural intermediate product given the Lyα forest data window function. The likelihood for the cross-spectrum can be converted into likelihood for any statistics of interest by applying chain rule to the derivatives. We have demonstrated the applicability of this method on an idealized set of mock data of realistic size and have justified the use of several optimization tricks.

Finally, we discuss a few qualitative issues.

A.   Coordinate system

We have defined our power spectrum by Eq. (9) to be the Fourier transform of the correlation function in observable coordinates, i.e., angle and redshift (ln λ), and always label our band parameters in these units. In contrast, analyses of redshift survey data usually use a fiducial cosmology to translate the observed coordinates into approximate comoving coordinates. Our reason for using observable coordinates is, at the core, that it is necessary given the algorithm that we want to use for the massive data compression step (pixels $\rightarrow P_\times$). We want to make an FFT or similar transform along the individual spectra, after which we simply cannot use transverse comoving separations at the pixel level – we are no longer using pixels. In contrast, if one is primarily interested in measuring statistics by summing over pairs of galaxies or pixels, comoving separation can be computed pair-by-pair. Our reason for wanting a radial transform is the usual one: Fourier modes at least approximately diagonalize the problem, naively reducing the computation time by a factor of $\sim N_{\text{pixels}}$ or more (we will discuss later why our analysis of a similar set of data does not in fact run faster than well-developed correlation function measurements).

The important question for us is: is there anything fatally wrong with using these coordinates? The standard objection would be, e.g., that if you measure a single correlation function by counting pairs in angular bins over a wide range in redshift, a given bin in angle will correspond to changing comoving separation so a feature like BAO will be smeared. As long as the fiducial model is sufficiently close to correct you can count pairs in comoving bins over an arbitrarily wide range in redshift without blurring any features that are fixed in comoving coordinates, as linear regime structures like BAO are. Our analysis, however, does not naively smear structure in this way. It is built from the beginning around well-defined interpolation in the coordinates, including measuring the power centered at an arbitrarily spaced set of points in redshift. We do not lose information by chopping the redshift range into chunks to be analyzed independently, and we automatically compute the covariance matrix between parameters at different redshifts. For sufficiently finely spaced points, the interpolation can resolve any structure, and we can always check that it is sufficient against mocks, and check for convergence in the real data analysis. With the linear interpolation we use in redshift, our error is smaller than a simple estimate of smearing based on, e.g., the change in angular position of a fixed comoving feature across a bin (e.g., 5.4% between typical band centers $z = 2.15$ and $z = 2.38$). The key thing to understand is: with well-designed interpolation between parameters labeled by any kind of redshift and radial and transverse separation coordinates, the potential value of comoving coordinates really boils down to providing a maximally smooth basis for interpolation.

There are some other reasons we might be happy to remove the layer of complication that comes with working in fiducial comoving coordinates: While the large-scale (linear regime) redshift evolution can be captured in comoving coordinates by the evolution of two numbers, $b_\delta(z)\sigma_8(z)$ and $f(z)b_\eta(z)\sigma_8(z)$,[18] the redshift evolution of clustering on smaller (non-linear) scales is more complicated, and it is generally desirable to use relatively finely spaced redshift bins in any case to resolve this evolution. Also, using a wavelength to define a redshift to define comoving coordinates requires that you specify a transition wavelength, e.g., Lyα. A small but non-negligible fraction of the absorption is caused by higher order hydrogen lines or metal lines (predominantly C-IV, Si-II, Si-III, Si-IV, O-VI, MgII. . .). The relative position of this absorption is fixed in ln λ, and "comoving positions" based on the Lyα wavelength are simply wrong. Generally, observation-related systematics live naturally in angle and wavelength space and know nothing about comoving coordinates. General-use spectra typically come with pixels evenly spaced in ln λ so we can FFT in these units without any re-pixelization.

---

[18] In galaxy clustering analyses one can use $b_\eta = 1$, but this is not the case in analyses of the Lyα forest [25, 33, 34].

Note that the key place where we need observable coordinates is the pixels$\rightarrow P_\times$ compression step. After that we could easily use the chain rule to convert to P_{3D} in fiducial comoving coordinates (we effectively do this with $\tilde{\mu}$). The potential damage of distorting features by insufficiently accurate interpolation will already be done, however, so that is no longer a motivation. While it might seem easier to use results labeled in fiducial comoving units, suggesting that if we published results from data we might want to at least shift at the end to these units, this is definitely wrong, i.e., it is not easier. Galaxy measurements really should be converted back to observable units in the end, even if the initial massive data compression is done in fiducial comoving units. Given power in observable units someone who wants to use it for broader cosmological model fitting just needs to make predictions for it in whatever model they want to test. Given power in fiducial comoving coordinates they need to figure out what fiducial model was used and convert back to observable coordinates before comparing (or, equivalently, apply ratios of distances model-by-model to convert from one set of comoving coordinates to another). The point is not that this is very hard, it's just that there is no ease-of-use argument for fiducial comoving coordinates. In fact, it is straightforward to label bands in a table with both types of coordinates, so that, e.g., someone who wants to make a quick plot with comoving labels can do that.

