Accelerating BAO Scale Fitting Using Taylor Series

Matthew Hansen, 1 * Alex Krolewski, 2,3 & Zachary Slepian 1,4

1 Department of Astronomy, University of Florida, Gainesville, Florida 32611, USA
2 AMTD Fellow, Waterloo Centre for Astrophysics, University of Waterloo, Waterloo ON N2L 3G1, Canada
3 Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada
4 Lawrence Berkeley National Laboratory, Berkeley, CA 94709, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The Universe is currently undergoing accelerated expansion driven by dark energy. Dark energy’s essential nature remains mysterious: one means of revealing it is by measuring the Universe’s size at different redshifts. This may be done using the Baryon Acoustic Oscillation (BAO) feature, a standard ruler in the galaxy 2-Point Correlation Function (2PCF). In order to measure the distance scale, one dilates and contracts a template for the 2PCF in a fiducial cosmology, using a scaling factor $\alpha$. The standard method for finding the best-fit $\alpha$ is to compute the likelihood over a grid of roughly 100 values of it. This approach is slow; in this work, we propose a significantly faster way. Our method writes the 2PCF as a polynomial in $\alpha$ by Taylor-expanding it about $\alpha = 1$, exploiting that we know the fiducial cosmology sufficiently well that $\alpha$ is within a few percent of unity. The likelihood resulting from this expansion may then be analytically solved for the best-fit $\alpha$. Our method is 48–85 $\times$ faster than a directly comparable approach in which we numerically minimize $\alpha$, and $\sim$12,000 $\times$ faster than the standard iterative method. Our work will be highly enabling for upcoming large-scale structure redshift surveys such as that by Dark Energy Spectroscopic Instrument (DESI).

1 INTRODUCTION

In the early, radiation-dominated universe, primordial overdensities in the relativistic plasma grew gravitationally and were opposed by the radiation pressure of the photons, creating spherical sound waves (Sakharov 1966; Peebles & Yu 1970; Sunyaev & Zeldovich 1970; Bond & Efstathiou 1984). These acoustic oscillations are a powerful cosmological probe, both in the Cosmic Microwave Background (CMB) anisotropies and in the Baryon Acoustic Oscillation (BAO) feature in the large-scale distribution of galaxies (Sakharov 1966; Eisenstein & Hu 1998; Linder 2003; Slepian & Eisenstein 2016). At recombination, the Universe becomes neutral and therefore transparent to photons. Slightly later, the photons decouple from the baryons and the sound waves halt. Velocity overshoot (i.e. that the density mode that grows to late times follows the spatial structure of the velocity at decoupling; Press & Vishniac 1980) leaves a sharp bump in the baryon density Green’s function (Slepian & Eisenstein 2016). Over time, matter collects around both the initial overdensity that sourced the sound wave, and the distant bump. The baryonic feature is a small perturbation compared to the central peak because the density at decoupling; Press & Vishniac 1980) leaves a sharp bump in the baryon density Green’s function (Slepian & Eisenstein 2016). Over time, matter collects around both the initial overdensity that sourced the sound wave, and the distant bump. The baryonic feature is a small perturbation compared to the central peak because the density that decays slow; in this work, we propose a significantly faster way. Our method writes the 2PCF as a polynomial in $\alpha$ by Taylor-expanding it about $\alpha = 1$, exploiting that we know the fiducial cosmology sufficiently well that $\alpha$ is within a few percent of unity. The likelihood resulting from this expansion may then be analytically solved for the best-fit $\alpha$. Our method is 48–85 $\times$ faster than a directly comparable approach in which we numerically minimize $\alpha$, and $\sim$12,000 $\times$ faster than the standard iterative method. Our work will be highly enabling for upcoming large-scale structure redshift surveys such as that by Dark Energy Spectroscopic Instrument (DESI).

The BAO feature was first clearly detected in the final release of the Two-degree-Field Galaxy Redshift Survey (2dFGS) (Cole et al. 2005) and the Sloan Digital Sky Survey (SDSS) Data Release 3 (DR3) (Eisenstein et al. 2005), though hints had been seen earlier (Percival et al. 2001). The BAO analyses of the DR7 SDSS Luminous Red Galaxy (LRG) and main galaxy samples established the BAO as a robust probe of cosmology (Percival et al. 2007; Eisenstein et al. 2007a). Subsequent analyses extended it to higher redshifts in the WiggleZ survey (Blake et al. 2011); $z \approx 0.1$ in the Six-degree-Field Galaxy Survey (6dFGS) (Beutler et al. 2011); improved precision by using reconstruction to partially reverse nonlinear smearing of the BAO peak (Eisenstein et al. 2007b; Padmanabhan et al. 2012); and much higher redshifts ($z \sim 2.5$) in the Ly-$\alpha$ forest (Busca et al. 2013; Slosar et al. 2013; Font-Ribera et al. 2014). The BAO was first detected in the 3-Point Correlation Function (3PCF) by Slepian et al. (2017) and in the bispectrum by Pearson & Samushia (2018); it has also now been detected in cosmic voids (Kitaura et al. 2016). The SDSS Baryon Oscillation Spectroscopic Survey (BOSS) was designed specifically to target a large sample of $z \sim 0.5$ LRGs for BAO measurements and yielded a 1 – 2% distance measurement at
z = 0.32 and z = 0.57 (Alam et al. 2017). Subsequently, the extended BOSS survey (eBOSS) extended the LRG sample of BOSS to z < 0.72, and added BAO measurements from Emission Line Galaxies (ELGs) and quasars at 1 < z < 2 to fill in the gap between the galaxies and the Ly-α forest (Alam et al. 2021). Ongoing and future surveys such as that by Dark Energy Spectroscopic Instrument (DESI), Euclid, and the Nancy Grace Roman Space Telescope (NGRST) promise considerably better precision. For instance, DESI should achieve

\[ \Delta \alpha = 0.1 \] (DESI Collaboration et al. 2016), and Euclid (Laureijs et al. 2011; Euclid Collaboration et al. 2020) and NGRST (Spergel et al. 2015) augur similar constraining power.

The BAO feature contains the most robust information about the cosmological parameters from galaxy clustering statistics. Some early studies directly fit ΛCDM models with varying parameters to the galaxy correlation function or power spectrum (Tegmark et al. 2006). On the other hand, other works instead fit a dilation parameter to the BAO feature and marginalize over the broadband shape (Eisenstein et al. 2005). This isolates the distance information, which is the most robust part of the BAO measurement, from the broadband shape, which is more challenging to model.

