Application of Bayesian Methods in Cosmological Data Analysis: Parameter Constraint Forecasts for Stage-IV Surveys and Bayesian Large-Scale Structure Inference

by

Byeonghee Yu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Uroš Seljak, Chair
Professor Martin White
Professor Mariska Kriek

Fall 2022

Application of Bayesian Methods in Cosmological Data Analysis: Parameter Constraint Forecasts for Stage-IV Surveys and Bayesian Large-Scale Structure Inference

Copyright 2022
by
Byeonghee Yu

Abstract

Application of Bayesian Methods in Cosmological Data Analysis: Parameter Constraint Forecasts for Stage-IV Surveys and Bayesian Large-Scale Structure Inference

by

Byeonghee Yu

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Uroš Seljak, Chair

The application of Bayesian methodology in cosmological data analysis has gained enormous popularity, as the Bayesian interpretation of statistics is particularly appealing to the field of cosmology in which its subject, the Universe, is unique. In the coming decade, unprecedented size of data observed from forthcoming Stage-IV experiments - e.g. galaxy surveys such as DESI, Euclid, Roman, and LSST and CMB surveys such as SO and CMB-S4 - will call for the development of more advanced statistical analysis tools, and the Bayesian framework is expected to provide a key to decoding information hidden in the dataset. This will enable us to unlock the fundamental mysteries of the Universe, which include the nature of dark matter and energy, the neutrino mass scale, and inflationary physics.

Within a Bayesian framework, this thesis develops numerical and statistical tools in preparation for Stage-IV cosmological surveys. First, we forecast the constraining power of combining LSST clustering and CMB-S4 lensing; we find that the constraint on the neutrino mass sum of 25meV can be achieved without optical depth information, and its constraint on the dark energy equation of state parameter is comparable to the LSST tomographic cosmic shear forecast. In the remainder of this thesis, we build an efficient, reliable analysis pipeline for growth of structure measurements from large-scale structure dataset, which can be useful for upcoming galaxy redshift surveys. This includes: hybrid covariance matrix generated by integrating the analytic disconnected part and the data-driven connected part, optimization-based numerical method for posterior inference, and the use of the halo perturbation theory model to provide RSD measurements from the power spectrum multipoles of SDSS-III BOSS DR12 galaxies. With the pipeline developed in this thesis, we find a tight constraint on $f\sigma_8$ corresponding to $S_8 = 0.821 \pm 0.037$ or an overall amplitude error of 4% at $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, within 0.3 sigma of Planck's $S_8 = 0.832 \pm 0.013$. We also show that on smaller scales ($k_{\rm max} = 0.4\ h{\rm Mpc}^{-1}$) the constraint improves considerably to an overall 2.7% amplitude error (with $S_8 = 0.786 \pm 0.021$), but there is some evidence of model misspecification. Such RSD measurements provide

one of the most powerful cosmological probes by testing dark energy and different gravity models. Finally, we discuss the fundamental plane effect, which is claimed to be an important systematics of RSD analyses, and show that its impact on growth of structure constraints is insignificant.

This thesis is dedicated to my Mom, Oghyeon Lee.

Your unconditional love and support are behind all my accomplishments.
I love you so much.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest appreciation and thanks to my Ph.D. advisor, Uroš Seljak, for his continuous support and guidance over the past six years. Uroš has been a great mentor, encouraging my research even during difficult times. I have benefited a lot from his great intuition, passion, and expertise in different fields such as cosmology, statistics, and machine learning. My experience in his group also opened the door to many different research opportunities. I was able to explore a wide range of research topics, particularly in Bayesian statistics and machine learning, and this helped me start a new career as a machine learning engineer.

I would also like to thank Blake Sherwin and Simone Ferraro for their guidance in my early research projects. Blake kindly mentored me during my first year and introduced many research skills which were very useful throughout my Ph.D. studies. Simone always welcomed me whenever I had questions and provided insightful discussions about research projects.

I am also thankful to my research collaborators: Yin Li, Sukhdeep Singh, Colin Hill, Elena Massara, Marcel Schmittfull, Lloyd Knox, and Yu Feng. And I am greatful to all former and current BCCP members, including Emanuele Castorina, Emmanuel Schaan, Michael Wilson, Pat McDonald, Adrian Bayer, Stephen Chen, Nick Hand, Benjamin Horowitz, Chirag Modi, James Sullivan, and Biwei Dai.

I would also like to thank my committee members - Martin White, Mariska Kriek, and William Holzapfel - for their commitment to serve on my committee.

Also, thanks to Neha, Best, Sam, Doyeon, Jihwan, Kyeongsik, Yangjin, Dongwon, and all my Berkeley friends. You guys made my grad school memories truly special.

And thanks to KyeongA & Garett, John & Jinhyo, Young Eun, Seokjae, Kootak & Chanhee, and all other friends at OAKCC. Hanging out with you all meant so much to me.

Special thanks to Hanna. Meeting you is the luckiest thing that has ever happened to me.

And I wouldn't be here without the love and support of my family.

Dad, thank you for always believing in me and encouraging my dreams. I was able to study in the States thanks to your help and support.

Mom, I would not have reached this far without you. You really are the person behind all my accomplishments. Thank you for being my mom.

Chapter 1

Introduction

In the coming decade and beyond, the field of cosmology is expected to enter a new era, driven by unprecedented size of data collected from forthcoming Stage-IV ground and space-based telescopes; we will finally be in the "Golden Age," as David Schramm envisioned earlier. Few decades ago, the Cosmic Background Explorer (COBE) Differential Microwave Radiometer (DMR) detected the intrinsic anisotropy of the Cosmic Microwave Background (CMB), and since then cosmology has evolved into a data-driven science. With more sensitive cameras surveying wider and deeper volumes of the Universe, current Stage-III experiments achieved some major progress, further validating the consensus view that we live in a flat, expanding universe mostly consisting of dark energy, cold dark matter (CDM), and ordinary baryonic matter. Much more powerful dataset will be provided by forthcoming Stage-IV galaxy surveys - such as the Dark Energy Spectroscopic Instrument (DESI) [72], the ESA Euclid satellite mission [158], the Roman Space Telescope [76], and the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) [6] and Stage-IV CMB surveys - such as the Simons Observatory (SO) [8] and CMB-S4 [4]. Observations from such next-generation surveys will enable us to investigate the mysteries of the nature of dark matter, the equation of state of dark energy, the neutrino mass scale, the physics of inflation, and much more.

Therefore, it is crucial to develop analysis tools which extract information encoded in the large dataset observed from such cosmological probes. A handful set of parameters describes the evolution of the Universe in the standard cosmological model, and we use statistical methods to infer their values from the data. However, we have only one Universe, and consequently the frequentist interpretation of statistics is not applicable, as we often cannot repeat the same observation or generate multiple Universes. Alternatively, we choose a Bayesian approach, where model parameters are described probabilistically with the data fixed, and such characteristics make Bayesian statistics suitable for cosmological data analysis. This thesis provides numerical and statistical tools for analysis of large astronomical dataset within a Bayesian framework, in preparation for future cosmological surveys.

The remainder of this thesis is organized as follows. In Chapter 2 and 3, we forecast the constraining power of combining Stage-IV galaxy and CMB experiments using the Fisher information matrix. Chapter 4 develops an analytic method to model the disconnected part

of the covariance matrix for large-scale structure (LSS) data analysis. In Chapter 5, we design a new optimization-based method for parameter inference called EL_2O, which is as accurate as MCMC but its computational cost is only a fraction of that of MCMC. Chapter 6 presents growth of structure measurements from the SDSS-III BOSS DR12 galaxies using the halo perturbation theory model and Bayesian numerical methods for parameter inference. In Chapter 7, we introduce an important systematic for RSD analyses, fundamental plane effects, and discuss its impact on growth of structure measurements presented in Chapter 6.

1.1 The Standard Cosmological Model - ΛCDM

With observations from different probes, such as the CMB, LSS, and Type Ia supernovae (SNe), cosmologists have found good consistency with the standard cosmological model, although there still remains a number of challenges. According to this model, the matter is mostly non-baryonic, cold, collisionless, and dissipationless (hence labeled as CDM, Cold Dark Matter), with the dark energy (often denoted by cosmological constant Λ) accounting for roughly 70% of the energy of the Universe and causing the accelerated expansion. Ordinary, baryonic matter comprises only 5% of the energy budget, and the radiation (i.e. photon and relativistic neutrino) energy density is many orders of magnitude lower. Hence, the standard cosmological model is known as ΛCDM model.

ΛCDM postulates that the Universe was initially extremely hot and dense and then has expanded and cooled down to the current state. Photon was tightly coupled to matter, electrons and baryons, and this nearly homogeneous mixture was in a hot plasma state and made up the early Universe. Its initial conditions are thought to be set during the epoch of exponential expansion, known as inflation, which happened almost immediately after the Big Bang. Such expansion inflated quantum fluctuations on sub-microscopic scales to macroscopic scales, and this seeded the growth of cosmic structure. Roughly few minutes after the Big Bang, it is predicted that helium, deuterium, and other heavier nuclei formed during the process called Big Bang nucleosynthesis, and predictions for their abundances in ΛCDM cosmology agrees well with the measurements.

At a redshift of $z \approx 3400$ (50,000 years after the Big Bang), radiation was no longer the dominant component of the energy budget, and the Universe started to become matter-dominated. At that period, Photon-electron interactions (via Thomson scattering) and electron-proton interactions (via Coulomb scattering) made photon and baryon tightly coupled as a fluid. After such transition, the Universe further expanded and cooled down to $T \approx 3000$K (380,000 years after the Big Bang), and protons and electrons bound together to form the neutral hydrogen atoms during the period of recombination. Subsequently, photons decouple from the matter and free-stream through the Universe, as the photon-electron scattering rate drops below the expansion rate at $z \approx 1100$. Such radiation, traveling from a spherical surface whose radius is the distance it has traveled since its last scattering from the matter, still reaches us, observed as the CMB. Before decoupling, the radiation pressure in the photon-baryon fluid balanced with gravity, resulting in acoustic oscillations in the fluid,

Parameter	Symbol	Value + 68% interval
Physical baryon density	$\Omega_b h^2$	0.02242 ± 0.00014
Physical CDM density	$\Omega_c h^2$	0.11933 ± 0.00091
Angular size of acoustic scale	$100\theta_{\rm MC}$	1.04101 ± 0.00029
Reionization optical depth	τ	0.0561 ± 0.0071
Primordial curvature fluctuation amplitude	$\ln(10^{10} A_s)$	3.047 ± 0.014
Scalar spectral index	n_s	0.9665 ± 0.0038
Hubble constant [km s^{-1} Mpc^{-1}]	H_0	67.66 ± 0.42
Present matter fluctuation amplitude	σ_8	0.8102 ± 0.0060

Table 1.1: The best-fit ΛCDM cosmological parameters and 68% confidence intervals from Planck CMB power spectra (TT+TE+EE+lowE), combined with CMB lensing reconstruction and BAO [11]. A set of six parameters in the top group are the base parameters, and other parameters - few are shown in the bottom group - can be derived from the model set by the base parameters.

and this pattern became frozen after decoupling, its imprint observed as baryon acoustic oscillations (BAO). The CMB blackbody radiation was at the temperature of ≈ 3000K when it formed, filling the Universe with red, uniform light. However, the CMB shifted to the infrared over the next few million years, and the Universe entered the era of the Dark Ages, as it became mostly not visible until the formation of the first stars, whose light reionized all atoms in the intergalactic medium. Overdense regions expand with the Hubble flow initially, then break away from it and collapses subsequently due to gravitational attraction. The picture of structure formation is hierarchical, bottom-up rather than top-down, with small overdensities collapsing first and merging into larger structures. Accordingly, CDM first forms microhalos and merge together to form larger halos, and their gravity attracts baryonic matter, which collapses to form stars and galaxies. More details about the large-scale structure formation are discussed in section 1.2. Finally, direct measurements of Type Ia supernovae, which serve as a standard candle measuring the cosmic expansion, suggest the accelerating expansion of the Universe, driven by the dark energy (or in the form of Λ) whose negative pressure opposes gravitation attractions. Roughly 10 billion years after the Big Bang, the Universe entered the dark energy-dominated era, but its physical origin and characteristics are largely unknown.

The standard ΛCDM model requires only six base parameters, such as: physical baryon density $\Omega_b h^2$, CDM density $\Omega_c h^2$, angular acoustic scale $100\theta_{\rm MC}$, reionization optical depth τ, the amplitude A_s primordial power spectrum of density perturbation, and scalar spectral index n_s. The model defined with such base parameters provides a good fit to the measured data, and this significantly increases the efficiency of parameter estimation [149].

FRW metric and Friedmann Equation

The modern cosmology is built upon the so-called cosmological principle: the Universe is homogeneous and isotropic on the largest sales, and in general relativity we develop the metric for such homogeneous and isotropic spacetime, called Friedmann-Robertson-Walker (FRW) metric:

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\Big[\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2)\Big]. \tag{1.1}$$

The coordinate system (r, θ, ϕ) denotes fixed spatial coordinates called comoving coordinates, and t is the proper time measured by the standard clock of a fundamental observer. The scale factor $a(t)$ describes the expansion in time, relating the comoving coordinates to physical coordinates. The constant k determines the curvature ($k > 0$ in a closed Universe, $k = 0$ in a flat Universe, and $k < 0$ in an open Universe).

The proper distance is the distance between two events in a frame where they occur simultaneously:

$$s(t) = a(t)\int_0^r \frac{\mathrm{d}r'}{\sqrt{1-kr'^2}} = a(t)\chi(r), \tag{1.2}$$

where $\chi(r)$ is the comoving distance defined as

$$\chi(r) = \begin{cases} \frac{1}{\sqrt{k}}\sin^{-1}(\sqrt{k}r) & \text{for } k > 0 \\ r & \text{for } k = 0 \\ \frac{1}{\sqrt{|k|}}\sinh^{-1}(\sqrt{|k|}r) & \text{for } k < 0. \end{cases} \tag{1.3}$$

Similarly, we define the conformal time $\tau(t)$ with respect to the proper time t:

$$\tau(t) = \int_0^t \frac{\mathrm{d}t'}{a(t')}. \tag{1.4}$$

Then, the FRW metric in equation 1.1 can be expressed as

$$\mathrm{d}s^2 = a^2(\tau)[d\tau^2 - d\chi^2 - r(\chi)(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)], \tag{1.5}$$

where $r(\chi)$ is the inverse of $\chi(r)$ in equation 1.3.

From the Einstein field equations, we then write the Friedmann equation which describes the evolution of the scale factor in terms of the energy density ρ and pressure p of a perfect fluid:

$$H^2(a) \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{1.6}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \tag{1.7}$$

where $H(a) \equiv \frac{\dot{a}}{a}$ is the Hubble parameter, and G is Newton's gravitational constant. We obtain the second Friedmann equation (1.7), or the acceleration equation, by differentiating the first equation and substituting the differential of the density with the expression from the fluid equation,

$$\dot{\rho} + 3H(\rho + p) = 0. \tag{1.8}$$

We define the equation of state parameter, $w = \frac{p}{\rho}$ and substitute this into the equation from the first law of thermodynamics, $\frac{\mathrm{d}\rho}{\mathrm{d}a} + 3\left(\frac{\rho+p}{a}\right) = 0$, yielding the equation $\rho \propto a^{-3(1+w)}$. Therefore, the equation of state parameter of each component determines how the corresponding energy density evolves with the scale factor. For instance, the pressure of the cosmological constant Λ is given as $p = -\rho$, implying $w = -1$.

The energy density required for $k = 0$ is defined as the critical density,

$$\rho_c(a) = \frac{3H^2(a)}{8\pi G}. \tag{1.9}$$

We define the ratio of the total energy density to the critical density as $\Omega(a) = \frac{\rho(a)}{\rho_c(a)}$, and the Friedmann equation can be rewritten as

$$\left(\frac{H(a)}{H_0}\right)^2 = \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \frac{\Omega_{k,0}}{a^2} + \frac{\Omega_{DE,0}}{a^{3(1+w)}}, \tag{1.10}$$

where H_0 is the Hubble parameter today, and $\Omega_{r,0}, \Omega_{m,0}, \Omega_{k,0}$ and $\Omega_{DE,0}$ refer to the present values of Ω for radiation, total matter, curvature, and dark energy, respectively.

1.2 Large-Scale Structure

In the standard cosmological model, on large scales massive filaments of galaxies are separated by voids in between and form a web-like structure called the cosmic web, or large-scale structure (LSS). Studying the LSS survey provides cosmologists valuable information about the origin and evolution of the Universe, particularly regarding primordial fluctuations, the dark matter distribution, the strength of gravity, and characteristics of dark energy.

Two-point statistics

Galaxies are considered as correlated if they are clustered, and we use the two-point correlation function to quantify the degree of their clustering. We first define the overdensity field as $\delta(\boldsymbol{x}) = \frac{n(\boldsymbol{x}) - \langle n \rangle}{\langle n \rangle}$, where $n(\boldsymbol{x})$ is the number density, and $\langle\,\rangle$ takes its ensemble average or expected value. Then, we can write the correlation function $\xi(r)$ as the excess probability of finding another galaxy at a separation r, relative to the probability expected for a random distribution

$$\xi(r) = \langle \delta(\boldsymbol{x}) \delta(\boldsymbol{x} + \boldsymbol{r}) \rangle. \tag{1.11}$$

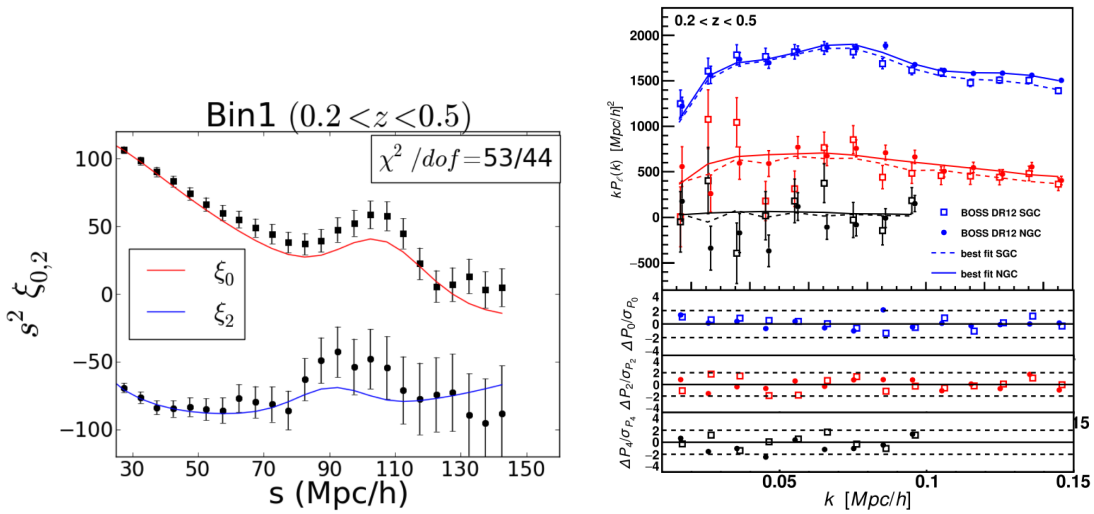


Figure 1.1: *Left*: Correlation function multipoles of the SDSS-III BOSS DR12 galaxy sample over the redshift range $0.2 < z < 0.5$, from [226]. *Right*: Power spectrum multipoles of the same galaxy sample, from [37]. Solid and dotted curves indicate best-fit theory models.

Note that $\xi(\boldsymbol{r}) = \xi(r)$ as we assume the Universe to be isotropic. The power spectrum $P(k)$ is the Fourier transform of $\xi(r)$ and can be written with respect to the Fourier transform of the overdensity field, $\delta(\boldsymbol{k})$, in the following way:

$$\langle \delta(\boldsymbol{k}) \delta^*(\boldsymbol{k}') \rangle = (2\pi)^3 \delta_D(\boldsymbol{k} - \boldsymbol{k}') P(\boldsymbol{k}), \tag{1.12}$$

where $\boldsymbol{k}$ is the wavenumber, and δ_D is the Dirac delta function. $P(\boldsymbol{k}) = P(k)$ due to the isotropy assumption. We can relate the correlation function to the power spectrum:

$$\xi(r) = \frac{1}{2\pi^2} \int \mathrm{d}k \; j_0(kr) \; k^2 P(k). \tag{1.13}$$

The power spectrum can also be expressed in terms of the multipole of order l:

$$P(k, \mu) = \sum_l P_l(k) \mathcal{L}_l(\mu), \tag{1.14}$$

$$P_l(k) = \frac{2l+1}{2} \int_{-1}^{1} \mathrm{d}\mu \; P(k, \mu) \mathcal{L}_l(\mu), \tag{1.15}$$

where $\mathcal{L}_l$ are Legendre polynomials of order l, and μ is the cosine of the angle between the wavenumber $\boldsymbol{k}$ and the line of sight. We can similarly define the correlation function multipoles. Figure 1.1 presents both correlation function multipoles (left panel) and power spectrum multipoles (right panel) of the galaxy distributions, measured by Sloan Digital Sky Survey (SDSS)-III Baryon Oscillation Spectroscopic Survey (BOSS), along with the best-fit

theory models. In Chapter 6, we use the power spectrum estimator, provided in the python software package `nbodykit` [104], to measure the galaxy power spectrum multipoles for the SDSS-III BOSS DR12 dataset. Then, we fit the halo perturbation theory-based model of [105] to the measured multipoles to constrain the cosmological parameters of our interest.

Redshift Space Distortions

As LSS surveys observe the redshifts of galaxies, not their true distances, the two-point statistics are measured via their locations inferred from the observed angular positions and redshifts. Hence, such conversion between different coordinate systems complicates the analysis of galaxy clustering. First, the assumption of a fiducial cosmological model is needed to convert the redshift and the angular position to the distance along the line of sight and the angular diameter distance, respectively. The mismatch between the fiducial and true cosmology results in a geometric distortion in the galaxy statistics, and this effect is known as the Alcock-Paczynski (AP) effect. However, there exists a more dominant effect on the two-point statistics, called Redshift Space Distortion (RSD), caused by the peculiar velocity of galaxies, which shift the measured redshifts and affect the line-of-sight position of galaxies accordingly. Therefore, RSD breaks the rotational symmetry of the galaxy correlations and thereby makes the galaxy clustering signal anisotropic.