B. Optimality

If we can compute $L(\mathbf{d}|\mathbf{p})$ exactly, we clearly have all the information we can have and our analysis is therefore optimal. Sub-optimality is a less well-defined concept in a Bayesian context compared to frequentist analysis. In the latter we deal with estimators, i.e. procedures that given data produce an estimate of the desired quantity in which we can trade complexity of the estimator computation with the output variance. An estimator can be suboptimal but still unbiased. In the Bayesian analysis however, likelihood is the basic concept and there is not such thing as "unbiased" likelihood. When we Taylor expand around $\mathbf{p}_0$ to estimate the likelihood, we are making an error in throwing out terms beyond quadratic order. This error will generally be larger in vicinity of the true parameters $\mathbf{p}_t$ (where we would like to use it) if $\mathbf{p}_0$ is farther away from the truth. As we saw in Section IV we had to expand $P^\times$ not around the model we would have liked to expand around, but instead around a model for which we can evaluate $\mathbf{C}_0^{-1}$ efficiently. In terms of the OQE picture above one could conclude that our estimate is somewhat sub-optimal and our error-matrix somewhat wrong. If we want to be more accurate, one way is to try to expand around $\mathbf{p}_0$ closer to $\mathbf{p}_t$ (i.e., use a better $\mathbf{C}_0^{-1}$ for weighting), but another way is to go to the next order in the Taylor series. While this is rarely if ever done, it is a very realistic possibility given our algorithm, because, once we have computed $\mathbf{M}_{I\alpha\beta}$ for each spectrum, the trace calculations that go into computing the Fisher matrix or higher derivatives are not very costly. The third order term could be used directly in subsequent uses of the results, or could be used to shift to a traditional Gaussian expansion around a more desirable model $\mathbf{p}_1 \sim \mathbf{p}_t$, i.e.,

$$\mathcal{L}_{,\alpha}(\mathbf{p}_1) = \mathcal{L}_{,\alpha}(\mathbf{p}_0) + \mathcal{L}_{,\alpha\beta}(\mathbf{p}_0)(\mathbf{p}_1 - \mathbf{p}_0)_\beta + \frac{1}{2}\mathcal{L}_{,\alpha\beta\gamma}(\mathbf{p}_0)(\mathbf{p}_1 - \mathbf{p}_0)_\beta(\mathbf{p}_1 - \mathbf{p}_0)_\gamma + \dots \tag{42}$$

$$\mathcal{L}_{,\alpha\beta}(\mathbf{p}_1) = \mathcal{L}_{,\alpha\beta}(\mathbf{p}_0) + \mathcal{L}_{,\alpha\beta\gamma}(\mathbf{p}_0)(\mathbf{p}_1 - \mathbf{p}_0)_\gamma + \dots, \tag{43}$$

In this paper, we do not go beyond second order. Our χ^2 values on mocks suggest $\sim 8\%$ underestimation of errors, suggesting there is not a lot of optimality to be gained by going to higher order. It would be nice to pick up this $\sim 8\%$ if it comes from correlations between close quasar pairs that we are not including in $\mathbf{C}_0^{-1}$ [9], but in any case we will inevitably want some kind of bootstrap-like cross-check of our errors, and this should be good enough, combined with an estimate from mocks, to convincingly calibrate a small error in the errors.

C. Speed

In spite of being massively faster than evaluating the OQE equations in pixel space, our algorithm still takes longer to run in practice than the well-developed pairwise pixel methods used to measure the Lyα correlation function in BOSS [9–12]. Clearly they are not evaluating the OQE equations – they make their own set of approximations to make their algorithm faster. Like us, they do not compute to wide transverse separations, and do not use off-diagonal $\mathbf{C}_0$ at all in their estimator, ignoring correlations even within pixels in the same spectra. Moreover, most of these analyses used rebinned pixels to speed up the code, and measured the correlations in a single redshift bin. We have a lot of overhead related to using a large number of parameters to resolve each coordinate. It is possible that this could be further reduced. A clear optimization which we plan as a next step is to have the number and positions of radial (and redshift?) bins adapt to separation so that the full shape of $P^\times$ is resolved where the information lies and not over-resolved elsewhere. The slow part of our method is the calculation of the second derivative or equivalently the error matrix. Roughly half of the cost comes from pre-computing the $\mathbf{M}_{I\alpha\beta}$ matrices and half from then computing the

cross-terms between quasars. If we extended to large angular separations, the latter would dominate. The equivalent part of the standard correlation function analysis would be the full evaluation of the 4-point function required to determine the error-matrix. Recent standard analysis avoid this by using errors determined by data-internal methods like bootstrap, sometimes comparing to direct calculations sped up by Monte Carlo subsampling of pair contributions to compute. We could also do this kind of subsampling.