Subsequently, this dilation method developed into a standard method to determine the BAO scale (Xu et al. 2012; Anderson et al. 2012, 2014; Ross et al. 2015). In this method, a correlation function or power spectrum template is produced at a fixed cosmology, and the data is used to constrain the dilation parameter, \( \alpha \), that stretches or squeezes the template; we show this schematically in Figure 1. The dilation is different in the directions perpendicular to and parallel to the line of sight, producing a characteristic distortion of the BAO feature called the Alcock-Paczyński effect if one uses the wrong cosmology (Alcock & Paczyński 1979). Thus often two dilation parameters are used, \( \alpha_\parallel \) and \( \alpha_\perp \), and they are sometimes referred to as the Alcock-Paczyński parameters.\(^1\) The result of the BAO fitting is a best-fit value of \( \alpha \) and the error on \( \alpha \) (or \( \alpha_\parallel \) and \( \alpha_\perp \) and their covariance matrix), and subsequently cosmological parameters are fit using \( \alpha \) and its covariance matrix. Often this method is referred to as the "compressed statistics" approach (the other piece of it is \( f \sigma_8 \), from Redshift Space Distortions (RSD), with \( f \) the logarithmic derivative of the linear growth factor and \( \sigma_8 \) the rms density field fluctuations on 8 h\(^{-1}\) Mpc scales). This approach implicitly assumes a prior (from existing measurements) on the shape of the power spectrum, typically determined from the best-fit parameters from the CMB. This is a reasonable approach because the CMB constraints on the power spectrum shape are far more powerful than the large-scale structure constraints.

Recently, advances in perturbation theory (PT) modelling for the broadband power spectrum have led to renewed interest in directly fitting ΛCDM models to the power spectrum or correlation function, bypassing the compressed statistics approach (Ivanov et al. 2020; d’Amico et al. 2020; Philcox et al. 2020; Chen et al. 2021). However, the compressed statistics approach remains valuable for a variety of reasons: it is model-independent and provides a straightforward dataset to which to compare alternative cosmological models; it separates information from the early Universe and late Universe; and the compressed likelihood has a much smaller data vector and is easier to handle. As a result, the compressed statistics will continue to play an important role in future analyses, and we focus on accelerating their inference from the clustering statistics in this work.

The standard approach for finding \( \alpha \) is to produce templates over a wide range (typically 0.8 to 1.2) of \( \alpha \), compute the log likelihood of each template, and select \( \alpha \) with the highest likelihood or the lowest \( \chi^2 \) (e.g. as in the BAOFitt software of Ross et al. 2015).\(^2\) On average, it takes about 2 to 3 seconds to find the optimal \( \alpha \). In this work, we present a more efficient analytic method for obtaining the optimal \( \alpha \). Our method is 45–85\% faster than numerically minimizing \( \chi^2 \) as a function of \( \alpha \) using the same 2PCF model that our method uses (but without the Taylor expansion), or \( 12,000 \) times faster than the standard method from BAOFitt iterating over many values of \( \alpha \), which is not perfectly comparable as it does use a slightly more complicated 2PCF model.

Our method uses Taylor series to achieve this speedup. We start by calculating the 2PCF including only linear bias, and then Taylor-expand that model about \( \alpha = 1 \) (taking advantage of the fact that the best-fit cosmology is known quite precisely and hence \( \alpha \) will not be very far from this value). In §2 we construct mock data and define our Taylor series model. We choose to simplify our model and not include the additional polynomial terms of e.g. Ross et al. (2015). In §3 and §4 we show two cases: one with a fixed linear bias, and one where we marginalize over the linear bias. In §5, we compare the timing of our method to both that of the standard method as well as that of a numerical method to minimize \( \chi^2 \) in the same model that we use.\(^3\) We conclude in §6. In Appendix A we outline how to marginalize over additional polynomial terms in the model, designed to remove any broadband effects; we leave implementation for future work.

\(^1\) Although sometimes only a single isotropic parameter \( \alpha \) is fit due to the low signal-to-noise of the measurement (Ross et al. 2015; Carter et al. 2018; de Mattia et al. 2020).

\(^2\) https://github.com/ashleyjross/BAOfit

\(^3\) The repository for implementation of our method: https://github.com/matthew3hansen/BAO_Scaling_Taylor_Series
2 DEFINING THE 2PCF MODEL AND MOCK DATA

We create a mock dataset roughly mimicking the properties of the DESI survey (Levi et al. 2013; DESI Collaboration et al. 2016) to test our approach. We start by generating a linear power spectrum using CCL\(^4\) (Chisari et al. 2019), which in turn calls CLASS (Blas et al. 2011) with densities (in units of the critical density) for the matter of \(\Omega_m = 0.315\), for the baryons \(\Omega_b = 0.045\), \(h \equiv H_0/(100 \text{ km/s/Mpc}) = 0.67\), primordial amplitude of fluctuations \(A_s = 2.1 \times 10^{-9}\), and scalar spectral tilt \(n_s = 0.96\). To create a correlation function with realistic behavior around the BAO scale, we model the galaxy power spectrum \(P_{gg}\) as a linear bias times \(b\) the infrared-resummed power spectrum \(P_{IR-resum}\) at wavenumber \(k\) and redshift \(z\):

\[
P_{gg}(k,z) = b^2 P_{IR-resum}(k,z).
\] (1)

Infrared resummation allows for more accurate modelling of the bulk displacements in the BAO feature than does standard perturbation theory (Senatore & Zaldarriaga 2015). The infrared-resummed power spectrum is generated from the linear power spectrum using the FASTPT code (McEwen et al. 2016; Fang et al. 2017).

We work at \(z = 0.7\), a representative redshift for the DESI LRG sample. This galaxy sample has \(b = 1.7/D(z = 0.7) = 2.42\), where \(D\) is the linear growth factor. This form for the bias evolution is typically assumed for DESI LRGs (DESI Collaboration et al. 2016) and roughly matches early clustering data (Kitanidis et al. 2020). For the purposes of constructing data to test the method, we drop higher-order biases. We then inverse Fourier-transform the galaxy power spectrum \(P_{gg}(k)\) to the correlation function, \(\xi(r)\), and work with the correlation function throughout this paper. We have

\[
\xi_{gg}(r) = \int \frac{k^2 dk}{2\pi^2} j_0(kr)P_{gg}(k),
\] (2)

where \(j_0(x) = \sin x/x\) is the spherical Bessel function of order zero.

We use the analytic expression for the correlation function’s covariance matrix (Cohn 2006),

\[
\text{Cov}(\xi(r), \xi(r')) = \frac{1}{V_{\text{eff}}^2} \int k^2 dk j_0(kr)j_0(k'r')P(k)^2
\]

\[
+ \frac{2}{V_{\text{eff}}^2 \pi^2} \int k^2 dk j_0(kr)j_0(k'r')\partial P(k)\]

\[
+ \frac{2}{V_{\text{eff}}^2 \pi^2} \int k^2 dk j_0(kr)j_0(k'r')(1 + \xi(r)).
\] (3)

\(\delta_D^{[1]}\) is the 1D Dirac delta function. The first two terms above are produced by the Gaussian density field and the final term by the shot noise in the galaxy distribution. \(V_{\text{eff}}\) is the effective volume, related to the survey volume \(V\) by

\[
V_{\text{eff}} = \left(1 + \frac{1}{b^2 V_{\text{eff}}}\right)^{-2} V.
\] (4)

The survey volume is taken as the volume of a spherical shell between \(z = 0.6\) and \(z = 0.8\), and the number density is \(\bar{n} = 6 \times 10^{-4} \text{ h}^3 \text{ Mpc}^{-3}\), appropriate for the DESI LRG sample (Zhou et al. 2020).