On large scales where linear perturbation theory can be applied, coherent infall of objects toward the overdensity leads to a squashing effect, known as the Kaiser effect, along the line of sight in the redshift space. We model such enhancement of the density field on large scales as $\delta^S = \delta^R(1 + f\mu^2)$, where $f = \mathrm{dln}D/\mathrm{dln}a$ is the logarithmic rate of the linear growth factor $D(a)$. Note that the superscript S and R denotes the redshift-space and real-space measurements, respectively. Accordingly, we obtain the power spectrum as $P^S(k,\mu) = (1+f\mu^2)^2 \cdot P^R(k)$. For biased tracers like galaxies, we write the number density as $\delta_{\rm gal} = b\delta_{\rm m}$ and the power spectrum as $P^S_{\rm gal}(k,\mu) = (b + f\mu^2)^2 \cdot P^R_{\rm m}(k) = b^2(1 + \beta\mu^2)^2 \cdot P^R_{\rm m}(k)$, where b is the bias of the tracer which characterizes the galaxy clustering relative to the matter field. Constraints on the parameter $\beta = f/b$ provide tests of general relativity and evidence of the standard cosmological model. As the term $(1 + \beta\mu^2)^2$ involves terms only up to μ^4, expanding the power spectrum with Legendre polynomial (equation 1.15) leads to non-vanishing monopole ($l = 0$), quadrupole ($l = 2$), and hexadecapole ($l = 4$) terms:

$$P_0(k) = \Big(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\Big)b^2 P_{\rm m}(k), \tag{1.16}$$

$$P_2(k) = \Big(\frac{4}{3}\beta + \frac{4}{7}\beta^2\Big)b^2 P_{\rm m}(k), \tag{1.17}$$

$$P_4(k) = \frac{8}{35}\beta^2 b^2 P_{\rm m}(k). \tag{1.18}$$

On small scales, the linear theory described above breaks down, as nonlinear effects which cannot be described by perturbation theory add complications to the analytic modeling. Satellite galaxies within halos have virialized, random motion, stretching the density field

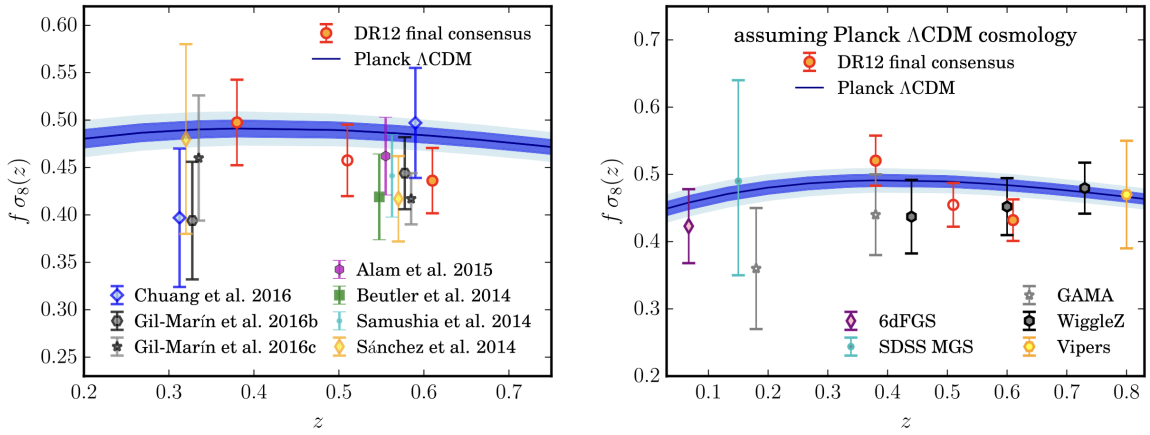


Figure 1.2: A compilation of $f(z)\sigma_8(z)$ measurements from the SDSS-III BOSS DR12 consensus paper [15]. *Left*: Consensus results (9.3% and 8.0% $f\sigma_8$ constraints on low-redshift ($0.2 < z < 0.5$) and high-redshift ($0.5 < z < 0.75$) galaxies), along with DR11 measurements and Planck ΛCDM prediction. *Right*: Comparison with the measurements form other surveys, such as 2dfGRS, 5dFGS, GAMA, WiggleZ, and Vipers.

and thereby reducing the clustering amplitude along the line of sight. This nonlinear effect is known as the Finger-of-God (FoG) effect. In previous literature, a damping factor $G(k\mu;\sigma_v)$, where σ_v is the velocity dispersion of the galaxy sample, is introduced. Typically this damping factor takes a form of a Gaussian or Lorentizan,

$$G(k\mu;\sigma_v) = \begin{cases} e^{-k^2\mu^2\sigma_v^2/2} & \text{(Gaussian)} \\ (1+k^2\mu^2\sigma_v^2/2)^{-2} & \text{(Lorentzian).} \end{cases} \tag{1.19}$$

Figure 6 in [193] shows how the predictions for different FoG damping functions, particularly the Lorentzian model, agree well with N-body results. However, we note that the perturbation theory-based approaches cannot fully capture the nonlinear effects, on small scales in particular, and therefore it may be necessary to use simulation-based modeling to account for small-scale information. The galaxy power spectrum model assumed in Chapter 6 uses N-body simulations to calibrate some of the key terms in the model.

Notwithstanding such difficulties in modeling, RSD has been widely used to constrain the growth rate of structure, which in turn test dark energy and different gravity models. In earlier studies, the ratio of the quadrupole to monopole,

$$\frac{P_2(k)}{P_0(k)} = \frac{4\beta/3 + 4\beta^2/7}{1 + 2\beta/3 + \beta^2/5}, \tag{1.20}$$

has been used as a estimator for β (and the growth rate f with known b), and this ratio is expected to be scale-independent as long as the linear theory is valid. However, this assumes

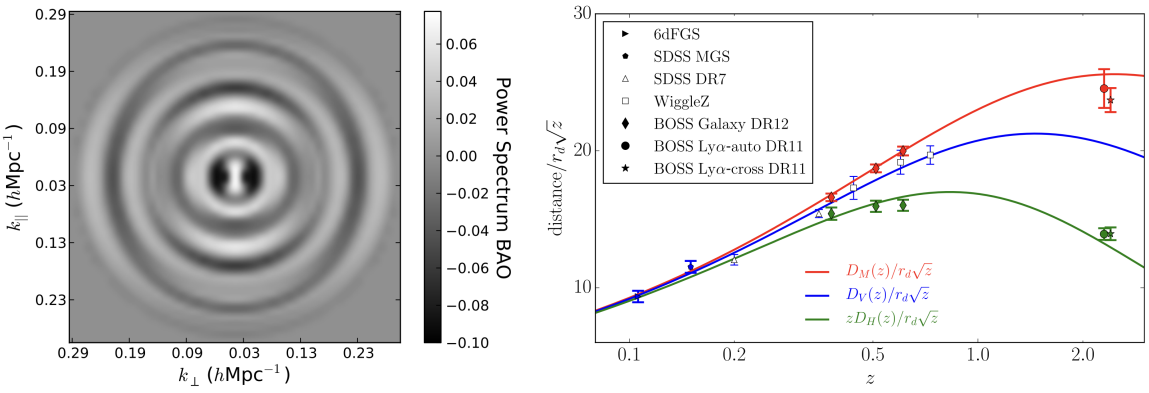


Figure 1.3: *Left*: BAO feature in the measured BOSS power spectrum. *Right*: Constraints on the distance from the BOSS galaxies and other surveys, along with Planck ΛCDM predictions. Both figures are from [15].

a linear bias and linear perturbation theory, which are poor approximations for RSD. Recent RSD analyses, therefore, has focused on using either simulation-based [157, 148, 285, 284] or perturbation theory-based [131, 68, 291, 58] approaches to model nonlinear corrections, measuring the growth rate via the parameter $f(z)\sigma_8(z)$. Figure 1.2 presents the SDSS-III BOSS DR12 final consensus results on the galaxies over the redshift range $0.2 < z < 0.75$. In this thesis, we provide RSD measurements using the halo perturbation theory model using the BOSS galaxy power spectrum and show that our results offer tight constraints on $f\sigma_8$, in a good agreement with both the DR12 consensus results and Planck amplitude.

Baryon Acoustic Oscillation

Baryon acoustic oscillations, or BAO, are the relics imprinted on the CMB anisotropy and the galaxy clustering, caused by sound waves from the pre-decoupling era [80]. Before decoupling, the overdensity pulled the matter together, thereby heating it up and creating an outward pressure. Gravitational attraction and radiation pressure balanced with each other, creating acoustic oscillations in the photon-baryon fluid. After decoupling, such features are imprinted on large-scale structures as oscillations in the power spectrum, as shown in the left panel of Figure 1.3. As a standard ruler, the BAO can be used to probe the Universe's expansion history and the dark energy properties. Back in 2005, [82] detected the BAO peak in the correlation function measurements of the SDSS Luminous Red Galaxies (LRGs) and provided a 4% meaurement of the ratio of the distances between $z = 0.35$ and $z = 1100$. More recently, [15] used measurements of the BOSS galaxies and reported constraints on the distance scale with a percent-level precision. The right panel of Figure 1.3 presents good agreements between different BAO measurements and the Planck ΛCDM predictions. In the next decade, the DESI is expected to achieve a subpercent-level precision on the BAO distance scale measurements. In Chapter 2 and 3, we use the forecasted uncertainties on the

distance-redshift relation from the DESI, which significantly improves both neutrino mass and dark energy equation of state parameter forecasts.

1.3 Statistical Data Analysis Tools in Cosmology

This section introduces different statistical techniques for data analysis in cosmology, mostly Bayesian methods used in this thesis. This includes the basics of Bayesian statistics, experiment design using the Fisher information matrix, and different methods - both sampling approach and variational approximations - for cosmological parameter inference.

Bayes' theorem and parameter inference

Bayesian statistics provides a subjective interpretation of probability which corresponds to the degree of belief in a hypothesis and can be updated with new observations. On the other hand, frequentist approach considers an objective measure of probability, which is directly related to the proportion of outcomes in a large number of repeated experiments. As we have only one Universe, cosmologists naturally choose the Bayesian method to extract useful information encoded in the large astronomical dataset, finding the best-fit parameters and their uncertainties given the cosmological model. In this section, we briefly discuss the procedure of parameter inference within a Bayesian framework.

First, we assume a model M which depends on a set of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_N)$, where N is the dimension of the parameter space. The Bayes' Theorem gives the posterior probability for $\boldsymbol{\theta}$, which shows the degree of belief about the value of $\boldsymbol{\theta}$ given J number of observed data $\boldsymbol{D} = (d_1, d_2, ..., d_J)$:

$$P(\boldsymbol{\theta}|\boldsymbol{D}, M) = \frac{P(\boldsymbol{D}|\boldsymbol{\theta}, M)P(\boldsymbol{\theta}|M)}{P(\boldsymbol{D}|M)}. \tag{1.21}$$

$P(\boldsymbol{D}|\boldsymbol{\theta}, M)$ is the likelihood (henceforth denoted as $L(\boldsymbol{D}|\boldsymbol{\theta})$), the probability of making an observation of the data, given the values of the parameters. $P(\boldsymbol{\theta}|M)$ is the prior probability distribution, which show the degree of belief about the values of $\boldsymbol{\theta}$, prior to providing the data. $P(\boldsymbol{D}|M) = \int P(\boldsymbol{D}|\boldsymbol{\theta}, M)P(\boldsymbol{\theta}|M)$ is a normalizing constant called the evidence or marginal, which normalizes the posterior and ensures that it behaves as a probability. Evidence is often ignored in parameter inference (reducing the Bayes' Theorem to: posterior = likelihood $\cdot$ prior) but is important for Bayesian model selection.

First, assume $N = 1$ for the sake of simplicity. With the known posterior distribution, we can compute the expectation value and uncertainty of θ. Often we use a point estimate called maximum a posteriori (MAP) to choose the value of θ which maximizes the posterior probability:

$$\theta_{\mathrm{MAP}} = \arg\max_{\theta} P(\theta|\boldsymbol{D}, M) = \arg\max_{\theta} \prod_{i=1}^{J} P(\theta|d_i, M). \tag{1.22}$$

$\theta_{\rm MAP}$, the MAP estimate of θ, finds the peak of the posterior probability density function. If we simply assume an uniform prior (i.e. $P(\theta|M) = \text{constant}$), the posterior probability $P(\theta|\boldsymbol{D}, M) \propto L(\boldsymbol{D}|\theta)$, and the MAP estimate is equivalent to the maximum likelihood (MLE) estimate, i.e. $\theta_{\rm MAP} = \theta_{\rm MLE} = \arg\max_\theta L(\boldsymbol{D}|\theta)$.

The MAP or MLE is only a point estimator, and the most widely used method to obtain an error estimate is the Laplace approximation: approximating the posterior with a Gaussian around the MAP. Under the Laplace approximation, the uncertainty of θ is given as:

$$\frac{1}{\sigma_\theta^2} = -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\ln P(\theta|\boldsymbol{D}, M)\Big|_{\theta=\theta_{\rm MAP}} \propto -\sum_{i=1}^{J}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\mathcal{L}(d_i|\theta)\Big|_{\theta=\theta_{\rm MLE}}, \tag{1.23}$$

where we define the log-likelihood as $\mathcal{L} = \ln L$.

More generally (for $N > 1$), assuming that the likelihood can be approximated with a single-peaked, multi-variate Gaussian, we can Taylor expand the log-likelihood around the MLE estimate in the following way:

$$\begin{aligned}\mathcal{L}(\boldsymbol{D}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{D}|\boldsymbol{\theta}_{\rm MLE}) &+ \sum_i \frac{\partial\mathcal{L}}{\partial\theta_i}\Big|_{\theta_i=\theta_{{\rm MLE},i}}(\theta_i - \theta_{{\rm MLE},i}) \\ &+ \frac{1}{2}\sum_{ij}(\theta_i - \theta_{{\rm MLE},i})\frac{\partial^2\mathcal{L}}{\partial\theta_i\partial\theta_j}\Big|_{\theta_i=\theta_{{\rm MLE},i}}(\theta_j - \theta_{{\rm MLE},j}) + ...\end{aligned} \tag{1.24}$$

The first term is a constant, and the second term vanishes as the MLE finds the peak of the likelihood distribution. Therefore, we can express the likelihood as:

$$\begin{aligned}L(\boldsymbol{D}|\boldsymbol{\theta}) &= L(\boldsymbol{D}|\boldsymbol{\theta}_{\rm MLE})\cdot\exp\Big[-\frac{1}{2}\sum_{ij}(\theta_i - \theta_{{\rm MLE},i})\frac{\partial^2\mathcal{L}}{\partial\theta_i\partial\theta_j}\Big|_{\theta_i=\theta_{{\rm MLE},i}}(\theta_j - \theta_{{\rm MLE},j})\Big] \\ &= L(\boldsymbol{D}|\boldsymbol{\theta}_{\rm MLE})\cdot\exp\Big[-\frac{1}{2}\sum_{ij}(\theta_i - \theta_{{\rm MLE},i})H_{ij}(\theta_j - \theta_{{\rm MLE},j})\Big],\end{aligned} \tag{1.25}$$

where H_{ij} is the Hessian matrix, which shows the correlation between the estimates of θ_i and θ_j.

Finally, with the model which depends on N different parameters, only a few may be of our interest, and others can be regarded as nuisance parameters. Say we are only interested in the first I number of parameters $(\theta_1, ..., \theta_I)$, and then we can marginalize over other unwanted parameters in the following way:

$$P(\theta_1, ..., \theta_I|\boldsymbol{D}, M) = \int \mathrm{d}\theta_{I+1}...\mathrm{d}\theta_N P(\boldsymbol{\theta}|\boldsymbol{D}, M). \tag{1.26}$$

With this, nuisance parameters are free to hold any values while constraining the parameters of our interest.

Designing an experiment

The Fisher information matrix is defined as an ensemble average of the Hessian matrix:

$$F_{ij} = \langle H_{ij} \rangle = -\Big\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j} \Big\rangle. \tag{1.27}$$

Here we replace the likelihood from the actual data with its ensemble average, which makes this method useful for predictions of future analyses. The inverse of the Fisher matrix is the covariance matrix, from which we can obtain the 1σ uncertainties of the parameters of our interest. The parameter uncertainty corresponds to $\sigma_{\theta_i} \geq \sqrt{(F^{-1})_{ii}}$, with other parameters marginalized over. If we have perfect knowledge of some of the parameters, we can fix them by removing their corresponding rows and columns from the Fisher matrix. Or we can add the prior information of the parameter, $\theta_{\mathrm{prior},i} \pm \sigma_{\theta_{\mathrm{prior},i}}$, from earlier experiments by adding the squared inverse of $\sigma_{\theta_{\mathrm{prior},i}}$ to the ith diagonal element of the Fisher matrix.

Assume that we have a model which depends on $\boldsymbol{\theta} = (\theta_1, ..., \theta_N)$, and we observe the data $\boldsymbol{D} = (d_1, ..., d_J)$ with noise $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_J)$, measured at points $\boldsymbol{x} = (x_1, ..., x_J)$. We can Taylor expand the log-likelihood around the fiducial model and compute the Fisher matrix at the fiducial parameters $\boldsymbol{\theta}_{\mathrm{fid}}$. Then, the resulting Fisher matrix is as follows:

$$F_{ij} = \sum_{k=1}^{J} \frac{1}{\sigma_k^2} \frac{\partial d_k}{\partial\theta_i}\Big|_{\theta_i=\theta_{\mathrm{fid},i}} \frac{\partial d_k}{\partial\theta_j}\Big|_{\theta_i=\theta_{\mathrm{fid},j}}. \tag{1.28}$$

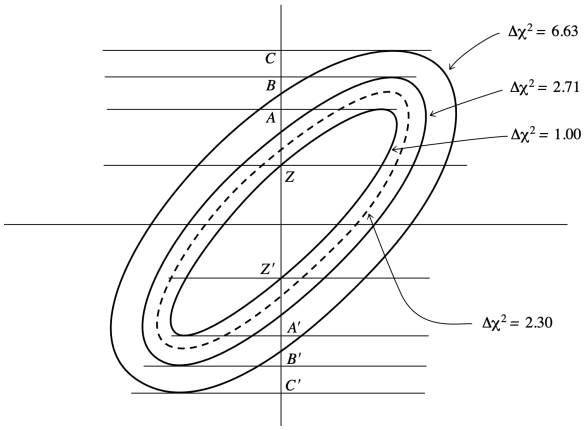
The advantage of this method is that actual data values are not needed, and we can instead compute the Fisher matrix element at the fiducial model before doing any experiments. Hence, we can use the Fisher matrix for experimental design, as it predicts the constraining power of future experiments from the expected noise levels.

Let the probability distribution be Gaussian, and we measure the mean $\mu = \langle \boldsymbol{d} \rangle$ and the covariance $\boldsymbol{C} = \langle (\boldsymbol{d} - \boldsymbol{\mu})^T (\boldsymbol{d} - \boldsymbol{\mu}) \rangle$. Then, the resulting Fisher matrix is given as [267, 114]:

$$F_{ij} = \frac{1}{2} \mathrm{Tr}\Big[\boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial\theta_i} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial\theta_j} + \boldsymbol{C}^{-1} \Big(\frac{\partial \boldsymbol{\mu}}{\partial\theta_i} \frac{\partial \boldsymbol{\mu}^T}{\partial\theta_j} + \frac{\partial \boldsymbol{\mu}}{\partial\theta_j} \frac{\partial \boldsymbol{\mu}^T}{\partial\theta_i} \Big) \Big]. \tag{1.29}$$

Hence, if the data distrubution is Gaussian, and the mean and covariance matrix are known, we can obtain the expected errors from the Fisher matrix before doing an actual experiment. [267] provides a simple example for obtaining the constraining power of a future CMB experiment, where the mean is zero, $\boldsymbol{\mu} = 0$, and the covariance $\boldsymbol{C}$ is diagonal. The observed data is the angular power spectrum of the CMB, C_l, measured from $l = 2$ to some cutoff l_{max} with the corresponding noise level σ_l. Assuming that the CMB fluctuations are Gaussian, we obtain the following Fisher matrix from equation 1.29:

$$F_{ij} = \sum_{l=2}^{l_{\mathrm{max}}} \Big(\frac{2l+1}{2} \Big) \frac{1}{\sigma_l^2} \Big(\frac{\partial C_l}{\partial\theta_i} \Big) \Big(\frac{\partial C_l}{\partial\theta_j} \Big). \tag{1.30}$$



$\Delta\chi^2$ as a Function of Confidence Level and Degrees of Freedom

	ν					
p	1	2	3	4	5	6
68.3%	1.00	2.30	3.53	4.72	5.89	7.04
90%	2.71	4.61	6.25	7.78	9.24	10.6
95.4%	4.00	6.17	8.02	9.70	11.3	12.8
99%	6.63	9.21	11.3	13.3	15.1	16.8
99.73%	9.00	11.8	14.2	16.3	18.2	20.1
99.99%	15.1	18.4	21.1	23.5	25.7	27.8

Figure 1.4: *Left*: Confidence regions for different values of $\Delta\chi^2$. The dashed ellipse represents a 1σ regon for two fitted paramers ($\nu = 2$), and 68.3%, 90%, and 99% ellipses for $\nu = 1$ project onto the intervals AA', BB', and CC', respectively. *Right*: A table containing the values of $\Delta\chi^2$ corresponding to different confidence levels and number of fitted parameters (ν). Both figure and table taken from [211].

In Chapter 2 and 3, we follow a similar procedure; assuming the Gaussian covariance matrix of the CMB lensing convergence and the LSST galaxy auto- and cross-power spectra, we calculate the Fisher matrix (equation 2.5) using the power spectrum and its derivatives obtained from the `camb` Boltzmann code [1].

Goodness of fit

Assuming the Gaussian likelihood distribution, equation 1.25 can be expressed in terms of chi-square χ^2, which can be defined as:

$$L(\boldsymbol{D}|\boldsymbol{\theta}) = L(\boldsymbol{D}|\boldsymbol{\theta}_{\rm MLE}) \cdot \exp(-\frac{1}{2}\chi^2), \text{ where } \chi^2 = \sum_{ij}(\theta_i - \theta_{{\rm MLE},i})H_{ij}(\theta_j - \theta_{{\rm MLE},j}). \quad (1.31)$$

Therefore, maximizing the likelihood L corresponds to minimizing the chi-square χ^2, and this procedure leads to finding the best-fit parameters of the model. However, this relation does not hold when the likelihood cannot be approximated with a Gaussian.

The best-fit parameters are a set of parameter values which gives the lowest χ^2, $\min(\chi^2)$. With this quantity, we can calculate the uncertainties of our fits by comparing the chi-square values of different parameter values to the best-fit. We define the quantity for evaluating the goodness of fit as $\Delta\chi^2 = \chi^2 - \min(\chi^2)$, and the right panel of Figure 1.4 provides a table containing the values of $\Delta\chi^2$ corresponding to 68.3%, 95.4%, and 99.73% (1σ, 2σ, and 3σ, respectively) confidence regions. The relation between $\Delta\chi^2$ and confidence regions depends on the number of fitted parameters, ν. For example, if we fit two parameters ($\nu = 2$), then the 1σ confidence region correspond to $\Delta\chi^2 = 2.30$, shown as the dashed ellipse in the left

panel of Figure 1.4. But if we take the confidence region for one of these fitted paramters, i.e. reducing ν to 1, then 1σ interval corresponds to $\Delta\chi^2 = 1$, between the lines A and A'.

Numerical methods for parameter inference

Parameter inference in cosmology often involves a large number of parameters, which makes analytic evaluation of the posterior difficult to attain. Moreover, if the prior and likelihood are not in the conjugate form, the analytic approach is intractable. Such difficulties lead to the usage of sampling methods for cosmological parameter inference.