D.   Application to fully realistic data

In this paper we have ignored a vast array of systematic effects that need to be dealt with in any real analysis. These include coaddition, non-Gaussian resolution matrix, per spectral pixel noise estimates, sky residuals, continuum fitting, metal contamination, etc. Our analysis is designed, however, to make it easier to deal with some of them compared to standard analysis. We can add marginalization over any systematic template data vector by projecting it out during weighting. For instance, by adding a large constant in our quasar covariances we effectively project out the distortion caused by fitting the continuum amplitude of each quasar, common in most measurements of the correlation function [9, 12, 35]. The marginalization increases the uncertainty in the low $k_{\parallel}$ power spectrum, but removes any bias. We have tested that we get identical results when we substitute $\delta \to \delta - \langle \delta \rangle$, where $\langle \delta \rangle$ is the averaged fluctuation in each line of sight.

Most metal contamination can be dealt with by subtracting the corresponding power measured on the red-side of the Lyα emission line. Finally, our formalism naturally allows for a unified analysis of 1D and 3D power spectra into a single master likelihood.

ACKNOWLEDGMENTS

[1] J. R. Bond, A. H. Jaffe, and L. Knox, *Estimating the power spectrum of the cosmic microwave background*, Phys. Rev. **D 57** (Feb., 1998) 2117–2137.

[2] U. Seljak, *Cosmography and Power Spectrum Estimation: A Unified Approach*, Astrophys. J. **503** (Aug., 1998) 492–+.

[3] J. R. Bond, A. H. Jaffe, and L. Knox, *Radical Compression of Cosmic Microwave Background Data*, Astrophys. J. **533** (Apr., 2000) 19–37.

[4] H. A. Feldman, N. Kaiser, and J. A. Peacock, *Power-spectrum analysis of three-dimensional redshift surveys*, Astrophys. J. **426** (May, 1994) 23–37.

[5] 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** (May, 1998) 555–+.

[6] M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton, J. A. Frieman, M. Fukugita, J. E. Gunn, A. J. S. Hamilton, G. R. Knapp, R. C. Nichol, J. P. Ostriker, N. Padmanabhan, W. J. Percival, D. J. Schlegel, D. P. Schneider, R. Scoccimarro, U. Seljak, H.-J. Seo, M. Swanson, A. S. Szalay, M. S. Vogeley, J. Yoo, I. Zehavi, K. Abazajian, S. F. Anderson, J. Annis, N. A. Bahcall, B. Bassett, A. Berlind, J. Brinkmann, T. Budavari, F. Castander, A. Connolly, I. Csabai, M. Doi, D. P. Finkbeiner, B. Gillespie, K. Glazebrook, G. S. Hennessy, D. W. Hogg, Ž. Ivezić, B. Jain, D. Johnston, S. Kent, D. Q. Lamb, B. C. Lee, H. Lin, J. Loveday, R. H. Lupton, J. A. Munn, K. Pan, C. Park, J. Peoples, J. R. Pier, A. Pope, M. Richmond, C. Rockosi, R. Scranton, R. K. Sheth, A. Stebbins, C. Stoughton, I. Szapudi, D. L. Tucker, D. E. V. Berk, B. Yanny, and D. G. York, *Cosmological constraints from the SDSS luminous red galaxies*, Phys. Rev. **D 74** (Dec., 2006) 123507–+, [astro-ph/].

[7] P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, R. Cen, D. Shih, J. Schaye, D. P. Schneider, N. A. Bahcall, J. W. Briggs, J. Brinkmann, R. J. Brunner, M. Fukugita, J. E. Gunn, Ž. Ivezić, S. Kent, R. H. Lupton, and D. E. Vanden Berk, *The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey*, ApJS **163** (Mar., 2006) 80–109, [astro-ph/].

[8] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, M. Viel, É. Aubourg, S. Bailey, J. Bautista, M. Blomqvist, A. Bolton, J. S. Bolton, N. G. Busca, B. Carithers, R. A. C. Croft, K. S. Dawson, T. Delubac, A. Font-Ribera, S. Ho, D. Kirkby, K.-G. Lee, D. Margala, J. Miralda-Escudé, D. Muna, A. D. Myers, P. Noterdaeme, I. Pâris, P. Petitjean, M. M. Pieri, J. Rich, E. Rollinde, N. P. Ross, D. J. Schlegel, D. P. Schneider, A. Slosar, and D. H.