The effective anisotropic power spectrum \(P_{eff}\) (i.e. power spectrum at roughly the average wavenumber \(k\) and angle to the line of sight \(\mu\)) is computed at \(k = 0.14 \text{ h Mpc}^{-1}\) and \(\mu = 0.6\) (DESI Collaboration et al. 2016) and is given by

\[
P_{\text{eff}} = \left(b + f(\mu)^2\right)^2 P_{\text{lin}}(k,z).
\] (5)

Our fiducial binning consists of 30 bins with width \(\Delta r = 5 \text{ h}^{-1}\) Mpc between \(r = 30 \text{ h}^{-1}\) Mpc and \(r = 180 \text{ h}^{-1}\) Mpc. We average the covariance over the bins, and show the fiducial data and binned covariance in Figure 2. In Figure 3, we show the log likelihood of the Taylor expanded model compared to the standard model. In Figure 4 and Figure 5, we produce 100 different mock data sets, each with a different value of \(\alpha\) corresponding to the labeled “True \(\alpha\”).

We also specifically choose not to add noise into the data that we use to test the best-fit \(\alpha\) in the Taylor series model. Adding noise would introduce error into our Taylor series model, since the best-fit \(\alpha\) would be equal to the true \(\alpha\) plus a small random number. This noise-produced offset in \(\alpha\) would interfere with our aim of comparing to the input \(\alpha\) specifically to ensure that our method’s Taylor expansion retain the desired accuracy in the recovered \(\alpha\). Since we know the value of \(\alpha\) is close to unity, we can Taylor-expand \(\xi(\alpha r)\) about \(\alpha = 1\) as

\[
\xi(\alpha r) = \xi(r) + \xi'(r)\Delta \alpha + \frac{1}{2} \xi''(r)\Delta \alpha^2 + \frac{1}{6} \xi'''(r)\Delta \alpha^3,
\] (6)

where \(\Delta \alpha \equiv \alpha - 1\) and prime denotes the partial derivative with respect to \(\alpha\). We Taylor-expand to third order in \(\alpha\), which allows for a consistent expansion to second order in the log likelihood, as will be seen. We show in §3 that a Taylor expansion to second order is insufficient to achieve the necessary accuracy in \(\alpha\).

A similar Taylor expansion was presented in Xu et al. (2013), though they expanded both the isotropic and anisotropic parts of the BAO scaling, and kept only first-order terms. Our work expands on this method, and shows that to achieve sufficient accuracy for current-generation surveys, we must expand to third order rather than first order.

2.1 Template Derivatives

Since we Taylor-expand our model to third order, we require the first, second and third derivatives of \(\xi(\alpha r)\) with respect to \(\alpha\), evaluated at \(\alpha = 1\). To enable efficiently taking these derivatives without resorting to numerical differencing, which can suffer from stability issues, we begin with our definition of \(\xi(\alpha r)\). We have:

\[
\xi(\alpha r) = \int \frac{k^2 dk}{2\pi^2} k j_0(\alpha kr)P_{\text{lin}}(k).
\] (7)

Taking \(\partial / \partial \alpha\), we find:

\[
\frac{\partial}{\partial \alpha} \xi(\alpha r) = \int \frac{k^2 dk}{2\pi^2} \frac{\partial}{\partial \alpha} [k j_0(\alpha kr)]P_{\text{lin}}(k).
\] (8)

We use the recursion relation for the spherical Bessel functions:\(^5\)

\[
j'_n(z) = -j_{n+1}(z) + \frac{n}{z} j_n(z), \quad n = 0, 1, 2, 3 \ldots
\] (9)

We find that

\[
\frac{\partial}{\partial \alpha} \xi(\alpha r) \bigg|_{\alpha = 1} = -r \int \frac{k^2 dk}{2\pi^2} k j_1(\alpha kr)P_{\text{lin}}(k).
\] (10)

Applying equation (9) again to obtain the second derivative, we find

\[
\frac{\partial^2}{\partial \alpha^2} \xi(\alpha r) \bigg|_{\alpha = 1} = r^2 \int \frac{k^2 dk}{2\pi^2} k^2 \left[ j_2(\alpha kr) - \frac{1}{k r} j_1(\alpha kr) \right] P_{\text{lin}}(k).
\] (11)

\(^4\) https://github.com/LSSTDESC/CCL

\(^5\) https://dlmf.nist.gov/10.51
The third derivative is then:

\[
\frac{\partial^3}{\partial \alpha^3} \xi(\alpha) \bigg|_{\alpha=1} = r^3 \int \frac{k^2dk}{2\pi^2} k^3 \left[ -j_3(kr) + \frac{2}{kr} j_2(kr) \right] P_{\text{lin}}(k) \\
+ r \int \frac{k^2dk}{2\pi^2} k^2 j_1(kr) P_{\text{lin}}(k) \\
+ r^2 \int \frac{k^2dk}{2\pi^2} k^2 \left[ j_2(kr) + \frac{1}{kr} j_1(kr) \right] P_{\text{lin}}(k).
\]

With these templates in hand, we may now proceed to evaluate the Taylor expansion (6) and then solve for the maximum-likelihood \( \Delta \alpha \).

### 3 OPTIMAL \( \alpha \) AT FIXED LINEAR BIAS

We now turn our focus to the log likelihood of our data with respect to the model. To analytically obtain the optimal value for \( \alpha \), we maximize the log likelihood by setting its derivative to zero. By retaining terms up to third order in \( \Delta \alpha \), we will obtain a quadratic in \( \alpha \) when taking the likelihood’s derivative, which we can then solve analytically for the optimal \( \Delta \alpha \). Primes will denote the derivatives with respect to \( \alpha \), \( \vec{d} \) denotes the data vector, and \( \vec{m} \) denotes the model vector. For these vectors, each element describes, respectively, the data and model correlation functions in a given radial bin.

The log likelihood is:

\[
-\ln \mathcal{L} = 0.5 \left[ \vec{d} - \vec{m} \right] \mathbf{C}^{-1} \left[ \vec{d} - \vec{m} \right]^T,
\]

where \( \mathbf{C} \) is the covariance matrix. We define \( \mathbf{C}^{-1} \) to have elements \( \mathcal{P}^{lm} \), where \( \mathcal{P} \) is the precision matrix. The superscripts \( l \) and \( m \) identify the radial bin of each of the data or model vectors involved. Using Einstein summation convention, the log likelihood becomes:

\[
-\ln \mathcal{L} = 0.5 \left[ d_l - m_l \right] \mathcal{P}^{lm} \left[ d_m - m_m \right].
\]

Since all of these elements are now scalars, their order no longer matters. Hence we may rewrite the log likelihood as

\[
-\ln \mathcal{L} = 0.5 \mathcal{P}^{lm} \left[ d_l d_m - d_l m_m - m_l d_m + m_l m_m \right].
\]

Now, taking the derivative of the log likelihood with respect to \( \Delta \alpha \), we find

\[
\frac{\partial}{\partial (\Delta \alpha)} \left[ -\ln \mathcal{L} \right] = 0.5 \mathcal{P}^{lm} \left[ -d_l m_m' - m_l' d_m + m_l' m_m + m_l m_m' \right] = 0.
\]
Our model vector has elements $m_i$ (the value at the $i$th radial bin) as:

$$m_i = B \xi(\alpha r_i),$$

where $\xi(\alpha r)$ is defined in equation (6) and $B \equiv b^2$, with $b$ the linear bias. Figure 3 shows the log likelihood as a function of $\alpha$ and as a function of $B$. These plots serve as our basis for approximating the model as a third-order Taylor series about $\alpha$. Figure 3 also argues that we need to Taylor expand the model to third order rather than second order. At the point of expansion $\alpha = 1$, the difference between third and second order is negligible near the maximum log likelihood. However, at $\alpha$ far from 1, third order is significantly more accurate than second order, as shown in the upper right panel of Fig. 3 at $\alpha = 1.07$. Thus, going to third order significantly increases the range of $\alpha$ over which our method is accurate.