The simplest numerical method is a grid-based approach, which sets up a grid of parameter values and explores the likelihood space. However, testing over a full grid of parameters is very time-consuming, as we also have to evaluate the model in regions where likelihood is very low. This leads to the development of Monte Carlo Markov Chain (MCMC) method, which does random walk in the parameter space and draws samples from more likely regions. A sequence of such samples constitutes the MCMC chain, and its density slowly converges to the target posterior distribution. As its name suggests, MCMC sampling uses a Markov Chain, a series of random variables for which the $(N+1)$th element depends only on the Nth element, and consequently this method is a sequential process in which the distribution of the current draws depends on the previous sample. The number of iterations required for convergence scales generally with the number of parameters, but its computational cost is much less than that of full grid searches. Drawing K number of samples in the MCMC chain, we can approximate the target posterior distribution as $P(\boldsymbol{\theta}|\boldsymbol{D}, M) \simeq \frac{1}{K}\sum_i^K \delta^{(N)}(\boldsymbol{\theta} - \boldsymbol{\theta}_i)$. Samples in the MCMC chain provides important statistics, such as the posterior mean and variance.

The Metropolis-Hastings algorithm is the simplest random walk method with a rule for accepting or rejecting drawn samples to reach the target posterior. First, we choose a proposal distribution $q(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)$ from which we draw a new sample $\boldsymbol{\theta}_{k+1}$ based on the current kth sample $\boldsymbol{\theta}_k$. For example, if we take a Gaussian $\mathcal{N}(0,\sigma)$ as the proposal distribution, $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \alpha$, where $\alpha \sim \mathcal{N}(0,\sigma)$. The choice of the appropriate proposal distribution requires some fine-tuning of the step size, as we would accept or reject too many samples if the distribution is too narrow or wide. Then, for each candidate sample, we calculate the following acceptance ratio:

$$\alpha = \frac{P(\boldsymbol{\theta}_{k+1}|\boldsymbol{D}, M)q(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)}{P(\boldsymbol{\theta}_k|\boldsymbol{D}, M)q(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})}. \tag{1.32}$$

New sample $\boldsymbol{\theta}_{k+1}$ is accepted if $\alpha > 1$, and otherwise it is acceptance with the probability of α. If rejected, $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$. As not all samples with low probability are rejected, there is a chance to explore regions with low probability. It is important to initialize the chain so that it does not require a large number of model evaluations to reach high probability regions. Often we discard some of early iterations, called the burn-in.

To determine if the MCMC chain has converged, we first inspect the chain values visually and check if the chain has not stuck at local minima. Furthermore, a metric such as Gelman-

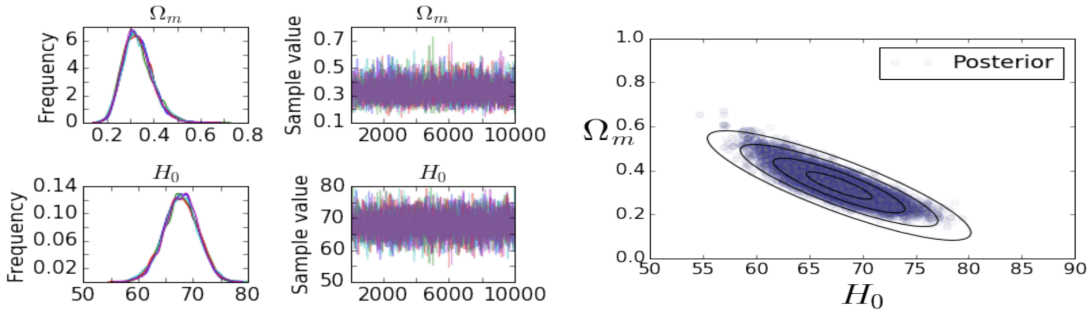


Figure 1.5: An example of posterior distributions from MCMC sampling. *Left*: 1-d density of the model parameters and the plot of chain values as a function of iterations. *Right*: 2-d posterior distribution with $1-4\sigma$ confidence regions. Figures from [198].

Rubin criterion is useful for the convergence test. With X number of chains, each of length Y, let $B = \frac{Y}{X-1}\sum_x(\langle\theta_{i,x}\rangle - \langle\theta_i\rangle)^2$, where $\langle\theta_{i,x}\rangle$ is the average of the parameter θ_i in the xth chain, and $\langle\theta_i\rangle$ is the global average of θ_i. B corresponds to the variance across the chain for θ_i. We further define $W = \frac{1}{X}\sum_x s_x^2$, where s_x^2 is the variance for θ_i within the xth chain. And $V = \frac{Y-1}{Y}W + \frac{1}{Y}B$, and this estimates the overall variance of θ_i. Finally, we define $R = \sqrt{V/W}$, and we check if $R \simeq 1$. For example, we can assume the convergence criteria to be $R < 1.1$. Or alternatively we can compute the auto-correlation between the samples and see if it lowers as we draw more samples. However, such metrics do not provides absolute convergence, and we need to use heuristics to determine when to stop the sampling procedure.

Figure 1.5 uses MCMC sampling to estimate the 1-d and 2-d posterior distributions of the present-day Hubble parameter H_0 and the total matter density Ω_m of the standard ΛCDM model. In the left panel, the chain values for each parameter are plotted, along with its density distributions. [198] reports Gelman-Rubin statistics for each paramter: R = 1.00045 for H_0 and 1.00044 for Ω_m, suggesting that the chain has reached convergence. The right panel plots the 2-d posterior distributions with $1-4\sigma$ confidence ellipses.

A simple Metropolis-Hastings algorithm may not be very efficient, if the proposal distribution does not agree with the posterior. Such inefficiencies lead to the development of more advanced samplers. Gibbs sampling, which reduces the dimension of the parameters by splitting $\boldsymbol{\theta}$ into blocks and separately sampling each, is one example [90]. Affine-invariant ensemble sampler, implemented in `emcee` Python package [87], uses the members of ensemble called walkers, and they run in parallel and exchange information at each step. Nested sampling [246], implemented in the algorithm such as `MultiNest` [85], is particularly useful for the calculation of the Bayesian evidence and sampling from a multimodal posterior distribution. Hamiltonian Monte Carlo (HMC) [35] uses the gradient of the sampled density to avoid the random walk sampling and efficiently approaches high probability regions. No-U-Turn Sampler (NUTS) [120] improves HMC in that it uses a recursive algorithm to eliminate the need of manual tuning required in HMC.

Figure 1.6: Results of optimizing approximate distributions from different families. By choosing a family of Gaussian distribution (orange), we only have two fitting parameters (faster optimization), but this leads to a high bias and consequently poor fit to the target posterior. Choosing a more complex distribution such as Gaussian mixture model (green), optimization takes longer, but it fits much better to the target. Figure from [221].

Variational inference as an alternative

MCMC sampling is widely used in cosmological parameter inference, but developing a faster yet reliable method has remained as a major challenge, as the computational cost of MCMC can be huge especially when the model evaluation is expensive. Approximate method such as variational inference (VI) provides a less-expensive alternative to MCMC. In VI, we approximate the target posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{D})$ with simple, analytically tractable distributions and hence transform parameter inference into an optimization problem.

In this section, let us assume that we only have a single parameter θ for the sake of simplicity. We propose a family of distributions $\mathcal{Q}$, called variational distributions, and then find a member $q(\theta) \in \mathcal{Q}$ closest to the target posterior $P(\theta|\boldsymbol{D})$. Figure 1.6 shows that finding an optimal distribution within $\mathcal{Q}$ does not guarantee a good fit; no distribution in $\mathcal{Q}$ may be close to the posterior, and therefore it is important to find $\mathcal{Q}$ complex enough to contain the target posterior yet simple enough to make the optimization feasible [221]. The goal of the optimization process is to find a member $q(\theta)$ by minimizing the distance to the target posterior, defined as the Kullback-Leibler (KL) divergence:

$$\begin{aligned} \mathrm{KL}(q(\theta)||P(\theta|\boldsymbol{D}, M)) &= \mathbb{E}_q[\ln q(\theta) - \ln P(\theta|\boldsymbol{D}, M)] \\ &= \ln P(\boldsymbol{D}|M) + \mathbb{E}_q \ln q(\theta) - \mathbb{E}_q[\ln P(\boldsymbol{D}|\theta, M) + \ln P(\theta|M)], \end{aligned} \tag{1.33}$$

where $\mathbb{E}_q$ denotes the expectation over $q(\theta)$. As the evidence, independent of the variational distribution, is a constant with respect to $q(\theta)$, minimizing the KL divergence corresponds to maximizing the rest of the terms, defined as evidence lower bound (ELBO) [48]:

$$\begin{aligned} \mathrm{ELBO}(q) &= \mathbb{E}_q[\ln P(\boldsymbol{D}|\theta, M) + \ln P(\theta|M)] - \mathbb{E}_q \ln q(\theta) \\ &= -\mathrm{KL}(q(\theta)||P(\theta|\boldsymbol{D}, M)) + \ln P(\boldsymbol{D}|M) \end{aligned} \tag{1.34}$$

[154] shows that KL(.) ≥ 0 between any distributions, and consequently the ELBO sets a lower bound for the log evidence as its name suggests: $\ln P(\boldsymbol{D}|M) \geq \text{ELBO}(q)$. The ELBO is useful because we cannot calculate the KL divergence directly; we instead optimize this quantity, which is equivalent to the KL divergence up to a constant. In VI, therefore, parameter inference becomes an optimization problem which aims to maximize the ELBO.

As shown in Figure 1.6, VI starts with specifying an appropriate $\mathcal{Q}$, and then we compute the ELBO and its derivatives, finally maximizing the ELBO with gradient-based methods. Black box variational inference (BBVI) avoids such model-specific calculations [212, 144], and automatic differentiation variational inference (ADVI) extends this by automatically solving the variational optimization (i.e. once the model is specified, ADVI finds an optimal VI algorithm automatically) [153]. This algorithm is integrated into `Stan` [56] and `PyMC3` [224] and widely used in scientific research.

In Chapter 5, we argue that full rank ADVI may not be useful in high-dimensional problems because of its computational inefficiency, and the stochastic nature of a KL divergence leads to the sampling noise. So we instead propose to replace the ELBO optimization with the optimization of a L_2 norm-based divergence, which converges significantly faster than KL divergence-based methods.

Chapter 2

Towards Neutrino Mass from Cosmology without Optical Depth Information

With low redshift probes reaching unprecedented precision, uncertainty of the CMB optical depth is expected to be the limiting factor for future cosmological neutrino mass constraints. In this chapter, we discuss to what extent combinations of CMB lensing and galaxy surveys measurements at low redshifts $z \sim 0.5 - 5$ will be able to make competitive neutrino mass measurements without relying on any optical depth constraints. We find that the combination of LSST galaxies and CMB-S4 lensing should be able to achieve constraints on the neutrino mass sum of 25meV without optical depth information, an independent measurement that is competitive with or slightly better than the constraint of 30meV possible with CMB-S4 and present-day optical depth measurements. These constraints originate both in structure growth probed by cross-correlation tomography over a wide redshift range as well as, most importantly, the shape of the galaxy power spectrum measured over a large volume. We caution that possible complications such as higher-order biasing and systematic errors in the analysis of high redshift galaxy clustering are only briefly discussed and may be non-negligible. Nevertheless, our results show that new kinds of high-precision neutrino mass measurements at and beyond the present-day optical depth limit may be possible.[1]

2.1 Introduction

An important goal in both particle physics and cosmology is to understand the physics underlying the neutrino mass [5]. The fact that neutrinos have a non-zero mass has been known since the discovery of neutrino oscillations; however, the absolute scale of this mass is uncertain, with oscillation experiments only giving a lower bound of ≈ 60 meV for the normal

[1]This chapter is taken from "Towards Neutrino Mass from Cosmology without Optical Depth Information," *Yu B., Knight Z., Sherwin B., Ferraro S., Knox L., Schmittfull M.* (arXiv: 1809.02120).

hierarchy and ≈ 100 meV for the inverted hierarchy. A measurement of the neutrino mass would not just reveal a new energy scale, it would set targets for terrestrial double beta decay experiments (thus potentially contributing to a determination of whether neutrinos are Dirac or Majorana particles) and might even give insight into the mass ordering. Perhaps the most exciting possibility is that the combination of cosmological and laboratory measurements reveals inconsistencies requiring new physics. A cosmological neutrino mass measurement would significantly contribute to efforts to understand physics in the neutrino sector.

The neutrino mass can be probed precisely in cosmology because properties of the cosmic neutrino background affect the growth of cosmic structure and the expansion history of the universe [160, 91]. A primary effect targeted by future experiments is the suppression of growth of small scale structure caused by a nonzero neutrino mass. The rest mass of the neutrinos, as they become non-relativistic, increases the neutrino contribution to the total mean energy density beyond what it would be in the massless case, thereby increasing the expansion rate and thus suppressing growth. A secondary effect is the scale dependence of this suppression: above the free-streaming scale the neutrinos act just like cold dark matter and therefore contribute to gravitational instability, with the net effect of canceling out the suppressive effect of the increased expansion rate [51, 175, 126, 142, 200]. This results in a fairly broad "step"-like feature in the matter power spectrum, where the size of the step is time dependent and grows approximately linearly with every e-fold of expansion[2].

To measure neutrino mass using this time-dependent suppression, the amplitude of structure at low redshift (probed by gravitational lensing, clusters, or redshift-space distortions) is typically compared with the initial, high redshift amplitude probed by the CMB. In particular, the small-scale suppression is about 4% between the redshift of recombination and today for the minimal mass of 60 meV.

With the design of increasingly powerful CMB surveys, such as CMB Stage-4 experiment (CMB-S4, [4]) and Simons Observatory (SO, [8]), it has become clear that the limiting factor for upcoming neutrino mass constraints will not be the precision of the measurement of CMB lensing or other low redshift probes, but instead the precision of the high redshift amplitude of structure at the CMB redshift $z \approx 1100$ [18]. This high-z amplitude A_s, in turn, is limited by how well we know the optical depth τ to the CMB, because the combination $A_s e^{-2\tau}$ (describing the amplitude of the CMB power spectrum) is what is measured by CMB surveys [18, 183]. Since it is unclear whether substantially improved τ constraints will be forthcoming, it is well motivated to seek methods by which the neutrino mass can be probed without relying on a knowledge of the CMB optical depth.

In this chapter we examine to what extent the combination of CMB lensing from future experiments with galaxy surveys such as LSST can be used to obtain competitive neutrino mass constraints *without optical depth information*. We further consider what future surveys are required to improve on optical depth limited neutrino mass constraints.

Our investigation is motivated by two effects that may allow neutrino mass constraints

[2]In linear theory, it can be shown that the size of the "step" feature in the power spectrum grows by $\frac{6}{5} f_\nu$ per $e-$fold of expansion, where $f_\nu = \Omega_\nu / \Omega_m$ is the fraction of mass in neutrinos.

without optical depth information. First, the time dependence of the neutrino mass suppression of structure growth can not only be seen by comparing the amplitude of fluctuations of the CMB and today; it can also be seen at low redshift alone using cross-correlations to probe the growth of structure over a sufficiently long redshift lever arm, given sufficiently precise measurements. Extremely high precision constraints on the amplitude of structure as a function of redshift were indeed forecast by [231], and we build on these results in our analysis (for similar recent forecasts also see, for example, [24, 184]). Second, there are other physical effects, such as the step feature in the shape of the matter power spectrum described previously, through which neutrino mass can be constrained with low redshift probes alone; we will consider these as well.

Our work follows a long list of papers that have forecasted constaints on neutrino mass to come from cosmological surveys [55, 24, 263, 176, 172, 54]. But it is the first to explore how the combination of CMB lensing and galaxy counts can be used to exploit the effects described in the previous paragraph to evade the impact of uncertainty about the optical depth to Thomson scattering.

We will begin by introducing our forecasting assumptions, before presenting and discussing our results.

2.2 Forecasting Method and Survey Systematics

A. Angular Power Spectra

We use the observed galaxy density field in ith tomographic redshift bin g_i and the CMB lensing convergence κ to construct the 2-point angular power spectra: $C_l^{\kappa\kappa}$, $C_l^{\kappa g_i}$, and $C_l^{g_i g_i}$. In the Limber approximation [168], we model all angular power spectra

$$C_l^{\alpha\beta} = \int \frac{dz H(z)}{\chi^2(z)} W^\alpha(z) W^\beta(z) P_{\delta_\alpha \delta_\beta}\left(k = \frac{l}{\chi(z)}, z\right), \tag{2.1}$$

where $\alpha, \beta \in (\kappa, g_1, ..., g_N)$, $H(z)$ is the Hubble parameter, $\chi(z)$ is the comoving angular-diameter distance to redshift z, $P(k, z)$ is the matter power spectrum at wavenumber k and redshift z, and N is the number of bins. δ_g is the CDM-baryon density contrast δ_{cb}, and δ_κ is the total matter density contrast $\delta_{cb\nu}$ including neutrinos. For the CMB lensing convergence, the redshift kernel $W^\kappa(z)$ is

$$W^\kappa(z) = \frac{3}{2H(z)} \Omega_m H_0^2 (1+z) \chi(z) \left(\frac{\chi_* - \chi(z)}{\chi_*}\right), \tag{2.2}$$

where χ_* is the comoving distance to the last scattering surface, and Ω_m and H_0 are the matter density and the Hubble parameter today, respectively. For the ith bin galaxy density field g_i, the kernel is

$$W^{g_i}(z) = \frac{b_i(z) dn_i/dz}{\int dz' (dn_i/dz')}, \tag{2.3}$$

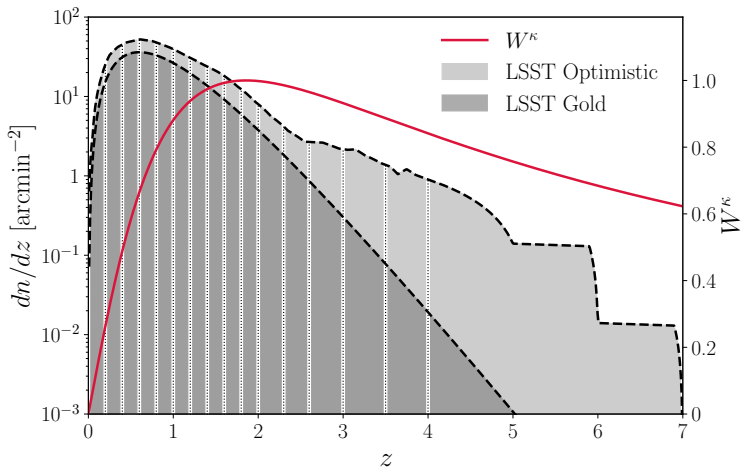


Figure 2.1: The redshift distribution of the CMB lensing convergence (red curve, normalized to a unit maximum) and LSST galaxy samples, both Optimistic (light gray) and Gold (dark gray). We assume 16 tomographic redshift bins in the range $0 < z < 7$, cross-correlation bin widths indicated with vertical dotted lines.

where dn_i/dz is the redshift distribution of the galaxies in the ith bin. We assume the linear galaxy bias is given by $b_i(z) = B_i(1+z)$ within each bin, where B_i is the overall bias amplitude in the ith bin [6] (we assume a fiducial value of $B_i = 1$). Fig. 2.1 compares the CMB lensing kernel with the redshift distribution of two different LSST samples, as further described in Section B. We use the publicly available `CAMB` Boltzmann code to calculate the power spectrum $P_{\delta_\alpha \delta_\beta}(k,\ z)$ [1, 161].

B. LSST Specifications

We assume two LSST number densities, as shown in Figure 2.1. The first sample is the $i < 25$ gold sample (henceforth referred to as "Gold"), corresponding to $\overline{n} = 40\,\mathrm{arcmin}^{-2}$ and $n(z) \propto 1/(2z_0)(z/z_0)^2 e^{-z/z_0}$ following [6] with $z_0 = 0.3$. As a second sample, we use a more optimistic $i < 27$ magnitude cut with $S/N > 5$ in the i band assuming three years of observations following [94] ("Optimistic"), and add Lyman break galaxies from redshift dropouts, whose number density we estimate by scaling recent HSC observations [195, 110] following [231]. This yields $\overline{n} \approx 66\,\mathrm{arcmin}^{-2}$ galaxies at $z = 0 - 7$. We decompose the LSST kernel into 16 tomographic bins, with redshift edges of $z = [0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1,\ 1.2,\ 1.4,\ 1.6,\ 1.8,\ 2,\ 2.3,\ 2.6,\ 3,\ 3.5,\ 4,\ 7]$, assuming that neighboring bins do not overlap. To reduce the sensitivity of our forecasts to uncertainties in non-linear modeling, including bias modeling, we keep the density perturbations in the near-linear regime by setting a $k_{\rm max}$ limit ($0.3\ h\mathrm{Mpc}^{-1}$ is assumed in Figure 2.2–2.6, with lower $k_{\rm max}$ shown in Table 2.1). For each bin, we convert this to $l_{\rm max} = k_{\rm max}\chi(\overline{z}_i)$, where $\overline{z}_i$ is the mean redshift of the ith bin. Imposing $k_{\rm max} = 0.3\ h\mathrm{Mpc}^{-1}$, we find that including non-linear corrections from Halofit [254,

262] has a negligible effect on our forecasts when all external datasets, such as primordial CMB and DESI information, are included. Hence, we use the linear matter power spectrum in all forecasts. We assume the survey area of 18,000 deg^2, which corresponds to $f_{\rm sky} \approx 0.4$. Finally, we neglect any redshift space distortion effects in the LSST power spectra.

C. CMB-S4 Specifications

For CMB lensing, we use a CMB-S4 experiment with the following configurations: beam FWHM $= 1'$, $\Delta_T = 1\mu K'$, and $\Delta_{E,B} = 1.4\mu K'$. We assume $f_{\rm sky} = 0.4$, with CMB-S4 fully overlapping with the LSST [4]. White noise is assumed, as we expect the impact of non-white noise to be small for lensing reconstruction from polarization-dominated experiments. With `quicklens` [109, 9], we compute the minimum variance quadratic estimator lensing reconstruction on the full sky with $l^{T,E,B}_{\rm min} = 50$, $l^{T}_{\rm max} = 3000$, and $l^{E,B}_{\rm max} = 5000$. We take into account the improvement from iterative lens reconstruction by rescaling the EB noise [118, 252]. In Table 2.1 and 2.2, we show forecasts assuming the resulting CMB-S4 lensing reconstruction noise. For the CMB lensing convergence κ, we set $l_{\rm min} = 30$ and $l_{\rm max} = 2000$.

Additionally, with the CMB-S4 specifications as described above, we compute the CMB-S4 Fisher matrix, using temperature and polarization power spectra from S4, to break the parameter degeneracies. We also consider Planck primary CMB data for $l > 30$ in the region not overlapping with the CMB-S4 ($f_{\rm sky} = 0.25$ accordingly) [4]. Since we aim here to investigate neutrino mass constraints without τ information, no prior on the optical depth to reionization τ is included, unless we explicitly note otherwise. Here we use the unlensed CMB power spectra because the lensing auto-power spectrum $C_l^{\kappa\kappa}$ already provides nearly all the CMB lensing information [251] and because then the source of lensing information is entirely clear.