Weinberg, *The one-dimensional Lyα forest power spectrum from BOSS*, <u>Astron. Astrophys.</u> **559** (Nov., 2013) A85, [`arXiv:1306.5896`].

[9] A. Slosar, A. Font-Ribera, M. M. Pieri, J. Rich, J.-M. Le Goff, É. Aubourg, J. Brinkmann, N. Busca, B. Carithers, R. Charlassier, M. Cortês, R. Croft, K. S. Dawson, D. Eisenstein, J.-C. Hamilton, S. Ho, K.-G. Lee, R. Lupton, P. McDonald, B. Medolin, D. Muna, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, Y. Piškur, E. Rollinde, N. P. Ross, D. J. Schlegel, D. P. Schneider, E. Sheldon, B. A. Weaver, D. H. Weinberg, C. Yeche, and D. G. York, *The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data*, <u>JCAP</u> **9** (Sept., 2011) 1–+, [`arXiv:1104.5244`].

[10] N. G. Busca, T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, D. Kirkby, J.-M. Le Goff, M. M. Pieri, A. Slosar, É. Aubourg, J. E. Bautista, D. Bizyaev, M. Blomqvist, A. S. Bolton, J. Bovy, H. Brewington, A. Borde, J. Brinkmann, B. Carithers, R. A. C. Croft, K. S. Dawson, G. Ebelke, D. J. Eisenstein, J.-C. Hamilton, S. Ho, D. W. Hogg, K. Honscheid, K.-G. Lee, B. Lundgren, E. Malanushenko, V. Malanushenko, D. Margala, C. Maraston, K. Mehta, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, P. Noterdaeme, M. D. Olmstead, D. Oravetz, N. Palanque-Delabrouille, K. Pan, I. Pâris, W. J. Percival, P. Petitjean, N. A. Roe, E. Rollinde, N. P. Ross, G. Rossi, D. J. Schlegel, D. P. Schneider, A. Shelden, E. S. Sheldon, A. Simmons, S. Snedden, J. L. Tinker, M. Viel, B. A. Weaver, D. H. Weinberg, M. White, C. Yèche, and D. G. York, *Baryon acoustic oscillations in the Lyα forest of BOSS quasars*, <u>Astron. Astrophys.</u> **552** (Apr., 2013) A96, [`arXiv:1211.2616`].

[11] T. Delubac, J. E. Bautista, N. G. Busca, J. Rich, D. Kirkby, S. Bailey, A. Font-Ribera, A. Slosar, K.-G. Lee, M. M. Pieri, J.-C. Hamilton, É. Aubourg, M. Blomqvist, J. Bovy, J. Brinkmann, W. Carithers, K. S. Dawson, D. J. Eisenstein, S. G. A. Gontcho, J.-P. Kneib, J.-M. Le Goff, D. Margala, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, P. Noterdaeme, R. O'Connell, M. D. Olmstead, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, N. P. Ross, G. Rossi, D. J. Schlegel, D. P. Schneider, D. H. Weinberg, C. Yèche, and D. G. York, *Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars*, <u>Astron. Astrophys.</u> **574** (Feb., 2015) A59, [`arXiv:1404.1801`].

[12] J. E. Bautista, N. G. Busca, J. Guy, J. Rich, M. Blomqvist, H. du Mas des Bourboux, M. M. Pieri, A. Font-Ribera, S. Bailey, T. Delubac, D. Kirkby, J.-M. Le Goff, D. Margala, A. Slosar, J. A. Vazquez, J. R. Brownstein, K. S. Dawson, D. J. Eisenstein, J. Miralda-Escudé, P. Noterdaeme, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, N. P. Ross, D. P. Schneider, D. H. Weinberg, and C. Yèche, *Measurement of baryon acoustic oscillation correlations at z = 2.3 with SDSS DR12 Lyα-Forests*, <u>Astron. Astrophys.</u> **603** (June, 2017) A12, [`arXiv:1702.0017`].

[13] A. Slosar, V. Iršič, D. Kirkby, S. Bailey, N. G. Busca, T. Delubac, J. Rich, É. Aubourg, J. E. Bautista, V. Bhardwaj, M. Blomqvist, A. S. Bolton, J. Bovy, J. Brownstein, B. Carithers, R. A. C. Croft, K. S. Dawson, A. Font-Ribera, J.-M. Le Goff, S. Ho, K. Honscheid, K.-G. Lee, D. Margala, P. McDonald, B. Medolin, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, P. Noterdaeme, N. Palanque-Delabrouille, I. Pâris, P. Petitjean, M. M. Pieri, Y. Piškur, N. A. Roe, N. P. Ross, G. Rossi, D. J. Schlegel, D. P. Schneider, N. Suzuki, E. S. Sheldon, U. Seljak, M. Viel, D. H. Weinberg, and C. Yèche, *Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9*, <u>JCAP</u> **4** (Apr., 2013) 26, [`arXiv:1301.3459`].