We now need to write out any terms in equation (16) involving the product of two models at different radial bins up to second order in $\Delta \alpha$. We substitute into equation (16) using equation (17) for the model, revealing what degree in $\Delta \alpha$ each term contains. This substitution will allow us to group the terms in the equation into a quadratic in terms of $\Delta \alpha$:

$$\frac{\partial}{\partial(\Delta \alpha)} [-\ln L] = a(\Delta \alpha)^2 + b(\Delta \alpha) + c,$$

with coefficients $a$, $b$ and $c$ as:

$$a = \frac{1}{4} B p^{lm} \left[ B \left( 3 \xi'_l \xi''_m + 3 \xi'_l \xi''_m + 3 \xi'_l \xi''_m + 3 \xi'_l \xi''_m \right) - d \xi''_m - d_m \xi''_m \right],$$

$$b = \frac{1}{2} B p^{lm} \left[ B \xi'_l \xi''_m + B \xi'_l \xi''_m + 2 B \xi'_l \xi''_m - d \xi''_m - d_m \xi''_m \right],$$

$$c = \frac{1}{2} B p^{lm} \left[ B \xi'_l \xi''_m + B \xi'_l \xi''_m - d \xi''_m - d_m \xi''_m \right].$$

We note here that $a$, $b$ and $c$ are scalars.

With these coefficients in hand, we may now solve the quadratic (18) for $\Delta \alpha$ using the quadratic formula, as

$$\Delta \alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

with $a$, $b$ and $c$ defined as in equation (19). The solution for $\Delta \alpha$, equation (20), is only valid $B$ is held constant.

In Figure 4, we hold $B$ fixed to a value of 5.88 and create 100 different mock data sets, each with a different value of $\alpha$ ranging from 0.9 to 1.1 in steps of 0.002. We then compare the value equation (20) returns for $\Delta \alpha$ to the inputted true value of $\Delta \alpha$. We note that the the recovered $\alpha$, $\alpha_R$, in Figure 4 is slightly different from the peak of the likelihood in the right panel of Figure 3. This is because the

Figure 3. Top left: The likelihood for the second-order and third-order Taylor expansions (respectively dotted blue and dashed red) agree extremely well with the full model (solid black) with the mock data scaled by $\alpha = 1.0$. The standard method (solid black) is one where $\alpha$ is varied with a spline interpolation. Top right: At $\alpha = 1.07$, the second-order Taylor expansion has a large discrepancy in the $\ln L$ compared to both the standard method and to the third-order Taylor expansion. Bottom left: The likelihood as a function of the bias squared, i.e. $B \equiv b^2$. The likelihood is peaked at the true value in the mock data ($b = 2.42, B = 5.88$). Bottom right: The likelihood of the third-order Taylor expansion after marginalizing over $B$, for a range of $\alpha$. 

---

**MNRAS 000, 1–11 (2021)**
likelihoods in Figure 3 include the higher-order $\Delta^3$ and $\Delta^4$ terms, but these are omitted in the Taylor series analytic solution.

This procedure’s goal is to confirm that our method of finding $\Delta\alpha$ is accurate over a wide range of input $\Delta\alpha$. Our method is accurate to 0.7% over an extremely wide range of $\alpha$, $0.9 < \alpha < 1.1$. This compares favorably to the errors on current and future surveys: BOSS constrains $\alpha$ to $\sim 1\%$ precision (Alam et al. 2017) and DESI will constrain $\alpha$ to $\pm 0.5\%$ precision (DESI Collaboration et al. 2016). In Figure 4, we add horizontal lines corresponding to 50% of the statistical error on these two surveys (i.e. 0.25% and 0.5%) as a guide to determine the $\alpha$ at which our method stops working. For 0.5% errors on the distance scale, our method for the third-order Taylor expanded model works from $0.91 < \alpha < 1.16$. For 0.25% distance errors, our method works from $0.92 < \alpha < 1.15$. In contrast, the second-order Taylor-expanded model works to 0.5% as $0.96 < \alpha < 1.055$, and 0.25% at $0.975 < \alpha < 1.03$. Moreover, the second-order Taylor-expanded model rapidly rises to very large errors at $\alpha_T < 0.95$. While the second-order model may be accurate enough for DESI given that $\alpha_T$ will be quite close to one, we prefer the third-order model as it is far more robust away from $\alpha = 1$ (and therefore may work better when, for instance, iterating over large numbers of mock data, some of which may have larger or smaller values of $\alpha$ just by chance).

This successful simplified case of finding the best fit $\Delta\alpha$ at a fixed $B$ means we can now extend our approach to more complicated models. In the present section, we have treated $B$ as a known constant, which is not entirely correct: typically the galaxy bias is unknown and marginalized over. Hence the present section serves as a proof of concept for our method. The next section, §4, treats the more realistic case wherein one marginalizes over the unknown value of $B$.

4 OPTIMAL $\Delta\alpha$ WITH MARGINALIZATION OVER LINEAR BIAS

We introduce marginalization over $B$ as a means of accounting for $B$ to mirror what is done in realistic BAO analysis, where the galaxy bias is unknown. Figure 3 shows the likelihood is peaked about the inputted $\alpha$ with marginalization over $B$. We use the same definitions of $L$, equation (15), and $\bar{B}$, equation (17). We write the likelihood as a quadratic in $B$ to facilitate marginalization,

$$L(B) = \exp \left(-aB^2 + bB + c\right),$$

with coefficients

$$a = \frac{1}{2} \mathcal{P}^{lm}_{\ell=1,m} \left[ \xi_m \xi_l + \frac{1}{2} \xi_m \xi_l' \Delta\alpha^2 + \frac{1}{2} \xi_m' \xi_l' \Delta\alpha^2 + \frac{1}{4} \xi_m'' \xi_l'' \Delta\alpha^4 \right],$$

$$b = -\frac{1}{2} \mathcal{P}^{lm}_{\ell=1,m} \left[ \xi_l + \xi_l' \Delta\alpha + \xi_l'' \Delta\alpha^2 + \frac{1}{2} \xi_l'' \Delta\alpha^2 + \frac{1}{6} \xi_l''' \Delta\alpha^3 \right],$$

$$c = -\frac{1}{2} \mathcal{P}^{lm}_{\ell=1,m} \left[ d_l d_m \right].$$

We choose to marginalize over $B$ with a uniform prior from $-\infty$ to $\infty$, as the resulting integral yields an analytic solution. Although negative values of $B$ are a unphysical solution, Figure 3 shows that the likelihood is very sharply peaked around the true value of $B$. Thus, the unphysical region from $-\infty$ to 0 makes a very small contribution to the total integral. Therefore, the result should be the same with tighter choices for the prior on $B$, or if we applied a prior on $b$ instead of $B$.