D. DESI Specifications

We include the forecasted galaxy baryon acoustic oscillation (BAO) information from the Dark Energy Spectroscopic Instrument (DESI) [64] which measures the distance-redshift relation at low redshift. (We neglect RSD and other broadband sources of information in the DESI galaxy power spectrum, but assume BAO reconstruction.) Including DESI measurements significantly improves neutrino mass forecasts by better constraining Ω_m and further breaking parameter degeneracies. We use the expected uncertainties on the distance ratio from 18 bins in the range $0.15 < z < 1.85$ with $\Delta z = 0.1$, given in [10, 18].

E. Fisher Matrix Analysis

If we have N tomographic galaxy redshift bins, our observables are $1 + N$ (lensing-lensing and galaxy-galaxy) auto-power spectra and $N + N(N-1)/2$ (lensing-galaxy and galaxy-galaxy) cross-spectra. For the CMB lensing convergence auto-spectra, we consider the lensing

reconstruction noise $N_l^{\kappa\kappa}$, and for the galaxy-galaxy auto-spectra, we take into account the shot noise $N_l^{gg} = 1/\overline{n}$.

The Gaussian covariance matrix of the CMB lensing convergence and the LSST galaxy auto- and cross-power spectra is given by

$$\begin{aligned}\mathrm{Cov}_{l_a,l_b}^{\alpha_1\beta_1,\alpha_2,\beta_2} = \frac{\delta_{l_a,l_b}}{f_{\rm sky}(2l_a+1)}\Big\{&(C_{l_a}^{\alpha_1\alpha_2}+N_{l_a}^{\alpha_1\alpha_2})\\ &(C_{l_a}^{\beta_1\beta_2}+N_{l_a}^{\beta_1\beta_2})+(C_{l_b}^{\alpha_1\beta_2}+N_{l_b}^{\alpha_1\beta_2})\\ &(C_{l_b}^{\beta_1\alpha_2}+N_{l_b}^{\beta_1\alpha_2})\Big\},\end{aligned} \tag{2.4}$$

where $\alpha_{1,2}, \beta_{1,2} \in (\kappa, g_1, ..., g_N)$.

We then construct the Fisher matrix

$$F_{ij} = \sum_{\substack{\alpha_1\beta_1,\\ \alpha_2,\nu_2}} \sum_l \frac{\partial C_l^{\alpha_1\beta_1}}{\partial\theta_i}[\mathrm{Cov}_l^{\alpha_1\beta_1,\alpha_2,\beta_2}]^{-1}\frac{\partial C_l^{\alpha_2\beta_2}}{\partial\theta_j}, \tag{2.5}$$

where $\vec{\theta} = \{B_i,\ H_0,\ \Omega_b h^2,\ \Omega_c h^2,\ n_s,\ A_s,\ \sum m_\nu,\ \tau\}$. B_i is the bias amplitude parameter of the ith bin. We take the fiducial values for τ and $\sum m_\nu$ to be 0.06 and a minimum value of 60 meV, respectively. Unless stated otherwise, we fix $w = -1$.

Finally, we combine the above Fisher matrix with the primordial CMB and BAO Fisher matrices and compute the marginalized constraints as $\mathrm{Cov}(\theta_i, \theta_j) = (F^{-1})_{ij}$.

2.3 Results and Interpretation

With the Fisher matrix formalism described above, Fig. 2.2 presents forecasts of 1σ constraints on the sum of neutrino masses, marginalized over ΛCDM parameters and linear galaxy biases in all redshift bins, for $k_{\rm max} = 0.3\ h\mathrm{Mpc}^{-1}$. No prior on the optical depth to reionization is included. With the LSST Optimistic sample split into 16 bins in the range $z = 0 - 7$, combining LSST clustering and CMB lensing from S4 gives $\sigma(\sum m_\nu) = 55$ meV. Adding the primordial CMB information (without any prior on τ), we can achieve a constraint of 33 meV, corresponding to a $\approx 1.8\sigma$ detection on the minimum value of $\sum m_\nu$ for the normal hierarchy. Using the parameter constraints from S4, we gain $\approx 7\%$ improvement in forecasts relative to the Planck primary Fisher matrix. Hereafter in this analysis, we use S4 primary CMB information with Planck co-added. Finally, with the the DESI BAO measurements added, we can achieve $\sigma(\sum m_\nu) = 24$ meV, reaching $\approx$ a 2.5σ measurement of the minimal sum of the neutrino mass, without any optical depth information.

In Fig. 2.2, we find that adding clustering information at higher redshift results in significantly better $\sum m_\nu$ constraints. A more pessimistic galaxy sample, LSST Gold, includes significantly less structures in high redshift and therefore yields only a minimal improvement in the constraints for $z > 3$. However, relative to the LSST Optimistic, the

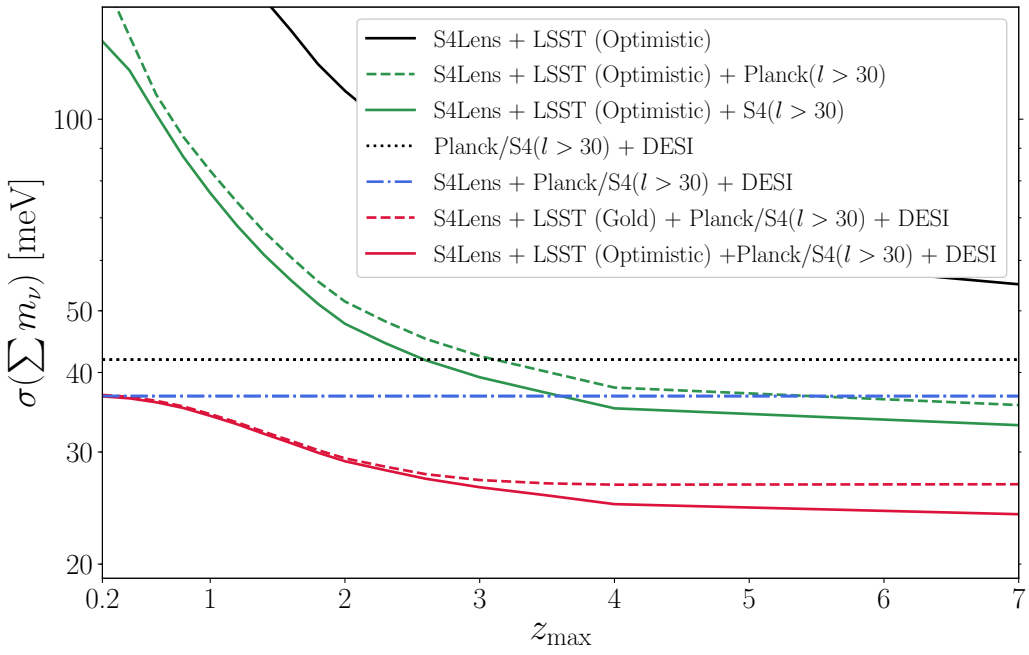


Figure 2.2: Forecasted 1σ constraints on the sum of the neutrino masses without optical depth information, for different experiment configurations: CMB-S4 lensing and LSST clustering (black) + primordial CMB data (green dotted for Planck and green solid for S4) + DESI BAO measurements (red solid for LSST Optimistic and red dotted for LSST Gold). S4 primary CMB (with Planck co-added) + DESI BAO gives $\sigma(\sum m_\nu) = 42$ meV, which further tightens to 37 meV with the reconstructed CMB lensing potential included. Including the LSST galaxies at higher redshift extends the redshift lever arm and increases the volume probed, which results in a significant improvement in the constraints.

$\sigma(\sum m_\nu)$ Gold sample constraints are not significantly worse when primary CMB and DESI information are included. We also consider the effect of having a broader redshift binning; with 6 bins in the same redshift range, $\sigma(\sum m_\nu)$ degrades by $\approx 15\%$.

Table 2.1 provides the 1σ constraints on the neutrino mass with different $k_{\rm max}$ limits, for both LSST Gold and Optimistic samples. Having just CMB lensing and LSST clustering, we find significant improvements as we assume a higher $k_{\rm max}$. However, with all external datasets included, we find only moderate dependence on $k_{\rm max}$, with a degradation of only $10-15\%$ when using $k_{\rm max} = 0.1\ h{\rm Mpc}^{-1}$ instead of $k_{\rm max} = 0.3\ h{\rm Mpc}^{-1}$. The dependence on CMB sensitivity is similar: Improved CMB sensitivity improves constraints from CMB lensing and LSST clustering alone significantly, but only mildly when including all other probes. We note that such modest improvements of the neutrino mass constraints with the S4 lensing reconstruction noise have been recognized previously [24].

We emphasize that the forecasts shown in Fig. 2.2 and Table 2.1 assume no prior information on the optical depth. We therefore conclude that the τ-less cross-correlation tomography

	$\sigma(\sum m_\nu)$ [meV] (Gold/Optimistic)		
$k_{\rm max}$	Lens + LSST	+ Planck/S4 T&P	+ DESI
0.05	307 / 243	94 / 68	32 / 29
0.1	176 / 129	68 / 53	31 / 27
0.2	107 / 71	47 / 38	28 / 25
0.3	84 / 55	40 / 33	27 / 24
0.4	79 / 49	38 / 31	26 / 23

Table 2.1: Forecasts of the neutrino mass constraints without optical depth information, for different LSST number densities and redshift distributions, $k_{\rm max}$ limits, and lensing reconstruction noise levels. For different combinations of data, constraints provided on the left assume the LSST Gold sample, and those on the right assume the LSST Optimistic sample.

	$\sigma(\sum m_\nu)$ [meV]		
	Lens + Planck/S4 T&P + DESI		
$k_{\rm max}$	$\sigma(\tau) = 0.01$	0.005	0.002
0.3	25	17	12

	$\sigma(\sum m_\nu)$ [meV] (Gold/Optimistic)		
	Lens + LSST + Planck/S4 T&P + DESI		
$k_{\rm max}$	$\sigma(\tau) = 0.01$	0.005	0.002
0.3	22 / 20	16 / 16	11 / 10

Table 2.2: Forecasts of the neutrino mass constraints with different flat priors on the optical depth assumed. *Top*: Combining CMB-S4 lensing, S4 primary CMB (with Planck co-added), and DESI BAO information. *Bottom*: LSST clustering added. As in Table 2.1, numbers on the left assume the Gold sample, and those on the right assume the Optimistic sample.

combining LSST clustering and CMB-S4 lensing provides a different and competitive way to measure the sum of the neutrino masses. This is better illustrated in Fig 2.3. We obtain slightly tighter bounds on $\sum m_\nu$ and τ (red solid curve) compared to the τ-limited bounds possible with CMB-S4 (blue dotted). Still, including a tight prior on τ constrains $\sum m_\nu$ better. Table 2.2 summarizes the effects of the optical depth measurements on the neutrino mass constraints in our forecasts. Assuming $k_{\rm max} = 0.3\ h{\rm Mpc}^{-1}$, adding a flat prior $\sigma(\tau) = 0.01$ improves our constraints by $15-20\%$. A better determination of τ reduces the uncertainty on the $\sum m_\nu$ detection; $\sigma(\tau) = 0.005$ tightens our 1σ constraint to 16 meV, and imposing the cosmic variance limit on the τ measurements brings $\sigma(\sum m_\nu)$ down to 10 meV, $\approx 6\sigma$ detection on the minimal sum of the neutrino masses (LSST Optimistic sample with S4

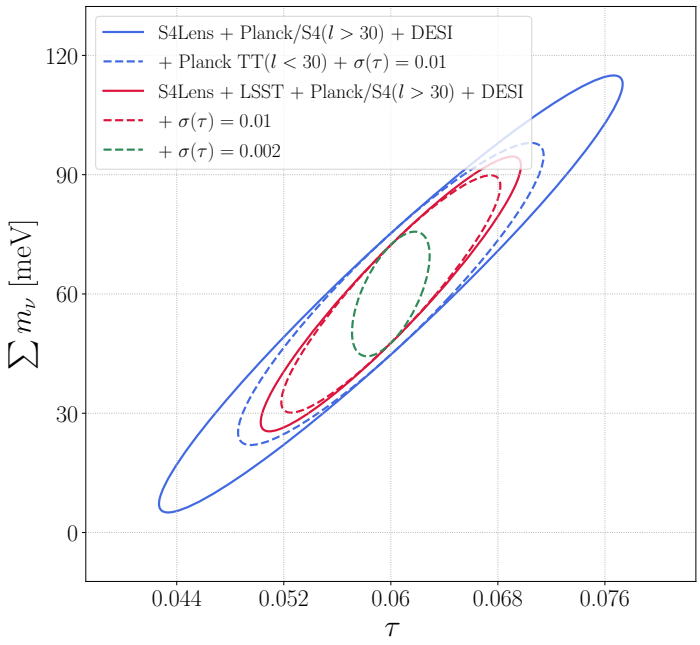


Figure 2.3: 1σ confidence ellipses in the $\tau - \sum m_\nu$ plane, with different combinations of datasets. The solid curves assume no prior on the optical depth, whereas the dotted curves include a flat prior on τ. We find that the combination of LSST clustering and CMB-S4 lensing without any τ information (red solid) can achieve constraints competitive with or slightly better than the τ-limited constraints possible with CMB-S4 (blue dotted).

lensing noise assumed).

What is the physical origin of these neutrino mass constraints without optical depth information? We consider two possible mechanisms by which the constraints could arise.

First, they could originate by probing neutrinos' effect on the growth of structure over a wider range of low redshifts. (We will henceforth refer to this as the "growth effect".) To illustrate this, we forecast the constraints on the amplitude of matter fluctuations σ_8 as a function of redshift, by defining a parameter A_i which quantifies how the measured power spectra deviate from the standard growth of structure: $P_{mm}(k, z_i) = A_i^2 P_{mm}^{\rm fiducial}(k, z_i)$, with $A_i = 1$ for the fiducial cosmology. Following [231], we consider broader redshift bins, $z = 0 - 0.5, 0.5 - 1, 1 - 2, 2 - 3, 3 - 4, 4 - 7$, and treat A_i in all 6 bins as a free parameter. Marginalizing over 6 ΛCDM parameters (H_0, $\Omega_b h^2$, $\Omega_c h^2$, n_s, A_s, τ) and linear biases in each bin and adding external datasets, such as primary CMB and DESI, we can convert A_i constraints to subpercent-level constraints on σ_8 at each redshift, as shown in Fig 2.4. This enables us to measure (to some extent) the tiny difference between high- and low- redshift amplitudes of structure, thereby leading to a better constraint on $\sum m_\nu$.

Since the precision to which the growth suppression alone can be measured appears moderate, we also consider other physical effects that can contribute to the constraints on

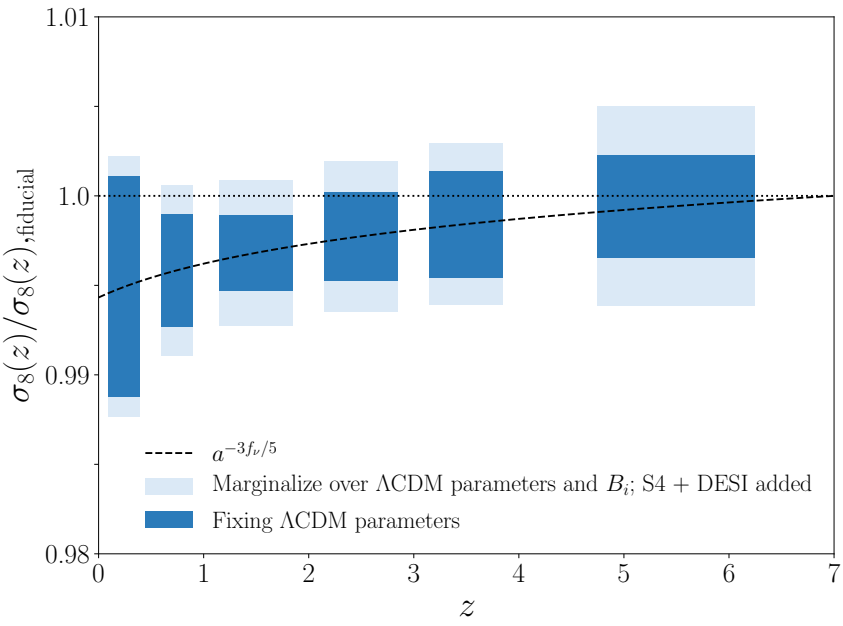


Figure 2.4: 1σ constraints on the matter amplitude σ_8 in 6 tomographic redshift bins, $z = 0-0.5, 0.5-1, 1-2, 2-3, 3-4, 4-7$, from the combination of LSST galaxies and CMB-S4 lensing. $k_{\mathrm{max}} = 0.3\ h\mathrm{Mpc}^{-1}$ is assumed. $\sigma_8/\sigma_{8,\mathrm{fiducial}} = 1$ corresponds to $\sum m_\nu = 0$. Massive neutrinos suppress the growth of density fluctuations, which can be shown by how the matter density contrast scales with the scale factor: $\delta_m \propto a^{1-\frac{3}{5}f_\nu}$ [4]. Assuming the minimal mass sum 60 meV, the black dotted curve plots such suppression. We either (1) marginalize over ΛCDM parameters and linear biases in each bin (light blue blocks) or (2) fix ΛCDM parameters (dark blue). In both scenarios, subpercent-level constraints on σ_8 can be achieved, leading to a significant improvement in the $\sum m_\nu$ detection.

neutrino mass. In particular, we consider constraints from the step-feature in the power spectrum induced by neutrino free streaming (i.e., the characteristic spectrum shape caused by growth suppression only below the free streaming scale); this should also improve with larger volume and a larger number of low-k modes, as surveys extend to higher redshift. We will label this effect the "spectrum shape effect".

In Fig. 2.5, we investigate the relative contribution of the growth and spectrum shape effects to the constraints on neutrino mass without optical depth information. We begin from an analysis including the full information arising from both galaxies and CMB lensing, in which we obtain constraints shown by the red line. To understand the relative contributions, will now remove either the growth effect or the spectrum shape effect. To remove the growth effect, we simply exclude all CMB lensing information ($C_l^{\kappa\kappa}$ and $C_l^{\kappa g}$ removed); this gives the constraints shown by the blue line. To remove the spectrum shape effect, we artificially remove the neutrino step feature by matching an $\sum m_\nu = 0$ power spectrum to the amplitude of the small-scale power spectrum at $k > 0.1h/\mathrm{Mpc}$. This way, the whole "featureless" power

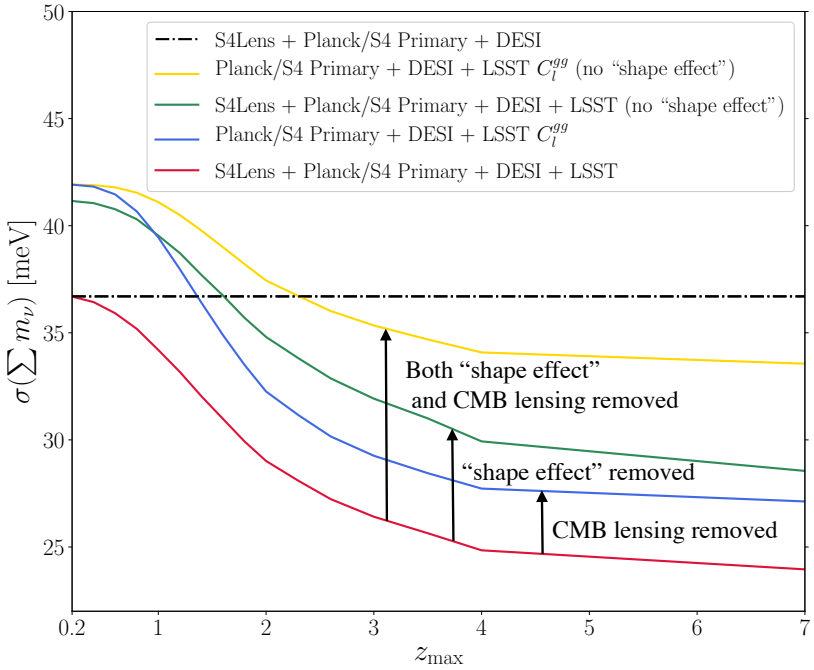


Figure 2.5: The relative contribution of the growth and spectrum shape effects to the $\sum m_\nu$ constraint without optical depth information. $k_{\rm max} = 0.3\ h{\rm Mpc}^{-1}$ assumed. From the full information combining both galaxies and CMB lensing (red curve), we remove either the growth effect by excluding all CMB lensing information (blue curve) or the spectrum shape effect by artificially removing the neutrino step feature (green curve). The removal of either effect substantially weakens our constraints, and removing both growth and shape effects (yellow curve) eliminates the majority of the constraining power of our data.

spectrum growth is suppressed in a redshift dependent way that mimics that caused by neutrinos. This gives the constraints shown by the green line. It can be seen that in both cases, constraints are weakened substantially; the effect sizes appear comparable, though the removal of the spectrum shape effect has slightly more impact [3]. Removing both the spectrum shape effect and the CMB lensing data eliminates the majority of the constraining power of our data; we thus conclude that both the shape of the galaxy power spectrum and the growth of cosmic structure, probed by high redshift galaxy and CMB lensing surveys, are responsible for the majority of our constraints on neutrino mass without optical depth information.

Since we find that our forecasts for neutrino mass errors do not degrade dramatically with the complete removal of CMB-S4 lensing information, we have also looked at constraints

[3]We note that when we reduce our default $k_{\rm max} = 0.3h/{\rm Mpc}$ to $k_{\rm max} = 0.1h/{\rm Mpc}$, the spectrum shape effect, which mainly arises from low k, becomes much more important than the growth effect, which requires many modes to get precise measurements of $\sigma_8(z)$.

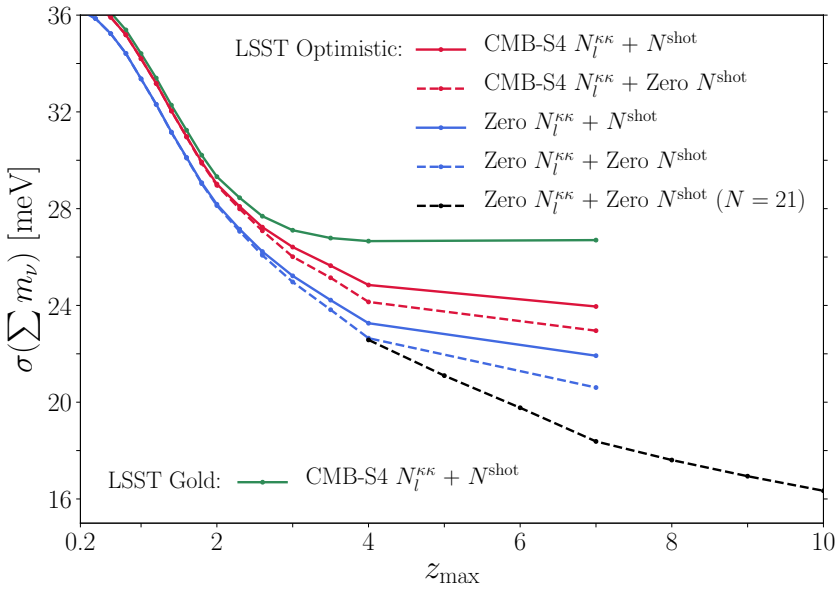


Figure 2.6: Forecasted 1σ constraints on $\sum m_\nu$ with different survey configurations. The solid curves include the LSST shot noise, and the dotted curves assume zero shot noise. Having more galaxies observed in higher redshift, LSST Optimistic (red, blue, and black curves) yields tighter $\sum m_\nu$ constraints relative to LSST Gold (green). Assuming $N = 22$ in the same redshift range, we can reach up to $\sigma(\sum m_\nu) = 16$ meV. Consequently, We conclude that having more bins in high redshift tightens our constraints considerably.

to come from a nearer-term survey with 7 times larger map noise at $\Delta_T = \Delta_{E,B}/2 = 7\mu$K$'$. This is similar to (though not necessarily equal to) the white noise level expected from the Simons Observatory, also to be situated on the Atacama Plateau, with a survey coverage of $f_{\rm sky} \simeq 0.4$ [8]. For $k_{\rm max} = 0.3h$/Mpc, our forecast for $\sigma(m_\nu)$/meV from CMB temperature, polarization and lensing combined with DESI BAO and the LSST Gold (Optimistic) sample degrades from 27 to 29 (24 to 26) when S4 is replaced with this nearer-term, noisier survey.