[14] M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein, B. Flaugher, K. Honscheid, R. Kron, O. Lahav, P. McDonald, N. Roe, D. Schlegel, and representing the DESI collaboration, *The DESI Experiment, a whitepaper for Snowmass 2013*, <u>ArXiv e-prints</u> (Aug., 2013) [`arXiv:1308.0847`].

[15] L. Hui, A. Stebbins, and S. Burles, *A Geometrical Test of the Cosmological Energy Contents Using the Lyα Forest*, <u>Astrophys. J. Let.</u> **511** (Jan., 1999) L5–L8, [`astro-ph/9807190`].

[16] M. Tegmark, *How to measure CMB power spectra without losing information*, <u>Phys. Rev.</u> **D 55** (May, 1997) 5895–5907.

[17] G. Efstathiou, *A maximum likelihood analysis of the low cosmic microwave background multipoles from the Wilkinson Microwave Anisotropy Probe*, <u>Mon. Not. Roy. Astron. Soc.</u> **348** (Mar., 2004) 885–896.

[18] M. S. Madhavacheril, P. McDonald, N. Sehgal, and A. Slosar, *Building unbiased estimators from non-Gaussian likelihoods with application to shear estimation*, <u>JCAP</u> **1** (Jan., 2015) 022, [`arXiv:1407.1906`].

[19] C. Alcock and B. Paczynski, *An evolution free test for non-zero cosmological constant*, <u>Nature (London)</u> **281** (Oct., 1979) 358–+.

[20] L. Knox, *Cosmic microwave background anisotropy window functions revisited*, <u>Phys. Rev.</u> **D 60** (Nov., 1999) 103516, [`astro-ph/9902046`].

[21] H.-J. Seo, S. Ho, M. White, A. J. Cuesta, A. J. Ross, S. Saito, B. Reid, N. Padmanabhan, W. J. Percival, R. de Putter, D. J. Schlegel, D. J. Eisenstein, X. Xu, D. P. Schneider, R. Skibba, L. Verde, R. C. Nichol, D. Bizyaev, H. Brewington, J. Brinkmann, L. A. Nicolaci da Costa, J. R. Gott, III, E. Malanushenko, V. Malanushenko, D. Oravetz, N. Palanque-Delabrouille, K. Pan, F. Prada, N. P. Ross, A. Simmons, F. de Simoni, A. Shelden, S. Snedden, and I. Zehavi, *Acoustic Scale from the Angular Power Spectra of SDSS-III DR8 Photometric Luminous Galaxies*, <u>Astrophys. J.</u> **761** (Dec., 2012) 13, [`arXiv:1201.2172`].

[22] J. Harnois-Déraps, U.-L. Pen, I. T. Iliev, H. Merz, J. D. Emberson, and V. Desjacques, *High-performance P^3M N-body code: CUBEP^{3}M*, <u>Mon. Not. Roy. Astron. Soc.</u> **436** (Nov., 2013) 540–559, [`arXiv:1208.5098`].

[23] A. Shirokov, *GRAvitational COSmology (GRACOS) code release announcement, for version 1.0.1a9*, <u>ArXiv e-prints</u> (Nov., 2007) [`arXiv:0711.4655`].

[24] N. Padmanabhan, U. Seljak, and U. L. Pen, *Mining weak lensing surveys.*, <u>New Astronomy</u> **8** (2003) 581–603.

[25] P. McDonald, *Toward a Measurement of the Cosmological Geometry at z ~ 2: Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets*, <u>Astrophys. J.</u> **585** (Mar., 2003) 34–51, [`astro-ph/`].

[26] A. Arinyo-i-Prats, J. Miralda-Escudé, M. Viel, and R. Cen, *The non-linear power spectrum of the Lyman alpha forest*, JCAP **12** (Dec., 2015) 017, [arXiv:1506.0451].

[27] M. Tegmark, A. N. Taylor, and A. F. Heavens, *Karhunen-Loeve Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?*, Astrophys. J. **480** (May, 1997) 22+.

[28] S. Alam, F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, E. Armengaud, É. Aubourg, S. Bailey, and et al., *The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III*, ApJS **219** (July, 2015) 12, [arXiv:1501.0096].

[29] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot, and et al., *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems*, AJ **142** (Sept., 2011) 72–+, [arXiv:1101.1529].