Performing the marginalization over $B$, we obtain

$$L_{\text{marg}} = \int_{-\infty}^{\infty} L(B) d\bar{B} = \frac{\pi}{\sqrt{a}} \exp \left(\frac{b^2}{4a} + c\right)$$

We choose to marginalize over $B$ with a uniform prior from $-\infty$ to $\infty$, as the resulting integral yields an analytic solution. Although negative values of $B$ are a unphysical solution, Figure 3 shows that the likelihood is very sharply peaked around the true value of $B$. Thus, the unphysical region from $-\infty$ to 0 makes a very small contribution to the total integral. Therefore, the result should be the same with tighter choices for the prior on $B$, or if we applied a prior on $b$ instead of $B$.

Performing the marginalization over $B$, we obtain

$$L_{\text{marg}} = \int_{-\infty}^{\infty} L(B) d\bar{B} = \frac{\pi}{\sqrt{a}} \exp \left(\frac{b^2}{4a} + c\right)$$

As in §3, we take the log of the marginalized likelihood, differentiate with respect to $\Delta\alpha$, and set it equal to zero. This again yields a quadratic in terms of $\Delta\alpha$, which is then easily solved. We have

$$\ln L_{\text{marg}} = \ln \left(\frac{\pi}{\sqrt{a}} \exp \left(\frac{b^2}{4a} + c\right)\right)$$

meaning that

$$\frac{\partial}{\partial \Delta\alpha} \ln L_{\text{marg}} = \frac{\partial}{\partial \Delta\alpha} \ln \left(\frac{\pi}{\sqrt{a}} \exp \left(\frac{b^2}{4a} + c\right)\right)$$

where

$$\frac{1}{a} \mid_{\Delta\alpha=0} = \frac{1}{c_a} - \frac{b_a}{c_a} \Delta\alpha + \frac{1}{2} \left(\frac{2b_a^2}{c_a^2} - \frac{2a_a}{c_a} \right) \Delta\alpha^2 + \cdots$$

To better organize the equations and recognize what terms we are able to drop, we group the coefficients, equation (22), with respect to $\Delta\alpha$. Since we started our process ignoring terms higher than second order in $\Delta\alpha$ in the derivative of the log likelihood, we continue to do so here in order to consistently capture all terms at second order.

In what follows, we use a specific variable naming convention to indicate whence each term originates. A variable called $a$ will denote that it comes from a factor of $(\Delta\alpha)^2$. A variable called $b$ will denote linearity in terms of $\Delta\alpha$. Finally, a variable called $c$ arises from the constant in the quadratic polynomial. This variable convention follows the familiar form of $a\Delta\alpha^2 + b\Delta\alpha + c$, so in an equation the dimensions in terms of $\Delta\alpha$ are easily checked. A subscript $a$ represents that that specific variable came from the $a$ coefficient of the $B$ quadratic polynomial. A subscript of $b$ represents that that specific variable belongs to the linear term in the $B$ quadratic polynomial, and so on. A subscript of $da$ means that the variable with that subscript came from the group of variables deriving from the partial of $a$ with respect to $\Delta\alpha$, and so on with $db$ and $dc$.

With this variable naming convention, the terms in equation (25) may be expanded and grouped in terms of $\Delta\alpha$. This enables us to distribute the terms and easily track the order of each term in $\Delta\alpha$. 

MNRA 000, 1–11 (2021)
We have
\[ a = \frac{1}{2} p_{\text{lm}} \left[ \xi_m \xi_1 + \left( \xi_m \xi'_1 + \xi'_m \xi_1 \right) \Delta \alpha \right. \]
\[ + \left. \frac{1}{2} \xi_m \xi''_1 + \xi'_m \xi'_1 + \frac{1}{2} \xi''_m \xi_1 \right] \Delta \alpha^2 \]
\[ + \left. \frac{1}{2} \xi'_m \xi''_1 + \frac{1}{2} \xi''_m \xi'_1 + \frac{1}{2} \xi''_m \xi_1 + \frac{1}{2} \xi''_m \xi''_1 \right] \Delta \alpha^3 \]
\[ = a_b \Delta \alpha^2 + b_a \Delta \alpha + c_a, \] (27)
where \( a_b, b_a \) and \( c_a \) are defined by the last line.

Taking the partial of \( a \) with respect to \( \Delta \alpha \), we obtain
\[ \frac{\partial a}{\partial \Delta \alpha} = \frac{1}{2} p_{\text{lm}} \left[ \xi_m \xi_1 + \left( \xi_m \xi'_1 + \xi'_m \xi_1 \right) \right. \]
\[ + \left. \frac{1}{2} \xi_m \xi''_1 + \xi'_m \xi'_1 + \frac{1}{2} \xi''_m \xi_1 \right] \Delta \alpha \]
\[ + \left. \frac{1}{2} \xi'_m \xi''_1 + \frac{1}{2} \xi''_m \xi'_1 + \frac{1}{2} \xi''_m \xi_1 + \frac{1}{2} \xi''_m \xi''_1 \right] \Delta \alpha^2 \]
\[ = a_{ab} \Delta \alpha^2 + b_{ab} \Delta \alpha + c_{ab}. \] (28)

We need to both expand \( b \) in order to recover a second-degree polynomial in terms of \( \Delta \alpha \):
\[ b = \frac{1}{2} p_{\text{lm}} \left[ -d_i \xi_m - d_m \xi'_i \right] \Delta \alpha \]
\[ + \left. \frac{1}{2} \left( -d_i \xi''_m - d_m \xi'''_i \right) \Delta \alpha^2 \right] \]
\[ = a_{b} \Delta \alpha^2 + b_{ab} \Delta \alpha + c_{b}. \] (29)

We also need to expand the partial of \( b \) with respect to \( \Delta \alpha \); doing so we obtain
\[ \frac{\partial b}{\partial \Delta \alpha} = \frac{1}{2} p_{\text{lm}} \left[ -d_i \xi_m - d_m \xi'_i \right] \Delta \alpha \]
\[ + \left. \frac{1}{2} \left( -d_i \xi''_m - d_m \xi'''_i \right) \Delta \alpha^2 \right] \]
\[ = a_{ab} \Delta \alpha^2 + b_{db} \Delta \alpha + c_{db}. \] (30)

We return briefly to the question of prior on \( B \). If we took the prior to be from \( 0 \) to \( \infty \), we would, in equation (23), get an error function for \( L_{\text{marg}} \). We could then Taylor expand it, much as we Taylor-expanded \( a^{-1} \) and \( a^{-2} \) in equation (26). However, because, as discussed below equation (23), the likelihood is sharply peaked in any case, we use a prior from -\( \infty \) to \( + \infty \), which yields the simpler forms used above.