Though we believe we have explained the origin of most of the combined probes' constraining power, other effects may contribute to some degree as well, such as: improved constraints on cosmological parameters such as the matter density, which may break degeneracies with neutrino mass, or constraints on the geometric factors probed by the relevant power spectra. We defer a detailed analysis attempting to quantify the impact of these other effects to future work.

The analyses described in this chapter might provide the best prospects for improved constraints in future experiments, since improving optical depth constraints further may be difficult. Fig. 2.6 explores possible improvements to our constraints, and shows that our forecasts are moderately limited by the CMB lensing reconstruction noise and the galaxy shot noise. Even though our LSST galaxy samples extend to $z = 7$, we consider one broad redshift

bin for $z = 4-7$, and only modest improvements can be achieved by including this broad bin. The black dotted curve in Fig. 2.6 assumes $N = 22$, with finer bins in high redshift: $\Delta z = 1$ in the range $z = 4-10$ (flat dn/dz assumed for $z > 7$). Then, assuming zero lensing noise and shot noise, we can achieve $\sigma(\sum m_\nu) = 16$ meV from the combination of LSST galaxies and CMB-S4 lensing (all external datasets also added). This suggests that our forecasts are primarily limited by the redshift extent of the galaxy surveys. Neutrino mass constraints thus provide some motivation for extending galaxy surveys to higher redshift, though the improvements are fairly slow and the analyses will be very challenging.

2.4 Conclusions and Outlook

We have forecast that the combination of LSST clustering and CMB-S4 lensing provides competitive neutrino mass constraints without optical depth information. Following [231], we use CMB lensing – galaxy survey cross-correlations, together with auto-power spectrum information, to cancel sample variance in part and thereby break parameter degeneracies.

For $k_{\rm max} = 0.3\ h{\rm Mpc}^{-1}$, the combination of CMB-S4 lensing with LSST galaxies, with external datasets such as Planck and S4 primordial CMB information and DESI BAO measurements included, can achieve $\sigma(\sum m_\nu) = 24$ meV, corresponding to a $\approx 2.5\sigma$ detection on the minimal mass 60 meV assuming the normal hierarchy. This suggests that the τ-less CMB lensing cross-correlation tomography provides an (at least partially) independent and competitive way to constrain the sum of the neutrino masses. Such improvements partially originate from sub-percent level constraints on the amplitude of structure at a number of different redshifts, which allow the measurement of the tiny difference between high- and low-redshift amplitudes of structure caused by neutrinos affecting structure growth; they also, in part, originate in constraints on the shape of the galaxy power spectrum, which benefit from the large volumes probed by high redshift surveys.

We demonstrate that including LSST galaxies at higher redshift leads to tighter constraints by extending the redshift lever arm. Comparing two LSST galaxy samples, we conclude that for a more pessimistic sample that includes less galaxies at high redshift the improvements in the constraints are only minimal for $z > 3$. We also assume zero lensing reconstruction noise and galaxy shot noise and find that the redshift lever arm and tomographic binning of the galaxy surveys (and the corresponding overlap with the lensing kernel) primarily limit our forecasts. In addition, we show that better measurements of the optical depth, if attainable and added to the analyses we describe, can improve the neutrino mass constraints further; in particular, including a cosmic-variance-limited optical depth measurement tightens $\sigma(\sum m_\nu)$ to 10 meV.

We caution that for our forecasts to hold, we need to be able to model the observed power spectrum in presence of massive neutrinos to better than $\sim 1\%$ level, which corresponds to the size of the suppression due to neutrinos in the range probed by LSST galaxies (see Fig. 2.4). For comparison, the size of the quadratic b_2 bias [23, 184], neglected in this analysis, can be a few percent correction to the galaxy power spectrum at $k = 0.1h{\rm Mpc}^{-1}$ and a $\sim 20\%$

correction at $k = 0.3h\mathrm{Mpc}^{-1}$, depending on redshift and on the mass of the host halos. We therefore anticipate needing to model and constrain scale-dependent corrections to the κg and gg power spectra from nonlinear bias terms to better than $\sim 10\%$ in order to achieve the required accuracy. Moreover, in this work we have neglected super-sample variance, and systematic errors in the analysis of high redshift galaxy clustering (such as photometric redshift uncertainties), which may limit how well we constrain the growth of structure. A full analysis of nonlinear biasing, photometric redshift errors and other systematic limitations is therefore well-motivated.

Nevertheless, if these systematic limitations can be controlled sufficiently well, our results show that novel high-precision neutrino mass measurements at and beyond the optical depth limit will be achievable with upcoming surveys.

Chapter 3

The Physical Origin of Dark Energy Constraints from Rubin Observatory and CMB-S4 Lensing Tomography

In this chapter, we seek to clarify the origin of constraints on the dark energy equation of state parameter from CMB lensing tomography, that is the combination of galaxy clustering and the cross-correlation of galaxies with CMB lensing in a number of redshift bins. We focus on the analytic understanding of the origin of the constraints. Dark energy information in these data arises from the influence of three primary relationships: distance as a function of redshift (geometry), the amplitude of the power spectrum as a function of redshift (growth), and the power spectrum as a function of wavenumber (shape). We find that the effects from geometry and growth play a significant role and partially cancel each other out, while the shape effect is unimportant. We also show that Dark Energy Task Force (DETF) figure of merit forecasts from the combination of LSST galaxies and CMB-S4 lensing are comparable to the forecasts from cosmic shear in the absence of the CMB lensing map, thus providing an important independent check. Compared to the forecasts with the LSST galaxies alone, combining CMB lensing and LSST clustering information increases the FoM by roughly a factor of 3-4 in the optimistic scenario where systematics are fully under control. We caution that achieving these forecasts will likely require a full analysis of higher-order biasing, photometric redshift uncertainties, and stringent control of other systematic limitations, which are outside the scope of this work, whose primary purpose is to elucidate the physical origin of the constraints. [1]

[1]This chapter is taken from "The Physical Origin of Dark Energy Constraints from Rubin Observatory and CMB-S4 Lensing Tomography," *Yu B., Ferraro S., Knight Z., Knox L., Sherwin B.* (arXiv: 2108.02801).

3.1 Introduction

Future surveys of the Cosmic Microwave Background (CMB) in intensity and polarization will produce high signal-to-noise ratio (SNR) CMB lensing maps over a large fraction of the sky. The survey conducted from the Atacama Plateau in Chile by CMB-S4 [4] will have significant spatial overlap with the deep and wide photometric galaxy catalogs to come from the Vera C. Rubin Observatory, an optical facility located at Cerro Pachón, also in Chile [6]. For the first ten years of operation, the Rubin Observatory will perform the Rubin Observatory Legacy Survey of Space and Time (LSST).

In this chapter we investigate physical origin of the constraints on the dark energy equation of state (EoS) parameter w that can be obtained by cross-correlating redshift-binned galaxy maps and a high SNR CMB lensing map. We choose to not include the forecasts of galaxy weak lensing (WL), in order to find out what can be achieved without WL and to create a complementary probe to those which do include WL. We have previously explored the physical origin neutrino mass constraints in a very similar setup in chapter 2.

Varying the EoS parameter of dark energy affects both the expansion rate of the Universe and the growth of large-scale structure (LSS), which impacts both the amplitude of the matter power spectrum and the angular position of the Baryon Acoustic Oscillation (BAO) features within it. The distribution of galaxies within a narrow redshift bin traces the distribution of matter at that redshift, and therefore its map and power spectrum contain information about expansion and growth in the same redshift range. On the other hand, the lensing of the CMB traces the distribution of matter over a wide range of redshifts, combined into a single map and power spectrum. Galaxy surveys measure the luminous matter, while lensing is sensitive to the underlying matter distribution, so we expect the cross-correlation between galaxies and lensing to provide a measurement of the relationship between luminous and dark matter [4], crucially breaking the intrinsic degeneracy between the amplitude of fluctuations and galaxy bias. Our goal is to describe the benefit of combining these two sources of information, particularly in how they together can inform us on the dark energy EoS parameter.

The cross-correlation of redshift-binned maps of galaxy number densities with CMB lensing is very useful due to the different ways in which galaxy clustering and CMB lensing are dependent on the galaxy bias. We use the standard definition of the linear galaxy bias as the ratio of the overabundance of galaxies to the overdensity of mass, $b(z) = \delta_g(\mathbf{r})/\delta(\mathbf{r})$; $\delta_g(\mathbf{r}) = (n(\mathbf{r}) - \bar{n})/\bar{n}$, where $n(\mathbf{r})$ is the density of galaxies at location $\mathbf{r}$ and $\bar{n}$ is its spatial average, and $\delta(\mathbf{r}) = (\rho(\mathbf{r}) - \bar{\rho})/\bar{\rho}$, where $\rho(\mathbf{r})$ is the mass density at a location $\mathbf{r}$ and $\bar{\rho}$ is its spatial average. We can then determine such linear and scale-independent bias, to within noise limitations, as the ratio between angular power spectra, $b_i \simeq C_l^{g_i g_i}/C_l^{\kappa g_i}$, where i runs over tomographic redshift bins. With improved constraints on galaxy bias at various redshifts, we can break the degeneracy between galaxy bias and the amplitude of the matter power spectrum $P(k, z)$, thereby better constraining the cosmological model parameters [203, 231].

[92] used the high depth and density of the DES survey to construct maps of galaxy number density in several photometric redshift bins. They then cross-correlated these maps with a CMB lensing map inferred from *Planck* and SPT data in order to determine both

the galaxy bias and the CMB lensing amplitude, in a process they called "CMB lensing tomography" (see for example [239, 47, 206] for early work and [209, 194, 152, 179, 69, 107, 145, 151, 89, 60] for more recent analyses).

CMB lensing tomography provides us with a means, complementary to tomographic cosmic shear, of reconstructing the mass distribution across the sky in coarse slices in redshift. Here we study the role, in reaching constraints on dark energy parameters, of not just the amplitude as a function of redshift, but also the shape of the matter power spectrum, and the distance-redshift relation that influences observables that are all seen in projection.

In this chapter we focus on the constraints on dark energy that can come from the CMB lensing tomography enabled by CMB-S4 lensing maps and LSST galaxy clustering. Current SNRs for the best-measured modes in CMB lensing maps are quite modest. The Planck lensing map [7], has a SNR per mode (on spherical harmonic modes with multipole moment l) approximately equal to 1 for $l \simeq 50$, and lower everywhere else. From CMB-S4 we expect SNRs of greater than unity for all modes with $l \lesssim 1000$ and as large as $\simeq 40$ for the best-measured modes. This increase in CMB lensing precision, together with LSST galaxy clustering, opens up the possibility of measuring the amplitude of structure to a high precision over a range of redshifts [4].

The roles of "geometry" (the distance-redshift relation) and "growth" (the amplitude of the matter power spectrum as a function of time) have been well-studied in the case of tomographic cosmic shear [3, 240, 290, 147, 287, 288, 181, 289]. Although often described as a probe of growth, distinguishing it from purely geometric probes such as the use of SNeIa as standard candles, these studies clarify that geometry is just as important as growth, if not more so, for constraints on dark energy parameters.

Several forecasts have been done for cosmological parameter sets which include Σm_ν and w, through several combinations of observables that include WL, high-SNR CMB lensing, and galaxy clustering. Early forecasts which included either WL or CMB lensing either did not include galaxy custering [141, 108, 186, 278], or did not include the cross-correlation between CMB lensing and galaxy clustering [225, 150, 79]. More recently, studies have gone to the opposite extreme. That is, recent studies have included the cross correlation between CMB lensing and galaxy clustering, as part of a robust and inclusive forecast that also includes the cross-correlations between cosmic shear and galaxy clustering, and between cosmic shear and CMB lensing. However, these studies did not attempt to forecast the benefits of the CMB lensing - galaxy clustering cross-correlation, without also including WL cross-correlations [137, 183, 230].

Two studies which we follow very closely are those of [231] and [283], in which we presented forecasts of Σm_ν that include CMB-S4 lensing, LSST galaxy clustering, and their cross-correlation, but did not forecast the dark energy figure of merit. The forecast of [231] includes a forecast of σ_8, the linear theory RMS of the mass distribution on scales of 8 Mpc/h, which effectively serves as a proxy for the amplitude of the matter power spectrum.

[229] considered the impact of photometric redshift uncertainties and the potential for self-calibration in a setup similar to ours. That work shows that the fraction of photometric redshift outliers can be constrained by the data itself, and it considers the effect of dynamical

Dark Energy and neutrinos, similar to this forecast. [83] explores constraints from cosmic shear, clustering and CMB lensing, including the effect of baryons and photo-z outliers. However, the physical origin of the constraints on Dark Energy, the main goal of the present chapter, was not explored in these works.

The remainder of our chapter is organized as follows: In Section 3.2, we present our cosmological model space and assumptions about observables and noise. Consistent with our focus on understanding the physics that leads to the forecasted constraints, our modeling of the data is quite simple, and in particular does not include sources of systematic error. In Section 3.3 we describe our forecasting formalism and in Sections 3.4 and 3.5 we present and discuss our results. Note that we discuss the physical origin of dark energy constraints in Section 3.2 and 3.4.

3.2 Model and Assumptions

In chapter 2, we assume the ΛCDM model extended to include massive neutrinos, and here we further extend it to include time-varying dark energy. The cross-correlations are computed in spherical harmonic space and take the form of angular power spectra.

We assume a fiducial model with the following parameters: $\Omega_b h^2 = 0.02226$, $\Omega_c h^2 = 0.1193$, $\tau = 0.063$, $A_s = 2.130 \times 10^{-9}$, $n_s = 0.9653$, $\theta_{MC} = 1.04087 \times 10^{-2}$, and $\Sigma m_\nu = 0.06$ eV. The neutrinos in this model are relativistic in the early universe and slow down as the universe expands, becoming non-relativistic at late times. We add time-varying dark energy by considering the dark energy equation of state parameter w to follow the common parameterization [61, 170] $w(a) = w_0 + (1-a)w_a$, where a is the scale factor of the expansion of the universe normalized such that $a = 1$ today, while noting that different parameterizations are possible [133, 78, 25, 63]. The fiducial values that we use are $w_0 = -1$, $w_a = 0$. We also include one galaxy bias parameter for each galaxy bin, as described in section 3.2.

Galaxy Binning

Following chapter 2, we employ two different assumptions for LSST galaxy redshift distributions in our analysis, as shown in figure 2.1. The first redshift distribution is the $i < 25$ Gold sample [6], which takes the analytic form $dN(z)/dz \propto (1/2z_0)(z/z_0)^2 e^{-z/z_0}$, with $z_0 = 0.3$ and a corresponding galaxy solid angle number density of $\bar{n} = 40$ arcmin^{-2}, and is hereafter referred to as the "LSST Gold" sample. The second distribution, which we refer to as the "LSST Optimistic", assumes a fainter observable magnitude limit of $i < 27$ with $S/N > 5$ in the i band and includes Lyman break galaxies from redshift dropouts, and this results in the increase the number density of observable galaxies to $\bar{n} = 66$ arcmin^{-2} [231]. This large number density may not be overly optimistic for our purposes as we do not rely on galaxy shape measurements, but just galaxy locations.

We divide LSST galaxy redshift distributions into 16 non-overlapping tomographic bins with the bin edges of z = [0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.3, 2.6, 3, 3.5, 4, 7].

Defining redshift bins with perfectly sharp edges in an actual galaxy survey, however, is not currently achievable, as the high number of galaxies can only have their redshifts determined photometrically, with an associated photometric redshift error. Thus our forecasts describe the ideal case in which the photometric redshift ("photo-z") error has been completely eliminated. This is a conceptually simpler case than one in which the redshift bins are more realistic and contain the photo-z error, and we leave the forecasting that includes such error to a future study.

Likewise, our idealized treatment neglects a number of effects that should be included in more realistic forecasts. These include effects of galactic dust, photometric redshift errors (e.g. [116, 229]), galaxy-galaxy blending [113], and magnification-induced correlations across redshift bins (e.g. [93]). Also, we neglect any non-Gaussian corrections to the covariance matrix [260, 150]. As stated above, we find that this simplified setting is helpful to elucidate the physical origin of the constraints, and we don't expect our conclusions to change when considering a more realistic forecast.

We use galaxies as tracers of matter and assume a single galaxy population. We use a linear galaxy bias model, which [67] show to be valid (in the Dark Energy Survey) at least on the scales where the linear growth of structure is a sufficiently accurate approximation. Similarly, we'll restrict our analysis to large scales (defined below), and use the linear matter power spectrum. More sophisticated modeling of non-linearities in matter and bias will be required in a more realistic analysis (see for example [184, 151, 145, 201]). Following [6] and [231], we assume $b(z) = 1 + z$ as our fiducial bias evolution.

For each redshift bin, we define a bias parameter calculated as a weighted average of this galaxy bias function over the redshift range of the bin:

$$b_i = \frac{1}{\left[\int \frac{dN_i(z')}{dz'} dz'\right]} \int \frac{dN_i(z)}{dz} b_i(z) dz, \tag{3.1}$$

where $b_i(z) = B_i b(z)$. B_i is effectively an amplitude of the bin bias, and $b(z)$ is the redshift dependent galaxy bias function. The B_i parameters are the ones that we use in our Fisher forecasting, with fiducial values of $B_{i,fid} = 1$. Similarly, $b_{i,fid} = b_i(B_{i,fid})$, which leads to $\sigma(B_i) = \sigma(b_i)/b_{i,fid}$ in the Fisher results.

Theoretical Power Spectra

With the CMB lensing convergence κ and a tomographic set of galaxy distribution map, we compute the following 2-point angular power spectra: $C_l^{\kappa\kappa}$, $C_l^{\kappa g_i}$, and $C_l^{g_i g_i}$, where g_i is the galaxy density field in the ith tomographic redshift bin.

The CMB lensing convergence in direction $\hat{\mathbf{n}}$ can be calculated as a line-of-sight integral over the fractional matter over-density $\delta(\mathbf{r}, z)$ at the comoving position $\mathbf{r}$ and redshift z:

$$\kappa(\hat{\boldsymbol{n}}) = \int d\chi W^{\kappa}(\chi) \delta\big(\chi\hat{\boldsymbol{n}}, z(\chi)\big), \tag{3.2}$$

where χ is the the comoving distance, and the lensing distance kernel is [66, 255, 47]:

$$W^{\kappa}(\chi) = \frac{3}{2}\Omega_m H_0^2 \frac{\chi}{a(\chi)} \frac{\chi_{\rm CMB} - \chi}{\chi_{\rm CMB}}, \tag{3.3}$$

where Ω_m is the matter fraction at the current time, H_0 is the current value of the Hubble parameter, $a(\chi)$ is the scale factor at comoving distance χ, and $\chi_{\rm CMB}$ is the comoving distance of the CMB's surface of last scattering.

For the galaxy density field in the ith bin, we can also calculate the following line-of-sight integral:

$$g_i(\hat{\boldsymbol{n}}) = \int d\chi W^{g_i}(\chi)\delta\big(\chi\hat{\boldsymbol{n}}, z(\chi)\big), \tag{3.4}$$

where the galaxy distance kernel is [47]:

$$W^{g_i}(\chi) = \frac{1}{\big[\int dz' \frac{dN_i(z')}{dz'}\big]} \frac{dz}{d\chi} \frac{dN_i(z)}{dz} b_i(\chi). \tag{3.5}$$

Note that we neglect the magnification bias.

Using the Limber approximation [167, 140], we model the angular power spectrum as:

$$C_l^{\alpha\beta} = \int dz \frac{d\chi}{dz} \frac{1}{\chi^2} W^{\alpha}(\chi) W^{\beta}(\chi) P_{\delta_\alpha \delta_\beta}\Big(\frac{l+1/2}{\chi}, z(\chi)\Big) \tag{3.6}$$

where $\alpha, \beta \in (\kappa, g_1, ..., g_N)$, and $P(k, z)$ is the matter power spectrum at wavenumber k and redshift z. Following [231] and [283], we use the CDM-baryon density contrast δ_{cb} for galaxy clustering and the total matter density contrast $\delta_{cb\nu}$ (which includes neutrinos) for lensing.

We use the publicly available CAMB code [161] and its Python wrapper [123] in order to calculate $P(k, z)$, as well as the unlensed primary CMB power spectra C_l^{TT}, C_l^{TE}, and C_l^{EE}. For such calculations, we assume the normal hierarchy, in which the third neutrino mass eigenstate ν_3 is heavier than the other two eigenstates. We use the fluid dark energy model implemented in the python wrapper of CAMB, available since its major update in version 1.0.

As in chapter 2, we impose the limit on the maximum wavenumber to be included in our analysis (see Table 3.1 for the corresponding $l_{\rm max}$ values of each redshift bin) so that perturbations can be assumed to remain in the linear regime, and therefore uncertainties due to non-linear modeling have negligible effects on our forecasts. The vertical lines in Figure 3.1 indicate the scale cuts assumed for the galaxy auto-power spectrum in different tomographic redshift bins. All forecasts presented in this work assume the linear matter power spectrum, as we find that the effects of adding non-linear corrections from HALOFIT to the power spectrum are only negligible with $k_{\rm max} = 0.1$ or 0.2 hMpc^{-1} imposed.

Separating Impacts of Distance, Growth, and Shape

The EoS parameter of dark energy affects both the cosmic distance scale and the growth factor. The relative importance of these two effects on constraining the EoS parameter has

Table 3.1: $l_{\rm max}$ values corresponding to two $k_{\rm max}$ limits, 0.1 and 0.2hMpc^{-1}, for the left edge of each tomographic redshift bin. ($l_{\rm max}$ is therefore set to be zero for the first bin.) We assume $l_{\rm min} = 30$ to account for the expected difficulty of attaining low-noise data on large angular scales.