[30] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind, V. Bhardwaj, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, A. S. Bolton, A. Borde, J. Bovy, W. N. Brandt, H. Brewington, J. Brinkmann, P. J. Brown, J. R. Brownstein, K. Bundy, N. G. Busca, W. Carithers, A. R. Carnero, M. A. Carr, Y. Chen, J. Comparat, N. Connolly, F. Cope, R. A. C. Croft, A. J. Cuesta, L. N. da Costa, J. R. A. Davenport, T. Delubac, R. de Putter, S. Dhital, A. Ealet, G. L. Ebelke, D. J. Eisenstein, S. Escoffier, X. Fan, N. Filiz Ak, H. Finley, A. Font-Ribera, R. Génova-Santos, J. E. Gunn, H. Guo, D. Haggard, P. B. Hall, J.-C. Hamilton, B. Harris, D. W. Harris, S. Ho, D. W. Hogg, D. Holder, K. Honscheid, J. Huehnerhoff, B. Jordan, W. P. Jordan, G. Kauffmann, E. A. Kazin, D. Kirkby, M. A. Klaene, J.-P. Kneib, J.-M. Le Goff, K.-G. Lee, D. C. Long, C. P. Loomis, B. Lundgren, R. H. Lupton, M. A. G. Maia, M. Makler, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, D. Margala, K. L. Masters, C. K. McBride, P. McDonald, I. D. McGreer, R. G. McMahon, O. Mena, J. Miralda-Escudé, A. D. Montero-Dorta, F. Montesano, D. Muna, A. D. Myers, T. Naugle, R. C. Nichol, P. Noterdaeme, S. E. Nuza, M. D. Olmstead, A. Oravetz, D. J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, W. J. Percival, I. Pérez-Fournon, I. Pérez-Ràfols, P. Petitjean, R. Pfaffenberger, J. Pforr, M. M. Pieri, F. Prada, A. M. Price-Whelan, M. J. Raddick, R. Rebolo, J. Rich, G. T. Richards, C. M. Rockosi, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martin, L. Samushia, A. G. Sánchez, C. Sayres, S. J. Schmidt, D. P. Schneider, C. G. Scóccola, H.-J. Seo, A. Shelden, E. Sheldon, Y. Shen, Y. Shu, A. Slosar, S. A. Smee, S. A. Snedden, F. Stauffer, O. Steele, M. A. Strauss, A. Streblyanska, N. Suzuki, M. E. C. Swanson, T. Tal, M. Tanaka, D. Thomas, J. L. Tinker, R. Tojeiro, C. A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, D. A. Wake, M. Watson, B. A. Weaver, D. H. Weinberg, B. J. Weiner, A. A. West, M. White, W. M. Wood-Vasey, C. Yeche, I. Zehavi, G.-B. Zhao, and Z. Zheng, *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, AJ **145** (Jan., 2013) 10, [arXiv:1208.0022].

[31] A. Font-Ribera, P. McDonald, and J. Miralda-Escudé, *Generating mock data sets for large-scale Lyman-α forest correlation measurements*, JCAP **1** (Jan., 2012) 1, [arXiv:1108.5606].

[32] J. E. Bautista, S. Bailey, A. Font-Ribera, M. M. Pieri, N. G. Busca, J. Miralda-Escudé, N. Palanque-Delabrouille, J. Rich, K. Dawson, Y. Feng, J. Ge, S. G. A. Gontcho, S. Ho, J. M. Le Goff, P. Noterdaeme, I. Pâris, G. Rossi, and D. Schlegel, *Mock Quasar-Lyman-α forest data-sets for the SDSS-III Baryon Oscillation Spectroscopic Survey*, JCAP **5** (May, 2015) 060, [arXiv:1412.0658].

[33] P. McDonald, J. Miralda-Escudé, M. Rauch, W. L. W. Sargent, T. A. Barlow, R. Cen, and J. P. Ostriker, *The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest*, Astrophys. J. **543** (Nov., 2000) 1–23.

[34] U. Seljak, *Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Ly-α forest*, JCAP **3** (Mar., 2012) 4, [arXiv:1201.0594].

[35] M. Blomqvist, D. Kirkby, J. E. Bautista, A. Arinyo-i-Prats, N. G. Busca, J. Miralda-Escudé, A. Slosar, A. Font-Ribera, D. Margala, D. P. Schneider, and J. A. Vazquez, *Broadband distortion modeling in Lyman-α forest BAO fitting*, JCAP **11** (Nov., 2015) 034, [arXiv:1504.0665].