Substituting equations (27) through (30) into equation (25), we are able to distribute terms and pull out a quadratic which can then be solved to give us the optimal value of \( a \). We have for the derivative of the log likelihood:
\[ \frac{\partial}{\partial \Delta \alpha} [\ln L_{\text{marg}}] = -\frac{1}{2} \frac{\partial a}{\partial \Delta \alpha} + \frac{b}{2a} \frac{\partial b}{\partial \Delta \alpha} - \frac{b^2}{4a^2} \frac{\partial a}{\partial \Delta \alpha} = \]
\[ -\frac{1}{2} \left[ \frac{1}{c_a} - \frac{b_a \Delta \alpha + 1}{2} \left( \frac{2a}{c_a} - \frac{2a}{c_a^3} \right) \Delta \alpha \right] \]
\[ \times \left[ a_{ab} \Delta \alpha^2 + b_{db} \Delta \alpha + c_{db} \right] + \frac{1}{2} \left[ a_b \Delta \alpha^2 + b_b \Delta \alpha + c_b \right] \left( \frac{1}{c_a} - \frac{b_a \Delta \alpha + 1}{2} \left( \frac{2a}{c_a} - \frac{2a}{c_a^3} \right) \Delta \alpha \right] \]
\[ \times \left[ a_{ab} \Delta \alpha^2 + b_{db} \Delta \alpha + c_{db} \right] - \frac{1}{4} \left( a_b \Delta \alpha^2 + 2b_b \Delta \alpha + c_b \right) \]
\[ \times \left( \frac{1}{2} \left( 6a^2 - 4 \frac{a_a}{c_a} \right) \Delta \alpha^2 + 2b_b \Delta \alpha + c_b \right) ] \times \left[ a_{ab} \Delta \alpha^2 + b_{db} \Delta \alpha + c_{db} \right]. \] (31)
We can extract a quadratic equation in terms of $\Delta \alpha$:

$$
\frac{\partial}{\partial \Delta \alpha} \left[ \ln L_{\text{marg}} \right] = \left[ - \frac{a_{da} + b_{db} b_{da}}{2c_a} - \frac{a_{db} c_{da}}{2c_a} + \frac{2b_{db} b_{db}}{2c_a} - \frac{2a_{da}}{4c_a} + \frac{2a_{db} b_{da}}{2c_a} + \frac{2b_{db} b_{db}}{2c_a} - \frac{2a_{da}}{4c_a} \right] \Delta \alpha^2 + \left[ - \frac{b_{db} c_{db}}{2c_a} - \frac{b_{db} c_{db}}{2c_a} + \frac{b_{db} c_{db}}{2c_a} - \frac{b_{db} c_{db}}{2c_a} \right] \Delta \alpha + \left[ c_{db} c_{db} - \frac{c_{db} c_{db}}{2c_a} - \frac{c_{db} c_{db}}{2c_a} - \frac{c_{db} c_{db}}{2c_a} \right].
$$

(32)

With the likelihood optimization equation now written as a quadratic, it is easily solved for the optimal $\Delta \alpha$.

Figure 5 displays the recovered values for the optimal $\alpha$ over a range of different inputted $\alpha$. In Figure 5, for 0.5% errors on the distance scale, our method with a third-order Taylor-expanded model works in the range of $0.94 < \alpha < 1.1$. If we demand at most 0.25% errors in the recovered $\alpha$, our method may be employed in the range $0.954 < \alpha < 1.05$. The range over which the third order method works to 0.5% or 0.25% is not much larger than the range over which the second order method works, but the third order method is considerably more robust to large errors in $\alpha$ just outside of that range. Furthermore, the accuracy achieved by the third order method is sufficient to make only a small systematic error contribution to the DESI measurement of $\alpha$.

5 PERFORMANCE COMPARISON

We compare the timing of our method to the timing of the standard method as implemented in the BAOfr package. The method used by BAOfr is described in Ross et al. (2015), Anderson et al. (2014) and Tojeiro et al. (2014), and has been employed in a variety of BAO studies. We use the BAOfr isotropic BAO fitting method, which also includes marginalization over polynomial broadband terms and splits the BAO feature into a wiggle piece and a no-wiggle piece. This method analytically finds the best-fit polynomial coefficients for the broadband correlation function, and places a prior on the bias parameter, $B_{\xi}$. The prior is a Gaussian in $\log[B_{\xi}/B_{\text{best}}]$ with width 0.4, where $B_{\text{best}}$ is the best-fit $B_{\xi}$ when $\alpha = 1$. While we discuss how to add polynomial terms to our method in Appendix A, our method as implemented is simpler than that of BAOfr. Due to this difference in methods, we also compare the timing of our method to a numerical minimization of the likelihood in equation (14), using a spline interpolation to define $\xi(\alpha r)$ rather than a Taylor series, as a more appropriate comparison.

The times for each method were measured on a 6-core Intel Core i7-8750H CPU at 2.20GHz, and reported in Table 1. The mean and standard deviation were computed using 7 runs and 10,000 loops each for the Taylor series and the numerical minimization of splined $\xi(\alpha r)$.

| Method                              | Fixed $B$ (mean ± std.) | Marginalized $B$ (mean ± std.) |
|-------------------------------------|-------------------------|---------------------------------|
| Third-order Taylor series           | 84 $\mu$s ± 782 ns      | 170 $\mu$s ± 2.04 $\mu$s       |
| plus analytic solution              |                         |                                 |
| BAOfr                              | ---                     | 2.04 s ± 30.6 ms                |
| Numerical minimization of $\xi(\alpha r)$ | 7.25 ms ± 103 $\mu$s | 8.2 ms ± 123 $\mu$s             |

Table 1. The average time for each method to find the optimal $\alpha$.

6 DISCUSSION AND CONCLUSIONS

In this work, we have shown that we can accelerate the time to fit the BAO dilation parameter $\alpha$ to measurements of the galaxy two-point correlation function, thus allowing for faster measurements of the cosmic distance scale as encoded by $\alpha$. We perform a third-order Taylor expansion about $\alpha = 1$ to allow us to analytically find the best-fit $\alpha$. We have shown the method works both at fixed linear bias and when marginalizing over linear bias. Our method is 12,000× faster than the standard method, in which many values of $\alpha$ are looped over to find the maximum likelihood $\chi^2$. While the BAOfr code for the standard method is somewhat more complicated than our fitting procedure, we note that our method is still 48 – 85× faster than numerically finding the maximum likelihood using scipy’s numerical optimizer.

The accuracy of our method is less than the expected statistical error on the DESI survey, ensuring that it will contribute as a negligible systematic error. The performance of the third-order expansion is considerably better than second-order: third-order allows us to recover $\alpha$ to within 0.8% across a 10% range in $\alpha$, whereas second-order yields a rapidly increasing error when $|\Delta \alpha| > 0.05$.

Further work remains to be done in providing an analytic solution to the optimal scaling factor. In this paper we have simplified our model to contain only the effects of the linear bias, and have outlined the math needed to include additional polynomial terms and marginalize over them, but not implemented it. However, even with the polynomial terms, this might still be considered an incomplete model. A direction of further work might be to expand our model to contain the effects of multiple bias terms, such as the quadratic, tidal tensor (McDonald & Roy 2009; Baldauf et al. 2012), and baryon-dark matter relative velocity biases (Yoo et al. 2011; Slepian et al. 2018; Blazek et al. 2016; Slepian & Eisenstein 2015; Beutler et al. 2017; Schmidt 2016). Moreover, we should modify our model to use the wiggle power spectrum rather than the infrared-resummed power spectrum in equation (1). However, we believe there is no reason to assume the methods we present in this paper would not carry over to a more complicated model.