$k_{\rm max}$ \ bin	1	2	3	4	5	6	7	8
0.1	0	57	108	153	193	229	261	289
0.2	0	114	216	307	387	459	522	579
$k_{\rm max}$ \ bin	9	10	11	12	13	14	15	16
0.1	315	338	358	386	411	439	469	495
0.2	630	676	717	773	822	879	939	991

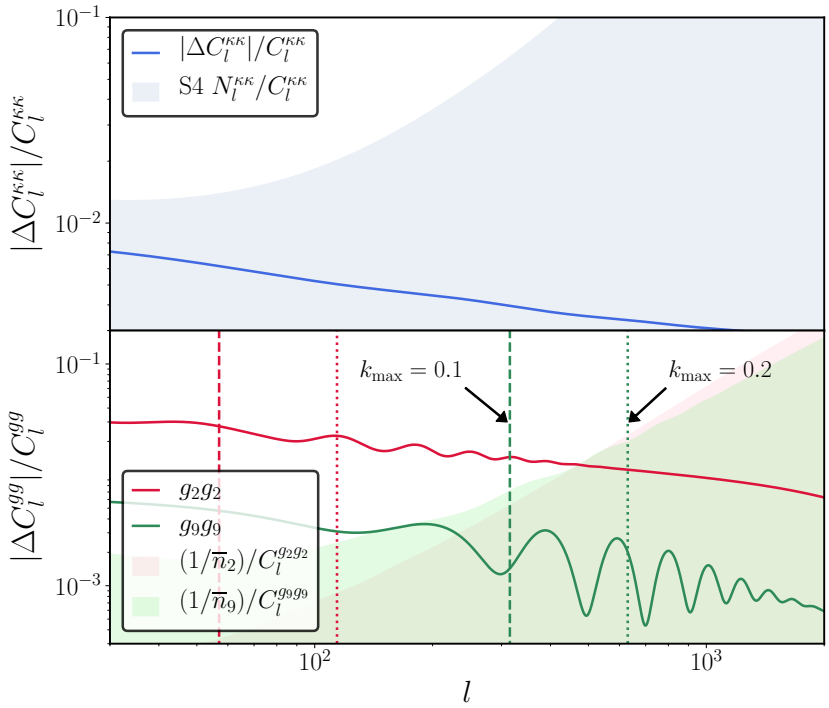


Figure 3.1: Fractional change of the auto-power spectra $C_l^{\kappa\kappa}$ and C_l^{gg} with respect to Δw_0: $|\Delta C_l|/C_l = \frac{1}{C_l}|\Delta w_0 \times \partial C_l/\partial w_0|$, where $\Delta w_0 = 0.05$. $\theta_{\rm MC}$ is held fixed. *Top*: Comparison between the S4 lensing reconstruction noise (blue shaded region) and the changes in the CMB lensing auto-power spectrum with respect to Δw_0 (blue curve). *Bottom*: The galaxy shot noise and the changes in the galaxy auto-power spectra with respect to w_0, in the 2nd (red) and 9th (green) tomographic redshift bin. Also shown are vertical lines indicating the $l_{\rm max}$ values for each redshift bin, corresponding to the two $k_{\rm max}$ values indicated in Table 3.1.

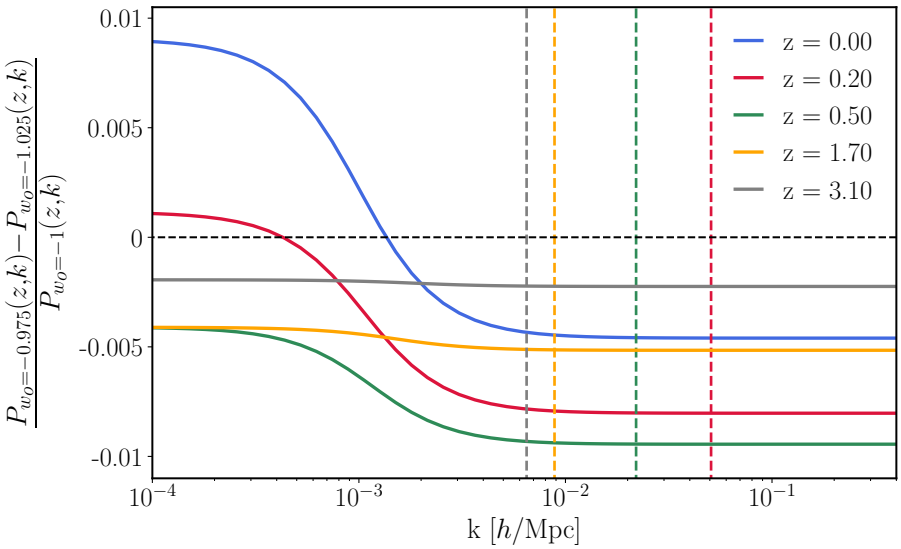


Figure 3.2: Fractional change of the matter power spectrum $P(z,k)$ with respect to w_0 ($\Delta w_0 = 0.05$), $\Delta P(z,k)/P(z,k)$, for five different redshifts within the range of our analysis. $\theta_{\rm MC}$ is held fixed. To preserve the distance to the last scattering surface, The vertical dashed lines correspond to the low-l cutoff ($l_{\rm min} = 30$) used in the Fisher forecasts, de-projected to the redshifts (from the right) 0.2, 0.5, 1.7, and 3.1. With the resulting k-limits, we remove dark energy perturbations on large scales from the analysis, thereby making the power spectrum shape effects negligible.

been discussed for WL by several groups [3, 240, 290, 147, 287, 288, 181], using various methods. As was shown by [240] and by [181], there is a partial cancellation of the geometry and growth effects for WL observables when w is varied.

In Figure 3.1 and 3.2, we vary w_0 while keeping the angular size of the sound horizon $\theta_{\rm MC}$, Ω_b, and Ω_m fixed in order to make minimal changes to primary CMB power spectra. In this scenario, if we increase w from its fiducial value, the dark energy density decreases with time. In order to keep $\theta_{\rm MC}$ fixed, and therefore the angular-diameter distance to last-scattering fixed, we increase the dark energy density at high redshifts. The result is that, compared to the fiducial model, with w_0 increased $H(z)$ is decreased at $z \lesssim 0.9$ and gently increased at $z \gtrsim 0.9$, asymptoting to zero increase deep in the dark-matter-dominated regime. One result is that the distances to all redshifts at $z \lesssim 1100$ are increased, asymptoting to zero change deep in the dark-matter dominated regime. Another is that at $z \gtrsim 0.9$, growth is slowed down. The impact on the growth reverses when $H(z)$ starts to become less than in the fiducial model at $z \simeq 0.9$. Figure 3.2 indeed shows $\Delta P(k) < 0$ for $\Delta w > 0$, with more power suppression at $z \approx 0.5$ than for any of the other redshift choices shown in the figure. On the other hand, increasing w leads to positive growth function, $D(z) > 0$, at all redshifts (asymptoting to zero at high redshift), thereby increasing the angular power spectrum C_l. We find that these two effects partly cancel each other leading to weaker cosmological constraints

than if growth and distance were measured separately, as we shall see below. Note that in Figure 3.2, $P(k)$ is suppressed at all redshifts, and therefore there is partial cancelation in all redshift bins.

One way of comparing the relative importance of distance and growth on constraining w in a Fisher forecast is to split w into two components: w_d, which only affects the distance, and w_g, which affects the growth only. Previous work has used this parametrization to distinguish between the two effects and systematic errors [290], or to investigate which has more constraining power [240, 288]. [287] extends this idea of splitting the parameter w to examine constraints on (w_0, w_a) for distance-only, growth-only, and full (growth+distance) cases, finding that the growth-only case has the least constraining power; with other cosmological parameters marginalized, the full case has the most constraining power, while fixing them make the distance-only case more powerful.

We instead apply a similar procedure to investigate the relative importance of distance and growth in our forecasts, but we take an additional step of separating the growth effect further into two separate components: the growth of amplitude of the power spectrum and the change in the shape of the power spectrum due to growth. The reason for such decomposition is the degeneracy between the power spectrum amplitude and the galaxy bias, and we break this degeneracy by incorporating observables which have different dependencies on galaxy bias[2].

Figure 3.2 shows the fractional change in $P(z, k)$ due with respect to Δw_0 for five different values of redshift. We only show the plots for w_0, but w_a derivatives are very similar in appearance. The absence of scale-invariance seen in this figure is a generic effect associated with $w \neq -1$ dark energy models, as pointed out by e.g. [30] and [271], who showed that dark energy perturbations appear on very large scales, depend in particular on the sound speed, and are model dependent.

At first glance, Figure 3.2 looks as if ignoring the shape of w derivatives is effectively flattening the large low-k features in the derivatives. However, these features occur on scales much larger than the maximum angular scale we include in the analysis. Due to the possibility of large-scale systematics, we also impose a low-l cutoff of $l_{\rm min} = 30$ for each redshift bin. We check that with the resulting k-limits, shown as vertical lines in Figure 3.2, we remove large perturbative features from the w derivatives. The portion of the power spectrum greater than these cut-off points differs from the shape-less version (identical to the value at $k = 0.01h{\rm Mpc}^{-1}$) by only several hundredths of a percent. Due to the smallness of this feature, we can safely neglect it.

When we calculate the partial derivatives of the observables with respect to the w_0, w_a parameters, $P(k, z)$ in each tomographic redshift bin is fixed to the value of $P(k, z_{\rm med})$, where $z_{\rm med}$ is the median redshift of each bin, to remove $w(a)$-dependent variations of the power spectrum across the width of each bin. Such variations, degenerate with the evolution of bias across the bin, can make our results artificially sensitive to those changes, especially given

[2]CMB lensing has no galaxy bias dependence whereas galaxy observations do, which allows the combination of the various observables $C_l^{g_i g_i}$, $C_l^{\kappa g_i}$, and $C_l^{\kappa\kappa}$ to distinguish a change in amplitude from the galaxy bias.

that we fix the bias evolution within each bin. We find that overlooking such procedure can lead to artificially rosy forecasts, increasing the constraining power by tens of percent.

3.3 Fisher Forecasting

We use the Fisher information matrix formalism to forecast constraints on the cosmological parameters of interest [267, 28].

Observables

We forecast the constraining power of cross-correlating CMB-S4 lensing with the galaxy clustering tomography observations of Rubin Observatory LSST (similar to [92]), and our observables are auto- ($C_l^{\kappa\kappa}$ and $C_l^{g_i g_i}$) and cross-spectra ($C_l^{\kappa g_i}$) from Section 3.2. We do not include $C_l^{g_i g_j}$, cross-spectra of galaxy tomographic bins, nor do we include the cross-spectrum $C_l^{T\kappa}$, which would be nonzero at low l due to the gravitational Integrated Sachs-Wolfe (ISW) effect.

For CMB lensing, we assume a CMB-S4 experiment with the telescope beam of Full-Width-Half-Maximum (FWHM) of $1'$ and a white noise level of $1\mu K'$ for temperature and $1.4\mu K'$ for polarization. We assume $f_{\rm sky} = 0.4$ and set the noise levels N_l^{TT}, N_l^{EE} in the primary CMB as a Guassian noise as:

$$N_l^{XX} = s^2 \exp\Big(l(l+1)\frac{\theta^2_{\rm FWHM}}{8\log 2}\Big), \tag{3.7}$$

where XX stands for TT or EE, s is the total intensity of instrumental noise in μKrad, and $\theta^2_{\rm FWHM}$ is the FWHM of the beam in radians [278]. For the CMB lensing reconstruction noise, we use the EB quadratic estimator method described in [127], implemented by the QUICKLENS [124] software package. Following [231], we rescale the EB noise to approximately match the expected improvement from iterative lens reconstruction for CMB-S4 [118, 252].

For the LSST, we assume that the survey covers an area on the sky of 18,000 deg^2, corresponding to $\approx 40\%$ of the sky, and that it fully overlaps with CMB-S4. The shot noise associated with the galaxy redshift distributions is $1/\bar{n}_i$, where $\bar{n}_i$ is the galaxy number density per redshift bin, calculated per bin from the ratio of the integrated area of dN_i/dz to that of the total $dN(z)/dz$ multiplied by the overall galaxy density $\bar{n}$.

Fisher Matrices

Assuming our observables from Section 3.3 are the power spectra of Gaussian random fields, we can compute the covariance matrix as:

$$\mathrm{Cov}(C_l^{\mu_1\nu_1}, C_{l'}^{\mu_2\nu_2}) = \frac{\delta_{ll'}}{(2l+1)f_{\rm sky}}\Big(C_l^{\mu_1\mu_2}C_l^{\nu_1\nu_2} + C_l^{\mu_1\nu_2}C_l^{\nu_1\mu_2}\Big), \tag{3.8}$$

	ΛCDM + m_ν free		+ w_0, w_a free									
(+ S4/Planck T&P)	$\sigma(\Sigma m_\nu)$ [meV]		$\sigma(w_0)$		$\sigma(w_a)$		$\sigma(w_p)$		FoM		$\sigma(\Sigma m_\nu)$	
S4Lens	69		0.25		0.95		0.14		7.5		83	
	$k_{\max} = 0.2$	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1
S4Lens + LSST Gold	36	50	0.12	0.18	0.37	0.50	0.050	0.085	54	23	55	61
+ DESI BAO	22	23	0.093	0.11	0.28	0.31	0.026	0.029	138	111	39	41
S4Lens + LSST Optimistic	31	41	0.10	0.15	0.33	0.46	0.042	0.077	71	28	48	54
+ DESI BAO	21	22	0.085	0.11	0.26	0.31	0.024	0.028	159	117	37	40

Table 3.2: Forecasts of the neutrino mass and dark energy constraints, for different experiment configurations and $k_{\max}$ limits. The first two columns assume the dark energy parameters are held fixed, while marginalizing over six ΛCDM parameters and linear bias amplitudes in tomographic bins, and the rest of the columns include w_0 and w_a as free parameters.

where $(\mu_1, \mu_2, \nu_1, \nu_2) \in \{\kappa, g_1, ..., g_N\}$. We assume that each C_l contains both signal and noise.

Then, the Fisher matrix is given by:

$$F_{ij} = \sum_{\substack{\mu_1,\nu_1,\\ \mu_2,\nu_2}} \sum_l \frac{\partial C_l^{\mu_1\nu_1}}{\partial\theta_i} \left[\mathrm{Cov}(C_l^{\mu_1\nu_1}, C_{l'}^{\mu_2\nu_2})\right]^{-1} \frac{\partial C_l^{\mu_2\nu_2}}{\partial\theta_j}, \tag{3.9}$$

where $\vec{\theta}$ is a set of cosmological model parameters and the bias amplitude parameters, B_i, from Section 3.2. We can combine the Fisher matrix in equation 3.9 with external datesets, such as the primordial CMB and BAO Fisher matrices, if needed and then invert the resulting matrix to determine the marginalized constraints on the parameters of our interest.

3.4 Forecast results

We frame our w_0, w_a forecasts in terms of the Dark Energy Task Force (DETF) Figure of Merit (FoM) [17], defined as the inverse of the area of an error ellipse in the w_0, w_a plane. Hence, a higher FoM corresponds to a smaller error.

Table 3.2 presents constraints on the the neutrino mass and dark energy equation of state, marginalized over ΛCDM parameters and linear galaxy bias amplitudes in all bins, for different experiment configurations and $k_{\max}$ limits. All forecasts include the primordial CMB information. We find that the relative merit of cross-correlating CMB lensing with galaxy clustering is huge; with $k_{\max} = 0.2h\mathrm{Mpc}^{-1}$, combining the galaxy clustering from the LSST Optimistic sample and CMB-S4 lensing can achieve the FoM of 71.

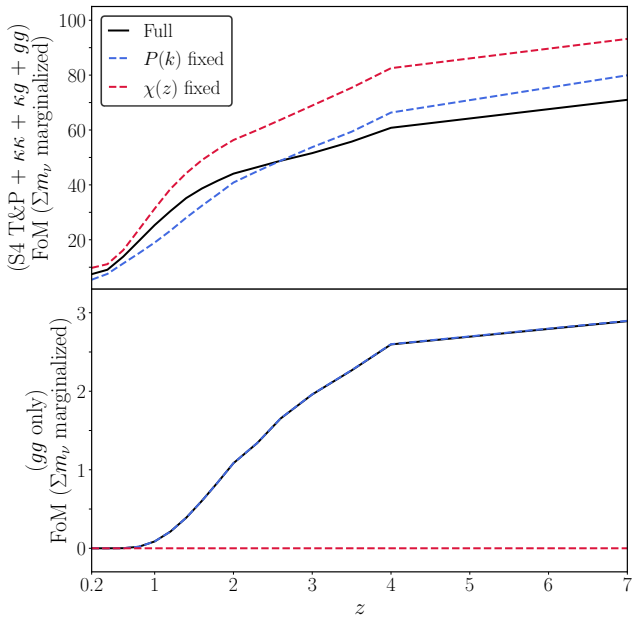


Figure 3.3: Forecasted DETF Figure of Merit, defined as $[\sigma(w_a)\sigma(w_p)]^{-1}$, with changes to either growth (blue) or geometry (red) disregarded. We also include the results with the "full" Fisher matrix (black), where we apply no restrictions (all effects included), for comparison. *Top*: FoM from the combination of S4 primordial CMB, S4 lensing, and LSST clustering (using Optimistic $dN(z)/dz$ with $k_{\rm max} = 0.2h{\rm Mpc}^{-1}$). We observe the partial cancellation between growth and geometry. *Bottom*: Only with LSST galaxies, ignoring changes to geometry (by fixing $\chi(z)$) makes the constraining power negligibly small.

In Figure 3.4, we show the DETF FoM with the sum of neutrino mass either marginalized (top panel) or fixed (middle and bottom). All forecasts demonstrate that including the galaxies in higher redshift bins significantly improves the constraints by extending the redshift lever arm and increasing the volume probed. We also show that with the DESI BAO measurements combined, a notable improvement in the FoM is acheived.

Growth or Geometry?

To investigate how sensitive our forecasts are to the distance-redshift relation (geometry) and the amplitude of the matter power spectrum as a function of redshift (growth), we make the following different types of forecasts: "$P(k)$ fixed", for which we do not allow the power spectrum to change as w changes, "$\chi(z)$ fixed", for which we do not let the distance-redshift relationship change as w changes, and "Full" for which we apply no restrictions.

Figure 3.3 shows the result of FoM forecasts for all three cases, using the LSST Optimistic $dN(z)/dz$ with $k_{max} = 0.2h{\rm Mpc}^{-1}$. We present FoM values as functions of redshift: at each redshift, only bins at and below that redshift are included, and FoM increases as we extend

the redshift lever arm and thereby include more galaxies, as expected.

The top panel of Figure 3.3 shows the forecasts of CMB lensing and LSST clustering combined. We note that both geometry and growth play a significant role, as can be seen by comparing the red and blue curves; the "Full" Fisher matrix (black curve), which include all of the effects, appears to have less constraining power than either geometry or growth, suggesting that there is a partial cancellation at play. A similar cancellation was noted for WL observables in [240, 181] and we find that this applies to clustering measurements as well. In short, we note that the partial cancellation between growth and geometry effects that has been noticed before also appears in our S4 lensing + LSST forecasts.

We also find that inclusion of CMB lensing increases the FoM by a factor of 3-4 (over the S4 primary CMB + gg result not shown in Figure 3.3), suggesting that cross-correalations between CMB lensing and galaxy clustering provide a very competitive dark energy probe. The bottom panel shows that with only LSST galaxies, "$\chi(z)$ fixed" case has a negligibly small constraining power, as we cannot gain any dark energy information if the distance-redshift relation is fixed, and the amplitude of power spectrum is degenerate with bias.

Comparisons with Galaxy Weak Lensing Forecasts

[286] presents LSST Cosmic Shear + *Planck* forecasts of $[\sigma(w_a)\sigma(w_p)]$, for various levels of photometric redshift error. Our forecasts assume zero uncertainty in the redshifts of the observed galaxies. Our redshift bin widths are $\Delta z = 0.2$ for the lowest redshift bins up until $z = 2$, then $\Delta z = 0.3 - 0.5$ out to redshift $z = 4$, and the final bin width of $\Delta z = 3.0$ from $z = 4$ to $z = 7$. To reduce the sensitivity to photo-z errors, we make the bin widths wider than the expected rms scatter in photo-z errors, but we acknowledge that the effects of photo-z errors are not entirely eliminated. We leave an analysis of quantifying such impacts to future work. [286] provides $[\sigma(w_a)\sigma(w_p)]$ as a function of $\sigma_z/(1+z)$, where σ_z is the rms photometric redshift error. For simplicity, we compare against the FoM corresponding to two specific values of redshift error: FoM ≈ 91 for $\sigma_z/(1+z) = 0$ and FoM ≈ 67 for $\sigma_z/(1+z) = 0.05$, and these forecasts appear as black horizontal lines in the middle panel of Figure 3.4, labeled as "Planck T&P + LSST Cosmic Shear." We find that our results are at a similar level to these cosmic shear forecasts. We also note that the middle and bottom panels of Figure 3.4 provides the FoM forecasts with fixed Σm_ν to make a fair comparison to the forecasts in [286].

However, the forecasts in [286] assume slightly different choices for the survey characteristics. For the LSST specifications, [286] assumes $f_{sky} = 0.48$ and uses a full-survey galaxy number density of $\bar{n} = 50$ galaxies/arcmin^2, while we use $f_{sky} = 0.4$ and $\bar{n} = 40$ galaxies/arcmin^2. Moreover, [286] uses a fiducial bias function of $b(z) = 1 + 0.84z$, whereas we use $1 + z$. The differences in $\bar{n}$, f_{sky}, and $b(z)$ should each affect the values of the FoM, but only to a small degree. Changing our forecast parameters to more closely match [286]'s would change our forecasted FoM values somewhat.