Appendix A: Marginalizing over unwanted parameters

We start with a log-likelihood that depends on parameters that we are interested in (labelled with greek indices α, β) and parameters that we would like to marginalize over (labelled with latin indices i, j):

$$\mathcal{L}\left(\mathbf{p}\right) = \mathcal{L}\left(\mathbf{p}_0\right) + \mathcal{L}_{,\alpha}\ \delta\mathbf{p}_\alpha + \mathcal{L}_{,i}\ \delta\mathbf{p}_i + \frac{1}{2}\mathcal{L}_{,\alpha\beta}\ \delta\mathbf{p}_\alpha\ \delta\mathbf{p}_\beta + \mathcal{L}_{,\alpha i}\ \delta\mathbf{p}_\alpha\ \delta\mathbf{p}_i + \frac{1}{2}\mathcal{L}_{,ij}\ \delta\mathbf{p}_i\ \delta\mathbf{p}_j\ . \tag{A1}$$

Here we would like to calculate the marginalized likelihood

$$L_m \propto \exp\left[\mathcal{L}_{,\alpha}\ \delta\mathbf{p}_\alpha + \frac{1}{2}\mathcal{L}_{,\alpha\beta}\ \delta\mathbf{p}_\alpha\ \delta\mathbf{p}_\beta\right] \int d^N\mathbf{p}\ \exp\left[\frac{1}{2}\mathcal{L}_{,ij}\ \delta\mathbf{p}_i\ \delta\mathbf{p}_j + \left(\mathcal{L}_{,i} + \mathcal{L}_{,\alpha i}\ \delta\mathbf{p}_\alpha\right)\delta\mathbf{p}_i\right]\ , \tag{A2}$$

where the integral is over unwanted parameters [19]. To perform integration, we use the following identity:

$$\int_{-\infty}^{\infty} \mathcal{D}\mathbf{x} \exp\left(-\frac{1}{2}\mathbf{x}^t\mathbf{A}\mathbf{x} + \mathbf{b}^t\mathbf{x} + C\right) = (2\pi)^{N/2}\det\mathbf{A}^{-1/2}\exp\left(\frac{\mathbf{b}^t\mathbf{A}^{-1}\mathbf{b}}{2} + C\right) \tag{A3}$$

obtained by completing the square

$$-\frac{1}{2}\left[\mathbf{x}^t\mathbf{A}\mathbf{x} - 2\mathbf{b}^t\mathbf{x}\right] = -\frac{1}{2}\left[(\mathbf{x}^T - \mathbf{A}^{-1}\mathbf{b})^T\mathbf{A}(\mathbf{x} - \mathbf{A}^{-1}\mathbf{b})\right] + \frac{1}{2}\mathbf{b}^t\mathbf{A}^{-1}\mathbf{b}. \tag{A4}$$

After integration, we get

$$L_m \propto (\det\mathcal{L}_{,ij})^{-1/2}\exp\left[\mathcal{L}_{,\alpha}\,\delta\mathbf{p}_\alpha + \mathcal{L}_{,\alpha\beta}\,\delta\mathbf{p}_\alpha\,\delta\mathbf{p}_\beta + \frac{1}{2}(-\mathcal{L}_{,ij}^{-1})\left(\mathcal{L}_{,i} + \mathcal{L}_{,\alpha i}\,\delta\mathbf{p}_\alpha\right)\left(\mathcal{L}_{,j} + \mathcal{L}_{,\beta j}\,\delta\mathbf{p}_\beta\right)\right]. \tag{A5}$$

Therefore, we can compute the marginalized first and second derivatives:

$$\mathcal{L}_{m,\alpha} = \mathcal{L}_{,\alpha} + (-\mathcal{L}_{,ij}^{-1})\left(\mathcal{L}_{,\alpha i}\,\mathcal{L}_{,j}\right) \tag{A6}$$

$$\mathcal{L}_{m,\alpha\beta} = \mathcal{L}_{,\alpha\beta} + (-\mathcal{L}_{,ij}^{-1})\left(\mathcal{L}_{,\alpha i}\,\mathcal{L}_{,\beta j}\right). \tag{A7}$$

Note that in the equations above we need to invert the submatrix of second derivatives with respect to unwanted parameters, $\mathcal{L}_{,ij}$. If the likelihood is not sensitive to one of these parameters, this matrix will be non-positive definite. When needed, we regularise those direction in parameter space by adding a small negative value to the diagonal elements $\mathcal{L}_{,ii}$.

Appendix B: 3D power spectrum window function

In section VII we have motivated a particular form of $\frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}}$:

$$\frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}} = \int \frac{d\mathbf{k}_\perp}{(2\pi)^2}\,e^{i\boldsymbol{\Delta}\theta_\alpha\mathbf{k}_\perp}w_\gamma(z_\alpha, \mathbf{k}_\perp, k_\parallel^\alpha) \tag{B1}$$

This form is exact in the limit of infinite number of (intermediate) $p^\times$ and p^{3D} parameters. Given that our results do not vary if we use a finer grid of parameters, we believe that this form is accurate enough for our analysis.

If we wanted to use a coarser parameterization, we would have to compute an effective window function. Even though this is not done in our analysis, in this appendix we describe how it could be done.