We note that a similar method could be developed for BAO in the 3PCF, as the relevant templates for the 3PCF already exist. Slepian & Eisenstein (2017) presents the 3PCF model in terms of double-spherical-Bessel-function integrals from which the analyti-
A marginal derivative can be found. In Umeh (2021), the anisotropic 3PCF is also presented in terms of double spherical Bessel function integrals. The double-spherical-Bessel-function-integral templates could also be Taylor-expended in $\alpha$ using the same recursion relations as in our work here.

Overall, we believe the present work is already a useful step towards a complete analytic approach to finding the optimal scaling factor for BAO analyses.

**ACKNOWLEDGEMENTS**

MH would like to thank ZS and AK for allowing me to pursue this area of research and donating countless hours of their time to helping me become a more thoughtful and complete researcher. MH would also like to thank Jiamin Hou, Bob Cahn, Oliver Philcox, Kiersten Meigs, Nina Brown, Jessica Chellino and Farshad Kamalinejad for useful discussions in our weekly group meetings. ZS acknowledges useful conversations with Daniel J. Eisenstein, Lado Samushia, Hee-Jong Seo, and Stephen Portillo.

**APPENDIX A: INCORPORATING MARGINALIZATION OVER POLYNOMIAL TERMS**

Often, in BAO fitting, one marginalizes over inverse power laws (or equivalently, polynomials in $1/r$) (e.g. Ross et al. 2015); this is intended to render the BAO constraints robust to broadband variations in the 2PCF, as might occur due to e.g. inhomogeneities in the imaging depth of the targeting survey. Broadband effects can also arise from mis-normalization of the average number density in the survey (failure of the integral constraint). Furthermore, the radial dependence of the templates associated with both quadratic and tidal tensor biases are fairly smooth, and it is expected that marginalizing over these inverse power laws will also render the BAO fitting robust against neglecting these bias terms in the 2PCF model.

Here, we show how such marginalization can be incorporated in our framework; though the math is extensive it is conceptually simple. We leave implementation of these formulae for future work, but present them here to show that indeed in principle such an effort would be straightforward.

Similarly to our approach in §4, we will marginalize over $B$, as well as the additional polynomial terms, $a_1, a_2, a_3$. Our new model is:

$$
\tilde{m}(r) = B\left(\xi(r) + \xi'(r) \Delta \alpha + \frac{1}{2} \xi''(r) \Delta \alpha^2 + \frac{1}{6} \xi'''(r) \Delta \alpha^3 \right) + \left(\frac{a_1}{r^2} + \frac{a_2}{r} + a_3\right).
$$

(A1)

The next several steps follow the same approach as in §4. We first expand out the likelihood to see the dependence of each term on $B$. We have

$$
\mathcal{L} = \exp\left[-\frac{1}{2} \chi^2\right] = \exp\left[-\frac{1}{2} \mathcal{P}^m \left( d_1d_m - d_1 B\left(\xi_m + \xi'_m \Delta \alpha + \frac{1}{2} \xi''_m \Delta \alpha^2 + \frac{1}{6} \xi'''_m \Delta \alpha^3 \right) + \left(\frac{a_1}{r^2_m} + \frac{a_2}{r_m} + a_3\right)\right] \right]
$$

\begin{align*}
&\quad - d_m \left[B\left(\xi_1 + \xi'_1 \Delta \alpha + \frac{1}{2} \xi''_1 \Delta \alpha^2 + \frac{1}{6} \xi'''_1 \Delta \alpha^3 \right) + \left(\frac{a_1}{r^2_1} + \frac{a_2}{r_1} + a_3\right)\right] \\
&\quad + \left[B\left(\xi_l + \xi'_l \Delta \alpha + \frac{1}{2} \xi''_l \Delta \alpha^2 + \frac{1}{6} \xi'''_l \Delta \alpha^3 \right) + \left(\frac{a_1}{r^2_l} + \frac{a_2}{r_l} + a_3\right)\right] \\
&\quad \times \left[B\left(\xi_m + \xi'_m \Delta \alpha + \frac{1}{2} \xi''_m \Delta \alpha^2 + \frac{1}{6} \xi'''_m \Delta \alpha^3 \right) + \left(\frac{a_1}{r^2_m} + \frac{a_2}{r_m} + a_3\right)\right].
\end{align*}

(A2)

Just as in §4, we may factor a quadratic in $B$ out of equation (A2). We have

$$
\mathcal{L}(B) = \exp\left(-aB^2 + bB + c\right),
$$

(A3)
and analytically performing the marginalization integral with a uniform prior in $B$, we find the marginalized likelihood as

$$L_{\text{marg}} = \int L(B) dB = \sqrt{\frac{2}{\pi}} \exp \left( \frac{b^2}{4a} + c \right).$$  \hspace{1cm} (A4)

The coefficients are

$$a = \frac{1}{2} P^{1m} \left[ \xi_m \xi_1 + \xi_m \xi_1' \Delta \alpha + \frac{1}{2} \xi_m \xi_1'' \Delta \alpha^2 + \xi_m \xi_1' \Delta \alpha + \xi_m \xi_1'' \Delta \alpha^2 \right.
+ \frac{1}{2} \xi_m \xi_1'' \Delta^3 + \frac{1}{2} \xi_m \xi_1' \Delta^2 + \frac{1}{2} \xi_m \xi_1'' \Delta^3 + \frac{1}{2} \xi_m \xi_1'' \Delta^3
+ \frac{1}{6} \xi_m \xi_m \Delta^3 \left. \right],$$

$$b = -\frac{1}{2} P^{1m} \left[ -d_l \left( \xi_m + \xi_m' \Delta \alpha + \frac{1}{2} \xi_m'' \Delta \alpha^2 + \frac{1}{2} \xi_m'' \Delta^3 \right)
- d_m \left( \xi_1 + \xi_1' \Delta \alpha + \frac{1}{2} \xi_1'' \Delta \alpha^2 + \frac{1}{2} \xi_1'' \Delta^3 \right)
+ \left( \xi_m + \xi_m' \Delta \alpha + \frac{1}{2} \xi_m'' \Delta \alpha^2 + \frac{1}{2} \xi_m'' \Delta^3 \right) \times \left( \frac{a_1}{r_l^2} + \frac{a_2}{r_l} + a_3 \right)
+ \left( \xi_1 + \xi_1' \Delta \alpha + \frac{1}{2} \xi_1'' \Delta \alpha^2 + \frac{1}{2} \xi_1'' \Delta^3 \right) \times \left( \frac{a_1}{r_m^2} + \frac{a_2}{r_m} + a_3 \right) \right].$$  \hspace{1cm} (A5)

Moving forward, we seek now to group $a$, $b$, and $c$ in terms of the additional polynomial terms. With this grouping, the integrals we need to perform next become clearer. Since $a$ does not depend on any of the additional polynomial terms, $a$ remains unchanged throughout the grouping process.