As an alternative way to compare our forecasts with the LSST cosmic shear forecast, we also present 1σ error ellipses in the w_0-w_a plane. Figure 3.5 includes our forecasts with

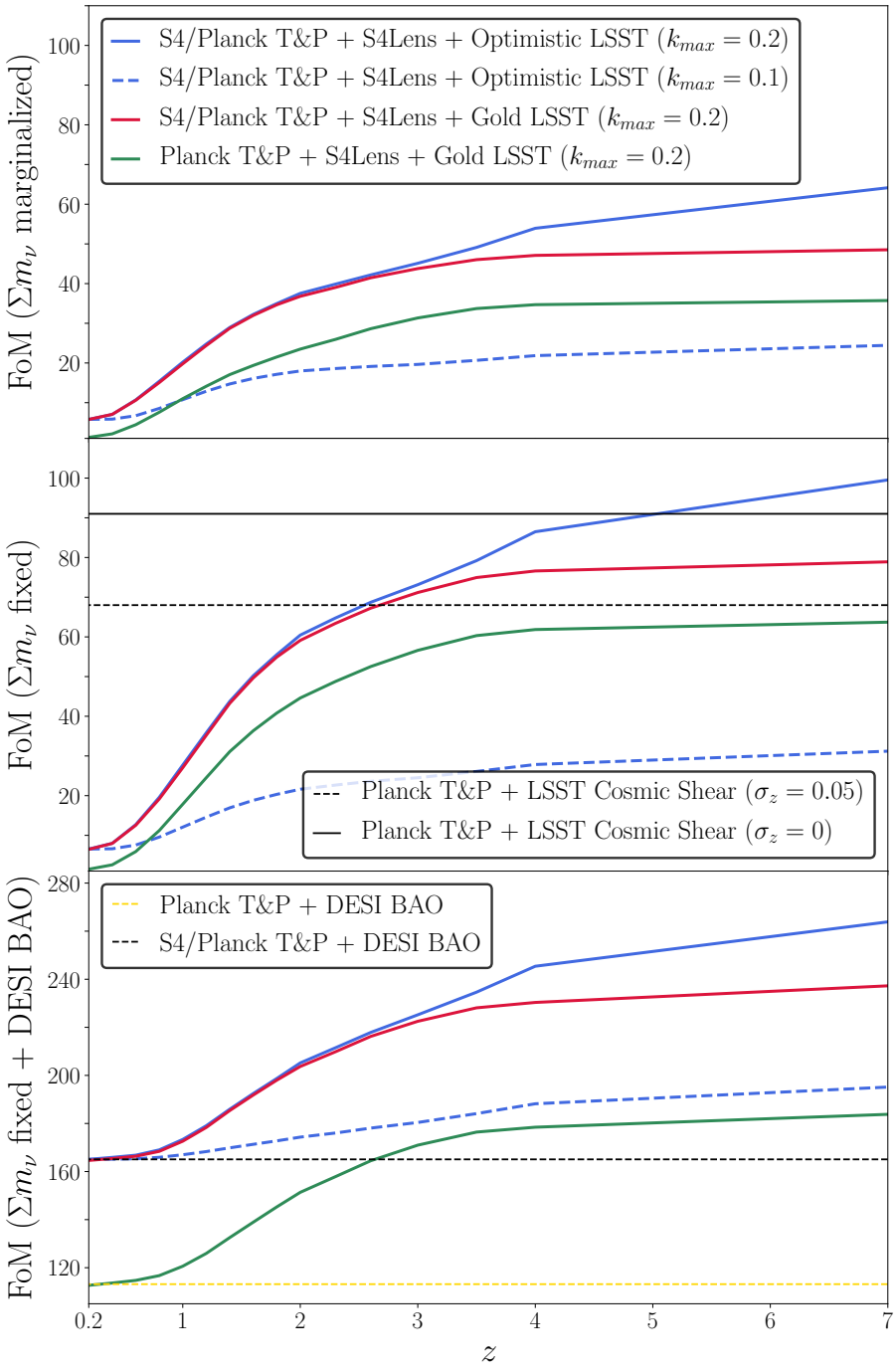


Figure 3.4: Forecasted DETF Figure of Merit for different experiment configurations and $k_{\max}$ limits. *Top*: Forecasts with the neutrino mass sum marginalized. Addition of galaxy bins at higher redshift extends the redshift lever arm, resulting in a greater constraining power. *Middle*: Forecasts with the neutrino mass sum fixed. Our results are at a similar level to the forecast with LSST weak lensing combined with Planck measurements [286] (black). *Bottom*: Forecasts with the DESI BAO measurements included. With the S4 primary CMB data, we gain a noticeable improvement in forecasts relative to the Planck data.

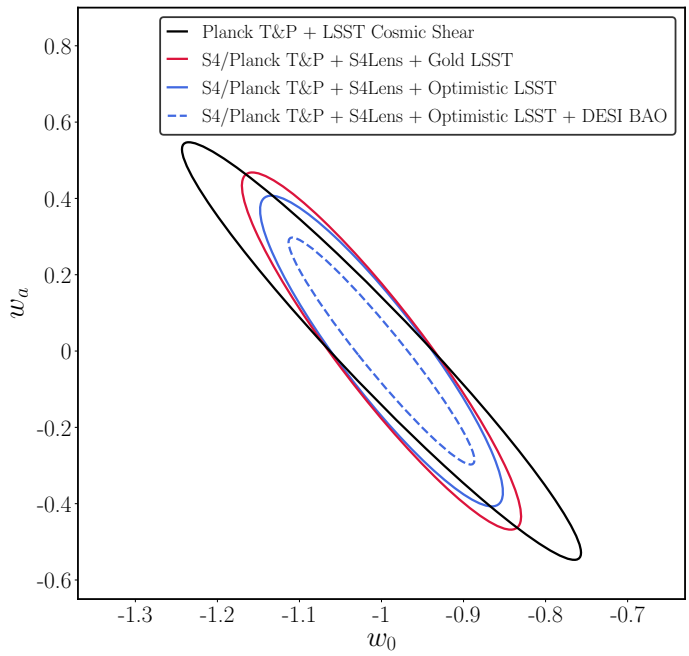


Figure 3.5: 1σ confidence ellipses in the $w_0 - w_a$ plane, with different survey configurations. We observe that dark energy constraints from LSST clustering in combination with CMB-S4 lensing (blue and red) are comparable to those from the LSST cosmic shear data (black).

different survey configurations (assuming $k_{max} = 0.2h\text{Mpc}^{-1}$), for both the LSST Gold and Optimistic samples. Plotted with them is the LSST cosmic shear forecast from [289], and this includes anticipated systematics, such as additive and multiplicative errors in the shear power spectra and uncertainty in the photometric redshift error distribution. This forecast is similar to the one in [286], though the LSST data model is updated, but it does not include galaxy clustering power spectra nor galaxy-galaxy lensing power spectra. Its error ellipsis appears to have a similar size as our forecast ellipses, consistent with the result shown in Figure 3.4.

3.5 Summary

We have studied the prospect for CMB Lensing Tomography to constrain the dark energy parameters by combining LSST redshift-binned galaxy clustering maps and CMB-S4 convergence map. Although one might expect that the dominant contribution to dark energy constraints would come from the determination of the matter power spectrum as a function of redshift, the observable statistical properties that we consider (auto- and cross-power spectra) are also sensitive to the distance-redshift relation. We find that a comparable amount of information about $w(a)$ comes from geometrical features as comes from growth.

To conduct such analysis, we need to take care not to artificially break degeneracies of galaxy biasing with the amplitude of the matter power spectrum. This danger is present due to the low dimensionality of our parameterization of the bias-redshift relation. In our analysis, we use one parameter for each bin, with a precisely fixed and known redshift dependence within each bin. To avoid an artificial breaking of degeneracy, we remove variation of the redshift dependence of $P(k, z)$ within a redshift bin. As the dark energy equation of state parameters vary, we adjust the amplitude of $P(k, z)$ at the center of each redshift bin, while keeping the shape unchanged.

We find that large angular scales are particularly important for the study of dark energy. Uncertainties associated with non-linear evolution and galaxy biasing on small scales lead us to ignore, in our forecasting, wavenumbers larger than k=0.2 h/Mpc. Another approach would be to increase the maximum k value as redshift increases, since the scale of non-linearity moves out to higher k. However, at higher redshift, the mean bias factor of the galaxies in the catalog increases, reducing tolerance to errors in the modeling of galaxy bias [184]. The choice of fixed maximum k means that there is effectively a maximum value of l for each redshift bin, which is an increasing function of redshift.

Finally, we present the DETF Figure of Merit for different experiment configurations and find that adding CMB lensing information to LSST clustering increases the FoM by roughly a factor of 3-4. We also show that our result is comparable to those from LSST tomographic cosmic shear, suggesting that the combination of CMB-S4 lensing and LSST clustering is a competitive probe of dark energy with very different systematics, and therefore highly complementary to the traditional analyses.

Chapter 4

Disconnected Covariance of 2-point Functions in Large-Scale Structure

Optimal analyses using the 2-point functions of large-scale structure probes require accurate covariance matrices. A covariance matrix of the 2-point function comprises the disconnected part and the connected part. While the connected covariance only becomes important on small scales, the disconnected covariance is dominant on large scales, where the survey window has a significant impact. In this chapter, we develop an analytical method to compute the disconnected covariance, accounting for the window effect. Derived under the flat-sky approximation, our formalism is applicable to wide surveys by swapping in the curved-sky window functions. Our method works for both the power spectrum and the correlation function, and applies to the covariances of various probes including the multipoles and the wedges of 3D clustering, the angular and the projected statistics of clustering and shear, as well as the cross covariances between different probes. We verify the analytic covariance against the sample covariance from the galaxy mock simulations in two test cases: (1) the power spectrum multipole covariance, and (2) the joint covariance of the projected correlation function and the correlation function multipoles. Our method achieve good agreement with the mocks, while at a negligible computational cost. Unlike mocks, our analytic covariance is free of sampling noise, which often leads to numerical problems and the need to inflate the errors. In addition, our method can use the best-fit power spectrum as input, in contrast to the standard procedure of using a fiducial model that may deviate significantly from the truth. We also show that a naive diagonal power spectrum covariance underestimates the signal-to-noise ratio compared to our analytic covariance. [1]

[1]This chapter is taken from "Disconnected Covariance of 2-point Functions in Large-Scale Structure," *Li Y., Singh S., Yu B., and Seljak B.* (arXiv: 1811.05714).

4.1 Introduction

Studies of the large-scale structure of the Universe have made rapid progress over the last two decades especially with high precision measurements using large galaxy surveys. These surveys have provided sensitive tests on the cosmological models using combinations of several independent probes including redshift space distortions (RSD) [43, 37, 213], baryon acoustic oscillations (BAO) [36, 20, 15], weak gravitational lensing [269, 136, 243, 115], supernovae [220] and strong lensing [53]. The next generation of galaxy surveys, LSST [173], Euclid [19], DESI [72], WFIRST [76], aim to make even more precise measurements by observing even larger volume of the universe.

With the increasing precision of the measurements, it is becoming increasingly important to have accurate methods to infer and extract information from these measurements. In this chapter we focus on the probes of matter and galaxy two-point functions. The two-point functions that measure clustering of galaxies or matter are still the primary source to extract cosmological information from the observables. One of the challenges in extracting this information is to accurately model the covariance matrices of these probes in order to obtain unbiased likelihoods of model parameters. Inaccurate covariances make the inferences sub-optimal and in extreme cases also introduce biases in the most likely values of the parameters. Common methods account for the effects of noise in numerical covariances by inflating the errors in the final constraints [112, 75, 238].

There are mainly three common approaches to quantify the covariance, including estimations using mock simulations, internal estimations from data using jackknife or bootstrapping methods, and analytic (or semi-analytic) modeling. Covariance estimation using mock simulations has become the most popular approach as it is in principle possible to include proper treatment of all the observable effects as well as nonlinear and multiscale physics. However, mock covariance require generating many independent realizations and then computing the two point functions over all realizations. This method becomes computationally very expensive, because the required number of simulations scales with the number of data points and with the desired accuracy of the parameter covariance [265]. To overcome these challenges, several fast methods to run approximate mocks have been developed (see e.g. [171, 49] for comparison of covariances from different mocks). Nonetheless, lack of correct physics in these mocks potentially leads to biased covariances [29]. In addition, it is difficult to guess a priori the right fiducial model at which the mocks should be generated. Any deviation of the model from the truth introduces errors in the covariance, either underestimating or overestimating the covariance depending on the relation of the fiducial model to the truth.

Computing covariance directly from the data has the potential to overcome these challenges as it includes all possible observational effects and can be estimated with relatively small computational overhead. This approach requires splitting the data into smaller mutually exclusive subsets, and the number of subsets required depends again on the number of data points and the desired accuracy of parameter covariance. However, splitting data into smaller subsamples limits the largest scales that can be used in the analysis. Furthermore, the subsamples are not totally independent (a requirement by the jackknife and bootstrap

methods), and as a result they bias the covariance [188]. They are also biased by their subvolume windows that are different from the survey window.

Compared to the mock and internal methods described above, calculating the covariance matrices analytically has multiple advantages: its predictions are noiseless; it can be evaluated at the best-fit cosmology with an iterative procedure, and it is computationally efficient. However, analytically modeling the covariance proves to be difficult due to the following reasons. The covariance of 2-point functions is related to 4-point functions that consists of the disconnected and the connected contributions. The disconnected covariance, including the Gaussian and Poisson errors, is a full-rank matrix and dominant on large scales [84]. It is sensitive to the window of the survey, which is non-trivial to model accurately. On the other hand, the connected covariance is a low-rank matrix, becomes important on small scales, and is hardly affected by the window function. It receives many contributions, including nonlinear mode-coupling [182, 233, 65, 111, 33, 185, 27, 155], the super-sample covariance [102, 125, 259, 260, 162, 163, 14, 164, 26], shot noises [182, 62, 253], and possible baryonic effects. However, even as the subject of extensive studies, the connected part remains intractable for accurate analytic prediction. Due to these challenges, semi-analytic approaches have been developed, by combining parametric covariance models with fewer number of mocks or the data [210, 190, 202, 100, 189]. These methods are able to relieve the computational burden by a factor of $\mathcal{O}(10)$, but their accuracy is subject to the quality of the parametric models. They also inherit other drawbacks from the mock and internal methods, such as fixation on fiducial cosmology, lack of correct physics, and missing large-scale modes.

In this chapter, we propose a hybrid method to tackle the covariance challenge. We develop a fully analytic method to compute the full-rank disconnected covariance with a proper treatment of the window effect. We demonstrate its accuracy and efficiency using galaxy mock simulations. Given the difficulty in modeling the connected contributions, and the fact that it is only important on small scales and not sensitive to the survey window, one should be able to calibrate it internally from the data, e.g. by low-rank approximation. By treating the disconnected and connected components separately, our hybrid approach enjoys the benefits of both the analytical and internal methods, and at the same time avoid their disadvantages. In this chapter we focus on the disconnected covariance, and leave the internal estimation of the connected part as a future project.

This chapter is structured as follows. In Sec. 4.2 we develop the methodology to analytically model the disconnected covariance matrix of the power spectrum or the correlation function of various large-scale structure probes, taking into account the survey window function. Our analytic method involves oscillatory integrals with two Bessel functions, which we solve in Sec. 4.3 with a novel quadrature algorithm. Using the mock simulations described in Sec. 4.4, we demonstrate the accuracy and efficiency of the analytic method in Sec. 4.5.

4.2 Formalism and Methodology

In this section we develop the formalism of the analytic disconnected covariance including the window effect. We start with the power spectrum covariance, and take the case of the power spectrum multipoles as the main example. Our derivation assumes flat-sky approximation, and is applicable to wide surveys by swapping in the curved-sky window functions. We then generalize our method to the case of the correlation function, and various other large-scale structure probes.

3D Power Spectrum Covariance

We begin by writing the number density of a discrete tracer field as a sum of Dirac delta functions

$$n_{\rm data}(\boldsymbol{x}) = \sum_i \delta^{\rm D}(\boldsymbol{x} - \boldsymbol{x}_i), \tag{4.1}$$

whose underlying number density is a continuous field $\bar{n}(\boldsymbol{x}) \equiv \langle n_{\rm data}(\boldsymbol{x})\rangle$, determined by the tracer evolution and the survey selection. $\langle\,\rangle$ takes the ensemble average or the expected value of a quantity. The survey selection is most conveniently captured by a synthetic random catalog, which in the simplest case is a Poisson process with mean $\langle n_{\rm rand}(\boldsymbol{x})\rangle = \bar{n}(\boldsymbol{x})/\alpha$, where α is a constant usually $\ll 1$ to reduce its shot noise. We can write down the overdensity fields of those data and random catalogs[2]

$$\begin{aligned}\delta_{\rm data} &= -1 + \frac{1}{\bar{n}(\boldsymbol{x})}\sum_{i\in\,\rm data}\delta^{\rm D}(\boldsymbol{x}-\boldsymbol{x}_i),\\ \delta_{\rm rand} &= -1 + \frac{\alpha}{\bar{n}(\boldsymbol{x})}\sum_{j\in\,\rm rand}\delta^{\rm D}(\boldsymbol{x}-\boldsymbol{x}_j).\end{aligned} \tag{4.2}$$

Following [84] (hereafter FKP), the galaxy overdensity is estimated as the difference between the data and random catalogs weighted by a weight function $w(\boldsymbol{x})$ (that maximizes the performance of an estimator like the FKP weight, and/or minimizes some systematic effects)

$$\begin{aligned}\delta_W(\boldsymbol{x}) &\equiv w(\boldsymbol{x})\big[n_{\rm data}(\boldsymbol{x}) - \alpha n_{\rm rand}(\boldsymbol{x})\big]\\ &\equiv W(\boldsymbol{x})\big[\delta_{\rm data}(\boldsymbol{x}) - \delta_{\rm rand}(\boldsymbol{x})\big] \equiv W(\boldsymbol{x})\delta(\boldsymbol{x}).\end{aligned} \tag{4.3}$$

On the second line we have rewritten the difference in number densities with the difference in overdensities, and defined a combined overdensity field $\delta \equiv \delta_{\rm data} - \delta_{\rm rand}$ that includes stochasticity from both data and random catalogs. We have also combined the survey selection and the weight function in the definition of the window function

$$W(\boldsymbol{x}) \equiv \bar{n}(\boldsymbol{x})w(\boldsymbol{x}). \tag{4.4}$$

[2]These overdensity fields of discrete catalogs are only formal, and they help to simplify the calculation of shot noise terms, e.g. $\langle\delta_{\rm data}\delta_{\rm rand}\rangle$ vanishes in (4.8).

Throughout this chapter W denotes the windows on the fields, and is to be distinguished from the other window factors introduced later for the 2-point functions or their covariances.

Under the flat-sky approximation, the power spectrum can be simply estimated by normalizing the squared Fourier-space overdensity before subtracting the shot noise

$$\hat{P}(\boldsymbol{k}) = \frac{|\delta_W(\boldsymbol{k})|^2}{\mathcal{W}_0} - P_{\rm shot}, \tag{4.5}$$

where the shot noise power spectrum is

$$P_{\rm shot} = \frac{\mathcal{S}_0}{\mathcal{W}_0}, \tag{4.6}$$

and the constant factors are given by[3]

$$\begin{aligned} \mathcal{W}_0 &= \int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})^2 w(\boldsymbol{x})^2 = \int_{\boldsymbol{x}} W(\boldsymbol{x})^2 = \int_{\boldsymbol{k}} |W(\boldsymbol{k})|^2, \\ \mathcal{S}_0 &= (1+\alpha) \int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x}) w(\boldsymbol{x})^2. \end{aligned} \tag{4.7}$$

Assuming flat sky and no redshift evolution, $\delta_{\rm data}(\boldsymbol{x})$ and $\delta_{\rm rand}(\boldsymbol{x})$ are statistically homogeneous and satisfy

$$\begin{aligned} \left\langle \delta_{\rm data}(\boldsymbol{k}) \delta_{\rm data}(-\boldsymbol{k}') \right\rangle &= (2\pi)^3 \, \delta^{\rm D}(\boldsymbol{k} - \boldsymbol{k}') P(\boldsymbol{k}) + \int_{\boldsymbol{x}} \frac{1}{\bar{n}(\boldsymbol{x})} e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}, \\ \left\langle \delta_{\rm rand}(\boldsymbol{k}) \delta_{\rm rand}(-\boldsymbol{k}') \right\rangle &= \int_{\boldsymbol{x}} \frac{\alpha}{\bar{n}(\boldsymbol{x})} e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}, \\ \left\langle \delta_{\rm data}(\boldsymbol{k}) \delta_{\rm rand}(-\boldsymbol{k}') \right\rangle &= 0. \end{aligned} \tag{4.8}$$

Here the Dirac delta is a result of the translation invariance, $P(\boldsymbol{k})$ is the 3D power spectrum of the tracer, and the shot noise terms involving $1/\bar{n}$ arise from the discrete nature of the data and random catalogs. As independent samples from $\bar{n}(\boldsymbol{x})$, the data and the random overdensity fields are not correlated, indicated by the third line in the above equations.

Plugging (4.3) and (4.8) into the ensemble average of the estimator in (4.5) we get

$$\left\langle \hat{P}(\boldsymbol{k}) \right\rangle = \frac{1}{\mathcal{W}_0} \int_{\boldsymbol{q}} P(\boldsymbol{k}-\boldsymbol{q}) |W(\boldsymbol{q})|^2 \simeq \frac{P(\boldsymbol{k})}{\mathcal{W}_0} \int_{\boldsymbol{q}} |W(\boldsymbol{q})|^2 = P(\boldsymbol{k}). \tag{4.9}$$

The first equality shows that the expectation of $\hat{P}$ is a convolution of the true power spectrum P with a window. To distinguish $\langle \hat{P} \rangle$ and P we refer to them as the convolved and the

[3] In this chapter we use the following shorthand notations for configuration-space and Fourier-space integrals

$$\int_{\boldsymbol{x}} \to \int \mathrm{d}^3\boldsymbol{x}, \qquad \int_{\boldsymbol{k}} \to \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}.$$

unconvolved power spectra, respectively. In the second equality we have assumed that the scales of interest are much smaller than the window size, i.e. $k \gg q$ for $\boldsymbol{q}$ where $W(\boldsymbol{q})$ is significant, so that a smooth power spectrum $P(\boldsymbol{k}-\boldsymbol{q}) \simeq P(\boldsymbol{k})$ can be taken out of the integral. Therefore the last equality proves that (4.5) is an unbiased estimation of the true power spectrum for modes much smaller than the survey scale. In other words, the unconvolved and the convolved power spectra differ on large scales due to the window.

The covariance function of the power spectrum estimator is define as

$$\mathrm{Cov}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] \equiv \big\langle \hat{P}(\boldsymbol{k})\hat{P}(\boldsymbol{k}')\big\rangle - \big\langle \hat{P}(\boldsymbol{k})\big\rangle\big\langle \hat{P}(\boldsymbol{k}')\big\rangle \tag{4.10}$$

Substituting (4.5) into the above equation, we split the covariance into the disconnected and connected pieces in the multivariate cumulant expansion

$$\begin{aligned}
\mathrm{Cov}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] &= \mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] + \mathrm{Cov}^{\mathrm{conn}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big], \\
\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] &= \frac{1}{\mathcal{W}_0^2}\big|\big\langle \delta_W(\boldsymbol{k})\delta_W(-\boldsymbol{k}')\big\rangle\big|^2 + (\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}'), \\
\mathrm{Cov}^{\mathrm{conn}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] &= \frac{1}{\mathcal{W}_0^2}\big\langle \delta_W(\boldsymbol{k})\delta_W(-\boldsymbol{k})\delta_W(\boldsymbol{k}')\delta_W(-\boldsymbol{k}')\big\rangle_{\mathrm{c}}.
\end{aligned} \tag{4.11}$$

The disconnected part $\mathrm{Cov}^{\mathrm{disc}}$ captures the part of the 4-point correlation arising from products of 2-point correlations, and the connected part $\mathrm{Cov}^{\mathrm{conn}}$ is from the excess correlation beyond $\mathrm{Cov}^{\mathrm{disc}}$. The subscript "c" on the ensemble average denotes the connected part or the cumulant of the 4-point function.