Equation B1 implies that $p^\times$ parameters are modeled by

$$P_\times^{\times\mathrm{fid}}(\boldsymbol{\Delta}\theta_\alpha) + p_\alpha^\times = P_\times^{3D\mathrm{fid}}(\boldsymbol{\Delta}\theta_\alpha) + \sum_\gamma p_\gamma^{3D}\frac{\partial p_\alpha^\times}{\partial p_\delta^{3D}}, \tag{B2}$$

where $P_\times^{\times\mathrm{fid}}$ is the fiducial cross-power spectrum, and $P_\times^{3D\mathrm{fid}}$ is the cross-power spectrum computed from the fiducial 3D power spectrum. The equivalent of equation 19 is therefore

$$p_\gamma^{3D} = \sum_\delta J_{\gamma\delta}^{-1}\sum_\alpha \left[p_\alpha^\times + P_\times^{\times\mathrm{fid}}(\boldsymbol{\Delta}\theta_\alpha) - P_\times^{3D\mathrm{fid}}(\boldsymbol{\Delta}\theta_\alpha)\right]\frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}}, \tag{B3}$$

with

$$J_{\gamma\delta} = \sum_\alpha \frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}}\frac{\partial p_\alpha^\times}{\partial p_\delta^{3D}}. \tag{B4}$$

We also know that

$$p_\alpha^\times = I_{\alpha\beta}^{-1}\int d^2\boldsymbol{\Delta}\theta w_\beta^\times(\Delta\theta)\left(P_\times(\Delta\theta) - P_\times^{\mathrm{fid}}(\Delta\theta)\right), \tag{B5}$$

———

[19] To be accurate, we compute the integral of the posterior probability, i.e., we include a prior for the unwanted parameters. As discussed at the end of section II, we use Gaussian priors around zero, with a width set to a thousand times the fiducial power evaluated at the center of the band.

where

$$I_{\alpha\beta} = \int d^2 \boldsymbol{\Delta}\theta \, w_\alpha^\times(\Delta\theta) w_\beta^\times(\Delta\theta) \tag{B6}$$

and $w^\times$ are the cross-power interpolation kernels in angular direction.[20]

We can now put together equations B3 and B5 to get:

$$p_\gamma^{3D} = \sum_{\delta,\alpha,\beta} J_{\gamma\delta}^{-1} I_{\alpha\beta}^{-1} \iiint d^2 \boldsymbol{\Delta}\theta d^2 k_\perp d^2 k'_\perp \left(P^{3D}(k_\perp, k_\parallel, z) - P^{3D\mathrm{fid}}(k_\perp, k_\parallel, z) \right) e^{i(\Delta\theta k_\perp + \Delta\theta_\alpha k'_\perp)} w_\beta^\times(\Delta\theta) w_\delta^{3D}(k'_\perp) \tag{B11}$$

The above provide the prescription on how to calculate the window function in the $k_\perp$ direction and can be simplified significantly by inserting concrete forms for windows w. The response in $k_\parallel$ and z direction is given simply by Equation 19. For sufficiently narrow bins, $p_\gamma^{3D} = P^{3D}(k_{\perp\gamma}, k_\parallel, z) - P^{3D\mathrm{fid}}(k_{\perp\gamma}, k_\parallel, z)$.

---

[20] The above equations are trivially generalizable to the case with non-uniform error bars using (Einstein convention assumed):

$$p_\gamma^{3D} = J_{\gamma\delta}^{-1} \left[p_\alpha^\times - P_\times^{3D\mathrm{fid}}(\boldsymbol{\Delta}\theta_\alpha) \right] C_{\alpha\beta}^{\times-1} \frac{\partial p_\beta^\times}{\partial p_\gamma^{3D}} \ , \tag{B7}$$

$$J_{\gamma\delta} = \frac{\partial p_\alpha^\times}{\partial p_\gamma^{3D}} C_{\alpha\beta}^{\times-1} \frac{\partial p_\beta^\times}{\partial p_\delta^{3D}}. \tag{B8}$$

and non-uniform density of skewers given by $f(\boldsymbol{\Delta}\theta)$:

$$p_\alpha^\times = I_{\alpha\beta}^{-1} \int d^2 \boldsymbol{\Delta}\theta w_\beta^\times(\Delta\theta) f(\Delta\theta) \left(P_\times(\Delta\theta) - P_\times^{\mathrm{fid}}(\Delta\theta) \right) \ , \tag{B9}$$

$$I_{\alpha\beta} = \int d^2 \boldsymbol{\Delta}\theta w_\alpha^\times(\Delta\theta) w_\beta^\times(\Delta\theta) f(\Delta\theta) \ . \tag{B10}$$