To deal with $b$, we define auxiliary variables as

$$\delta = -\frac{1}{2} P^{1m} \left[ -d_l \left( \xi_m + \xi_m' \Delta \alpha + \frac{1}{2} \xi_m'' \Delta \alpha^2 + \frac{1}{2} \xi_m'' \Delta^3 \right)
- d_m \left( \xi_1 + \xi_1' \Delta \alpha + \frac{1}{2} \xi_1'' \Delta \alpha^2 + \frac{1}{2} \xi_1'' \Delta^3 \right)
+ \left( \xi_m + \xi_m' \Delta \alpha + \frac{1}{2} \xi_m'' \Delta \alpha^2 + \frac{1}{2} \xi_m'' \Delta^3 \right) \times \left( \frac{a_1}{r_l^2} + \frac{a_2}{r_l} + a_3 \right)
+ \left( \xi_1 + \xi_1' \Delta \alpha + \frac{1}{2} \xi_1'' \Delta \alpha^2 + \frac{1}{2} \xi_1'' \Delta^3 \right) \times \left( \frac{a_1}{r_m^2} + \frac{a_2}{r_m} + a_3 \right) \right].$$

$$\gamma_{a1} = -\frac{1}{2} P^{1m} \left[ \frac{1}{r_l^2} \left( \xi_m + \xi_m' \Delta \alpha + \frac{1}{2} \xi_m'' \Delta \alpha^2 + \frac{1}{2} \xi_m'' \Delta^3 \right)
+ \frac{1}{r_m^2} \left( \xi_1 + \xi_1' \Delta \alpha + \frac{1}{2} \xi_1'' \Delta \alpha^2 + \frac{1}{2} \xi_1'' \Delta^3 \right) \right].$$

$$\gamma_{a2} = -\frac{1}{2} P^{1m} \left[ \frac{1}{r_l^2} \left( \xi_m + \xi_m' \Delta \alpha + \frac{1}{2} \xi_m'' \Delta \alpha^2 + \frac{1}{2} \xi_m'' \Delta^3 \right)
+ \frac{1}{r_m^2} \left( \xi_1 + \xi_1' \Delta \alpha + \frac{1}{2} \xi_1'' \Delta \alpha^2 + \frac{1}{2} \xi_1'' \Delta^3 \right) \right].$$

$$\gamma_{a3} = -\frac{1}{2} P^{1m} \left( \xi_m + \xi_m' \Delta \alpha + \frac{1}{2} \xi_m'' \Delta \alpha^2 + \frac{1}{2} \xi_m'' \Delta^3 \right)
+ \left( \xi_1 + \xi_1' \Delta \alpha + \frac{1}{2} \xi_1'' \Delta \alpha^2 + \frac{1}{2} \xi_1'' \Delta^3 \right) \right],$$  \hspace{1cm} (A6)

and find that

$$b = \gamma_{a1} a_1 + \gamma_{a2} a_2 + \gamma_{a3} a_3 + \delta.$$  \hspace{1cm} (A7)

The coefficient $c$ has cross-terms with respect to the additional polynomial terms, so grouping in terms of the individual polynomial terms only is not possible. We hence require cross-term coefficients as well; we find

$$c = \frac{1}{2} P^{1m} \left[ \frac{a_1}{r_l^2} + \frac{a_2}{r_l} + \frac{a_3}{r_l} + \frac{d_l}{r_l} \right] a_1
+ \left( \frac{a_1}{r_m^2} + \frac{a_2}{r_m} + \frac{a_3}{r_m} + \frac{d_l}{r_m} \right) a_2
- (d_m + d_l) a_3 + d_d a_3.$$  \hspace{1cm} (A8)

We now rewrite equation (A4) using matrices to enable us to integrate it via the solution for the $N$-dimensional Gaussian integral with a linear term:

$$\int \exp \left( -\frac{1}{2} \sum_{i,j=1}^{N} A_{ij} x_i x_j + \sum_{i=1}^{N} B_i x_i \right) d^N x = \frac{\sqrt{2\pi}^N}{\det A} \exp \left( \frac{1}{2} \tilde{B} A^{-1} \tilde{B} \right).$$  \hspace{1cm} (A9)

With this formula, equation (A9), in mind, the matrix $A$ and the vector $	ilde{B}$ are defined as:

$$A = \left[ \begin{array}{cccc}
\frac{\gamma_{a1}^2}{4a} + \frac{1}{r_m r_l^2} & \frac{2 \gamma_{a1} \gamma_{a2}}{4a} + \frac{1}{r_m r_l} & \frac{2 \gamma_{a1} \gamma_{a3}}{4a} + \frac{1}{r_m} + \frac{1}{r_l^2} \\
\frac{2 \gamma_{a1} \gamma_{a2}}{4a} & \frac{\gamma_{a2}^2}{4a} + \frac{1}{r_m r_l^2} & \frac{2 \gamma_{a2} \gamma_{a3}}{4a} + \frac{1}{r_m} + \frac{1}{r_l^2} \\
\frac{2 \gamma_{a1} \gamma_{a3}}{4a} & \frac{2 \gamma_{a2} \gamma_{a3}}{4a} + \frac{1}{r_m} + \frac{1}{r_l^2} & \frac{\gamma_{a3}^2}{4a} + 1
\end{array} \right],$$

and

$$\tilde{B} = \left[ \frac{2 \gamma_{a1} \delta}{4a} - \left( \frac{d_l}{r_m} + \frac{d_m}{r_l} \right) \frac{\gamma_{a2} \delta}{4a} - \left( \frac{d_l}{r_m} + \frac{d_m}{r_l} \right) \frac{\gamma_{a3} \delta}{4a} - (d_l + d_m) \right].$$

In our case $N = 3$ and equation (A9) gives the solution to the marginalization over each of the additional polynomial terms. We leave the final step, extremizing the resultant equation to find the optimal $\alpha$, for future work, but it is a straightforward matter of computing all of the required derivatives and solving the resulting quadratic in $\Delta \alpha$.

REFERENCES

Alam S., et al., 2017, MNRAS, 470, 2617
Alam S., et al., 2021, Phys. Rev. D, 103, 083533
Alcock C., Paczyński B., 1979, Nature, 281, 358
Anderson L., et al., 2012, MNRAS, 427, 3435
Anderson L., et al., 2014, MNRAS, 441, 24
Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, Phys. Rev. D, 86, 083540
Bashinsky S., Bertschinger E., 2001, Phys. Rev. Lett., 87, 081301
Bashinsky S., Bertschinger E., 2002, Phys. Rev. D, 65, 123008
Beutler F., et al., 2011, MNRAS, 416, 3017
Beutler F., Seljak U., Vlah Z., 2017, MNRAS, 470, 2723
Blake C., Glazebrook K., 2003, ApJ, 594, 665
Blake C., et al., 2011, MNRAS, 418, 1707
Blas D., Lesgourgues J., Trim T., 2011, Journal of Cosmology and Astroparticle Physics, 2011, 034–034
Blazek J. A., McEwen J. E., Hirata C. M., 2016, Phys. Rev. Lett., 116, 121303
Bond J. R., Efstathiou G., 1984, ApJ, 285, L45