In the conventional terminology, $\mathrm{Cov}^{\mathrm{disc}}$ is referred to as the Gaussian part and $\mathrm{Cov}^{\mathrm{conn}}$ is named the non-Gaussian part. This is true for a continuous random field, in which case the terms "disconnected" and "Gaussian" can be used interchangeably. However, the conventional names are not accurate and can be confusing for point processes like a galaxy catalog. Since Gaussian and Poisson contributions enter both $\mathrm{Cov}^{\mathrm{disc}}$ and $\mathrm{Cov}^{\mathrm{conn}}$, $\mathrm{Cov}^{\mathrm{disc}}$ is not purely Gaussian and $\mathrm{Cov}^{\mathrm{conn}}$ is not completely free of Gaussian contribution. Therefore in this chapter we rename this covariance decomposition for clarification.

Let's first look at the disconnected piece. From (4.3) and (4.8)

$$\big\langle \delta_W(\boldsymbol{k})\delta_W(-\boldsymbol{k}')\big\rangle = \int_{\boldsymbol{k}''} P(\boldsymbol{k}'')W(\boldsymbol{k}-\boldsymbol{k}'')W(\boldsymbol{k}''-\boldsymbol{k}') + (1+\alpha)\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})w(\boldsymbol{x})^2 e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}, \tag{4.12}$$

which includes both Gaussian and Poisson contributions. For modes much smaller than the survey scale, the Gaussian term is only important when $\boldsymbol{k}''$ is close to both $\boldsymbol{k}$ and $\boldsymbol{k}'$, thus $P(\boldsymbol{k}'')$ approximates $P(\boldsymbol{k})$ and $P(\boldsymbol{k}')$. So we can approximate the integral by taking the power spectrum out of the convolution while preserving the $\boldsymbol{k} \leftrightarrow \boldsymbol{k}'$ exchange symmetry, and then plug it into (4.11)

$$\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] \approx \left|\frac{P(\boldsymbol{k}) + P(\boldsymbol{k}')}{2}\frac{\mathcal{W}(\boldsymbol{k}-\boldsymbol{k}')}{\mathcal{W}_0} + P_{\mathrm{shot}}\frac{\mathcal{S}(\boldsymbol{k}-\boldsymbol{k}')}{\mathcal{S}_0}\right|^2 + (\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}'), \tag{4.13}$$

where we have introduced the following window factors

$$\mathcal{W}(\boldsymbol{q}) = \int_{\boldsymbol{x}} \mathcal{W}(\boldsymbol{x}) e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \equiv \int_{\boldsymbol{x}} W(\boldsymbol{x})^2 e^{-i\boldsymbol{q}\cdot\boldsymbol{x}},$$
$$\mathcal{S}(\boldsymbol{q}) = \int_{\boldsymbol{x}} \mathcal{S}(\boldsymbol{x}) e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \equiv (1+\alpha) \int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x}) w(\boldsymbol{x})^2 e^{-i\boldsymbol{q}\cdot\boldsymbol{x}}, \tag{4.14}$$

that modulate Gaussian and Poisson pieces, respectively. $\mathcal{W}$ and $\mathcal{S}$ are to be distinguished from W, the window on the field. Notice that the constants $\mathcal{W}_0$ and $\mathcal{S}_0$ introduced earlier in (4.7) are special cases of $\mathcal{W}(\boldsymbol{q})$ and $\mathcal{S}(\boldsymbol{q})$ at $\boldsymbol{q} = 0$.

The expansion of (4.13) contains quadratic combinations $\mathcal{W}^2$, $\mathcal{W}\mathcal{S}$, and $\mathcal{S}^2$, therefore we further define the window factor $\mathcal{Q}$'s as the auto and cross 2-point correlation of $\mathcal{W}$ and $\mathcal{S}$. In Fourier space

$$\mathcal{Q}_{\mathcal{W}}(\boldsymbol{q}) \equiv \mathcal{W}(\boldsymbol{q})\mathcal{W}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{W}}(\boldsymbol{s}) e^{-i\boldsymbol{q}\cdot\boldsymbol{s}},$$
$$\mathcal{Q}_{\mathcal{S}}(\boldsymbol{q}) \equiv \mathcal{S}(\boldsymbol{q})\mathcal{S}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{S}}(\boldsymbol{s}) e^{-i\boldsymbol{q}\cdot\boldsymbol{s}},$$
$$\mathcal{Q}_{\times}(\boldsymbol{q}) \equiv \mathcal{W}(\boldsymbol{q})\mathcal{S}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\times}(\boldsymbol{s}) e^{-i\boldsymbol{q}\cdot\boldsymbol{s}}, \tag{4.15}$$

with the configuration-space $\mathcal{Q}(\boldsymbol{s})$'s being the correlation functions of $\mathcal{W}(\boldsymbol{x})$ and $\mathcal{S}(\boldsymbol{x})$, e.g.

$$\mathcal{Q}_{\times}(\boldsymbol{s}) = \int_{\boldsymbol{x}} \mathcal{W}(\boldsymbol{x}+\boldsymbol{s})\mathcal{S}(\boldsymbol{x}). \tag{4.16}$$

So they can be measured as the correlation functions or power spectra of the properly weighted random catalogs, which we describe in more details in Sec. 4.2.

The final expression for the disconnected covariance given the power spectrum P and the window function $\mathcal{Q}$'s is

$$\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] \approx \frac{1}{\mathcal{W}_0^2}\Big\{P(\boldsymbol{k})P(\boldsymbol{k}')\mathcal{Q}_{\mathcal{W}}(\boldsymbol{k}-\boldsymbol{k}') + \big[P(\boldsymbol{k}) + P(\boldsymbol{k}')\big]\Re\big[\mathcal{Q}_{\times}(\boldsymbol{k}-\boldsymbol{k}')\big] + \mathcal{Q}_{\mathcal{S}}(\boldsymbol{k}-\boldsymbol{k}')\Big\} + (\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}'). \tag{4.17}$$

Notice that under the same $P(\boldsymbol{k}) \approx P(\boldsymbol{k}')$ approximation we have further simplified the expression by combining both $P(\boldsymbol{k})P(\boldsymbol{k})$ and $P(\boldsymbol{k}')P(\boldsymbol{k}')$ terms into $P(\boldsymbol{k})P(\boldsymbol{k}')$ while preserving the exchange symmetry. In general, the $\mathcal{Q}(\boldsymbol{k}-\boldsymbol{k}')$ windows have non-vanishing width determined by the survey size, and its shape characterizes the correlation between neighboring modes at $\boldsymbol{k}$ and $\boldsymbol{k}'$. Also different $\mathcal{Q}$ windows generally have different shapes, and neglecting this difference would lead to inaccurate $\mathrm{Cov}^{\mathrm{disc}}$, which we show later in Fig. 4.7.

Now turning to the connected piece $\mathrm{Cov}^{\mathrm{conn}}$, it is composed of a mixture of the non-Gaussian, Poisson, and Gaussian contributions. The non-Gaussian part arises from the

gravitational mode-coupling, including the trispectrum piece [182, 233] and the super-sample covariance (SSC) [102, 125, 260]. The remaining parts of $\mathrm{Cov}^{\mathrm{conn}}$ are various shot noise terms due to the discrete nature of the tracer field [182], and involves Poisson, non-Gaussian (bispectrum), and Gaussian (power spectrum) components. Overall $\mathrm{Cov}^{\mathrm{conn}}$ is more difficult and complicated to model analytically than $\mathrm{Cov}^{\mathrm{disc}}$. However, it is a smooth function of $\boldsymbol{k}$ and $\boldsymbol{k}'$ and can be well approximated with a low-rank eigen-decomposition [111, 185], which allows it to be measured from the data with the internal covariance estimators. Leaving the internal estimation of $\mathrm{Cov}^{\mathrm{conn}}$ as a future project, in this chapter we can obtain $\mathrm{Cov}^{\mathrm{conn}}$ from the mock simulations by subtracting C^{disc} from the mock sample covariance. We find this empirical $\mathrm{Cov}^{\mathrm{conn}}$ is indeed smooth and a low-rank component with a principal component analysis.

Diagonal Limit

We expect $\mathrm{Cov}^{\mathrm{disc}}$ in (6.12) to reduce to the familiar diagonal form when certain condition is met. In the limit where $|\boldsymbol{k}-\boldsymbol{k}'|$ is much greater than the window scale, e.g. when they are from different bins with very wide bin width, $\mathcal{Q}(\boldsymbol{q})$ approaches $(2\pi)^3\,\delta^{\mathrm{D}}(\boldsymbol{q})\mathcal{Q}(\boldsymbol{s}=0)$, and $\mathrm{Cov}^{\mathrm{disc}}$ reduces to a diagonal covariance:

$$\mathrm{Cov}^{\mathrm{diag}}\big[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\big] = (2\pi)^3\,\delta^{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}')\bigg\{\frac{P(\boldsymbol{k})^2}{V_{\mathcal{W}}}+\frac{2P(\boldsymbol{k})P_{\mathrm{shot}}}{V_{\times}}+\frac{P_{\mathrm{shot}}^2}{V_{\mathcal{S}}}\bigg\}+(\boldsymbol{k}'\leftrightarrow-\boldsymbol{k}'),\tag{4.18}$$

where V's are effective volumes defined as follows

$$\begin{aligned}
V_{\mathcal{W}} &\equiv \frac{\mathcal{W}_0^2}{\mathcal{Q}_{\mathcal{W}}(\boldsymbol{s}=0)} = \frac{\big[\int_{\boldsymbol{x}}\bar{n}(\boldsymbol{x})^2w(\boldsymbol{x})^2\big]^2}{\int_{\boldsymbol{x}}\bar{n}(\boldsymbol{x})^4w(\boldsymbol{x})^4},\\
V_{\mathcal{S}} &\equiv \frac{\mathcal{S}_0^2}{\mathcal{Q}_{\mathcal{S}}(\boldsymbol{s}=0)} = \frac{\big[\int_{\boldsymbol{x}}\bar{n}(\boldsymbol{x})w(\boldsymbol{x})^2\big]^2}{\int_{\boldsymbol{x}}\bar{n}(\boldsymbol{x})^2w(\boldsymbol{x})^4},\\
V_{\times} &\equiv \frac{\mathcal{W}_0\mathcal{S}_0}{\mathcal{Q}_{\times}(\boldsymbol{s}=0)} = \frac{\int_{\boldsymbol{x}}\bar{n}(\boldsymbol{x})^2w(\boldsymbol{x})^2\int_{\boldsymbol{x}'}\bar{n}(\boldsymbol{x}')w(\boldsymbol{x}')^2}{\int_{\boldsymbol{x}}\bar{n}(\boldsymbol{x})^3w(\boldsymbol{x})^4}.
\end{aligned}\tag{4.19}$$

We call (4.18) the diagonal limit, and take it as an ansatz for $\mathrm{Cov}^{\mathrm{disc}}$ for any $\boldsymbol{k}$ and $\boldsymbol{k}'$ even though the derivation only holds far enough from the diagonal. This diagonal covariance accounts for the size of the window which is captured by $\mathcal{Q}(\boldsymbol{s}=0)$, but ignores the shape of $\mathcal{Q}(\boldsymbol{s})$, and as a result biases the signal-to-noise ratio shown later in Sec. 4.5.

In the diagonal limit, the $\mathrm{Cov}^{\mathrm{disc}}$ matrix of the band-power, binned in spherical k shells following the later Sec. 4.2, is

$$\mathrm{Cov}^{\mathrm{diag}}\big[\hat{P}(k_i),\hat{P}(k_j)\big] = 2\delta_{ij}^{\mathrm{K}}\int_{k_{i-\frac{1}{2}}}^{k_{i+\frac{1}{2}}}\frac{4\pi k^2\mathrm{d}k}{V_{k_i}}\bigg\{\frac{P(k)^2}{N_{\mathcal{W}}}+\frac{2P(k)P_{\mathrm{shot}}}{N_{\times}}+\frac{P_{\mathrm{shot}}^2}{N_{\mathcal{S}}}\bigg\},\tag{4.20}$$

where $\delta^{\rm K}$ is the Kronecker delta, and N's are the effective numbers of modes associated with the effective volumes:

$$N_{\mathcal{W}} = \frac{V_k V_{\mathcal{W}}}{(2\pi)^3}, \quad N_{\mathcal{S}} = \frac{V_k V_{\mathcal{S}}}{(2\pi)^3}, \quad N_{\times} = \frac{V_k V_{\times}}{(2\pi)^3}. \tag{4.21}$$

Often only $N_{\mathcal{W}}$ is used to compute the diagonal covariance matrix, ignoring the differences among N's, that however could cause larger biases in the signal-to-noise ratio (see Sec. 4.5). Eq. (4.18) and the constant effective volumes generalize the scale-dependent effective volume commonly used in survey forecast [266] for any weights $w(\boldsymbol{x})$.

Multipole Covariance

In redshift space, because of the azimuthal symmetry about the line of sight (LOS), the power spectrum does not have azimuthal dependence, i.e. $P(\boldsymbol{k}) = P(k, \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})$, where $\hat{\boldsymbol{n}}$ is the LOS direction. It is natural to decompose it into multipoles

$$P(\boldsymbol{k}) = \sum_{\ell} P_\ell(k)\, \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}}), \tag{4.22}$$

in which $\mathcal{L}_\ell$ is the Legendre polynomial of degree ℓ. The multipole moments can be obtained by[4]

$$P_\ell(k) = (2\ell+1)\int_{\hat{\boldsymbol{k}}} P(\boldsymbol{k})\, \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}}). \tag{4.23}$$

In real space, all high-order multipoles usually vanish, leaving only the monopole.

(4.22) and (4.23) also apply to their estimators $\hat{P}(\boldsymbol{k})$ and $\hat{P}_\ell(k)$, so the disconnected covariance of multipoles follows straightforwardly from (6.12)

$$\begin{aligned}\mathrm{Cov}^{\rm disc}\big[\hat{P}_\ell(k), \hat{P}_{\ell'}(k')\big] \approx \frac{2(2\ell+1)(2\ell'+1)}{\mathcal{W}_0^2}\int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})\,\mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}'\cdot\hat{\boldsymbol{n}})\Big\{ \\ P(\boldsymbol{k})P(\boldsymbol{k}')\mathcal{Q}_{\mathcal{W}}(\boldsymbol{k}-\boldsymbol{k}') + \big[P(\boldsymbol{k})+P(\boldsymbol{k}')\big]\Re\big[\mathcal{Q}_\times(\boldsymbol{k}-\boldsymbol{k}')\big] + \mathcal{Q}_{\mathcal{S}}(\boldsymbol{k}-\boldsymbol{k}')\Big\}.\end{aligned} \tag{4.24}$$

Since $P(\boldsymbol{k}) = P(-\boldsymbol{k})$ for auto power spectrum, here we only consider the even multipoles, for which the $\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}'$ term in (6.12) simply doubles the first one giving the factor of 2 in the above equation.

[4] In this chapter we use the following shorthand notation for averaging (instead of integrating) over 4π solid angle of any vector $\boldsymbol{v}$: $\int_{\hat{\boldsymbol{v}}} \to \int \frac{\mathrm{d}\Omega_{\boldsymbol{v}}}{4\pi}$.

We compute the three terms in the curly brackets of (4.24) separately. First we consider the P^2 term, expand the power spectra in multipoles, plug in $\mathcal{Q}_\mathcal{W}$ from (6.11) and derive

$$\begin{aligned}
&\int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})\,\mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}'\cdot\hat{\boldsymbol{n}})P(\boldsymbol{k})P(\boldsymbol{k}')\int_{\boldsymbol{s}}\mathcal{Q}_\mathcal{W}(\boldsymbol{s})e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{s}}\\
&=\sum_{\ell_1\ell_2\ell_3\ell_4} P_{\ell_1}(k)P_{\ell_3}(k')(2\ell_2+1)\begin{pmatrix}\ell & \ell_1 & \ell_2\\ 0 & 0 & 0\end{pmatrix}^2(2\ell_4+1)\begin{pmatrix}\ell' & \ell_3 & \ell_4\\ 0 & 0 & 0\end{pmatrix}^2\\
&\qquad\times(-i)^{\ell_2-\ell_4}\int 4\pi s^2\mathrm{d}s\sum_{\ell''}\begin{pmatrix}\ell_2 & \ell_4 & \ell''\\ 0 & 0 & 0\end{pmatrix}^2\mathcal{Q}_{\mathcal{W}\ell''}(s)j_{\ell_2}(ks)j_{\ell_4}(k's). \quad (4.25)
\end{aligned}$$

Similarly we obtain the cross term between power spectrum and shot noise ($P\times P_\mathrm{shot}$)

$$\begin{aligned}
&\int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})\,\mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}'\cdot\hat{\boldsymbol{n}})\big[P(\boldsymbol{k})+P(\boldsymbol{k}')\big]\Re\Big[\int_{\boldsymbol{s}}\mathcal{Q}_\times(\boldsymbol{s})e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{s}}\Big]\\
&=\sum_{\ell_1\ell_2} P_{\ell_1}(k)(2\ell_2+1)\begin{pmatrix}\ell & \ell_1 & \ell_2\\ 0 & 0 & 0\end{pmatrix}^2(-i)^{\ell_2-\ell'}\int 4\pi s^2\mathrm{d}s\sum_{\ell''}\begin{pmatrix}\ell_2 & \ell' & \ell''\\ 0 & 0 & 0\end{pmatrix}^2\mathcal{Q}_{\times\ell''}(s)j_{\ell_2}(ks)j_{\ell'}(k's)\\
&\quad+(\ell,k\leftrightarrow\ell',k'), \quad (4.26)
\end{aligned}$$

and the third (P^2_shot) term

$$\begin{aligned}
&\int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})\,\mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}'\cdot\hat{\boldsymbol{n}})\int_{\boldsymbol{s}}\mathcal{Q}_\mathcal{S}(\boldsymbol{s})e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{s}}\\
&\qquad=(-i)^{\ell-\ell'}\int 4\pi s^2\mathrm{d}s\sum_{\ell''}\begin{pmatrix}\ell & \ell' & \ell''\\ 0 & 0 & 0\end{pmatrix}^2\mathcal{Q}_{\mathcal{S}\ell''}(s)j_\ell(ks)j_{\ell'}(k's). \quad (4.27)
\end{aligned}$$

Note that in the above derivations the disconnected covariance only depends on the multipole moments of the window $\mathcal{Q}$ due to the redshift-space symmetry

$$\mathcal{Q}_\ell(s)=(2\ell+1)\int_{\hat{\boldsymbol{s}}}\mathcal{Q}(\boldsymbol{s})\,\mathcal{L}_\ell(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}). \quad (4.28)$$

As shown in (6.11) and (4.16), Q factors are really the 2-point functions of the $\mathcal{W}$ and $\mathcal{S}$ windows, so their multipoles can be readily measured from a random catalog of a survey. Given $\mathcal{Q}_\ell(s)$ from the randoms and $P_\ell(k)$ from the data, we can numerically evaluate (4.25), (4.26) and (4.27) before summing them up in (4.24) to obtain the disconnected covariance for power spectrum multipoles. The only remaining difficulty lies in the numerical integrals involving two spherical Bessel functions, to which we provide a novel solution in Sec. 4.3.

Window Functions

In deriving the disconnected covariance we have assumed the flat-sky approximation. However, many current and future surveys have wide sky coverages, making their window functions fundamentally curved-sky entities. Furthermore, the redshift-space clustering between two points depends on the LOS direction that varies with the positions of the pair of points. To account for this, the Yamamoto estimator proposed by Ref. [279] measures the multipoles of galaxy clustering with respect to the midpoint direction of the two galaxies. An alternative estimator also suggested by Ref. [279] uses the position of one galaxy of the pair as the reference direction, and can be measured with a suite of efficient FFT-based algorithms [104, 258, 41, 234, 247].

Fortunately, it is easy to adapt our formalism to account for the curved-sky effects both in the window function and the power spectrum. For modes smaller than the window scale, our flat-sky equations remain a good approximation to capture their local power-window coupling. The LOS dependence can be simply incorporated in our window functions, e.g. in (4.28) by letting the fixed line-of-sight direction $\hat{\boldsymbol{n}}$ to depend on $\boldsymbol{x}$ and/or $\boldsymbol{x}+\boldsymbol{s}$. Likewise, we can use the model fit to the power spectrum multipoles measured with a LOS-dependent estimator, to compute $\mathrm{Cov}^{\mathrm{disc}}$.

We use `nbodykit`, which calls `corrfunc` [245], to measure the window functions via fast pair counting. With a random catalog that describes the survey geometry and weights, we count the weighted pair in bins of both separation s and the polar angle cosine μ, and then normalize the counts by the shell volume. For example

$$\mathcal{Q}_{\mathcal{S}}(s,\mu) \approx \frac{3\int_{\boldsymbol{x},\boldsymbol{s}} \mathcal{W}(\boldsymbol{x}+\boldsymbol{s})\mathcal{S}(\boldsymbol{s})}{2\pi(s_i^3-s_{i-1}^3)(\mu_j-\mu_{j-1})} \approx \frac{3\alpha^2\sum_{ab}\bar{n}(\boldsymbol{x}_a)w(\boldsymbol{x}_a)^2 w(\boldsymbol{x}_b)^2}{2\pi(s_i^3-s_{i-1}^3)(\mu_j-\mu_{j-1})}. \tag{4.29}$$

We take 441 logarithmic s bins ranging from $1\,\mathrm{Mpc}/h$ to $3.4\,\mathrm{Gpc}/h$, and 100 μ bins from 0 to 1. The `corrfunc` package measures the midpoint polar cosine

$$\mu = \frac{\hat{\boldsymbol{s}}\cdot(\boldsymbol{x}+\boldsymbol{s}/2)}{|\boldsymbol{x}+\boldsymbol{s}/2|}. \tag{4.30}$$

To relate the double integral to the pair summation, we have simply replaced $\int_{\boldsymbol{x}}\bar{n}$ with $\alpha\sum_a$. Given the $\mathcal{Q}$ windows in s and μ bins, we then transform them into multipoles to obtain an estimate of $\mathcal{Q}_\ell$ with (4.28). All the odd order multipoles vanish because of the choice of midpoint for measuring μ. With the same discretization trick, we can estimate the normalization constants in (4.7) by replacing the integral with a summation over a random catalog, e.g.

$$\mathcal{W}_0 \approx \alpha\sum_a \bar{n}(\boldsymbol{x}_a)w(\boldsymbol{x}_a)^2. \tag{4.31}$$

We have seen in (6.9) that the expectation of the estimated power spectrum is subject to a convolution with the window, and the unconvolved and convolved power spectra differ significantly on large scales. For accurate $\mathrm{Cov}^{\mathrm{disc}}$, rather than directly using the estimated $\hat{P}$,

we should use an unconvolved model fit. To achieve this we convolve the power spectrum model from [105] with the mask [277] before fitting it to the data (more detailed description in Sec. 4.5 and illustration in Fig. 4.2). We then use the model before convolution to compute $\mathrm{Cov}^{\mathrm{disc}}$.

Corollary Covariances

We generalize our formalism below for more applications.

Cross Correlation and Cross Covariance

In Sec. 4.2 and 4.2 we have considered the simplest case for 2-point function disconnected covariance: the covariance of auto correlations of a single tracer. In more general applications, one may need the cross covariance between different sets of auto and cross correlations between different tracers. For the most general case, we can derive the equations for cross covariance of cross 3D power spectra. The derivation parallels that of Sec. 4.2 and preserves the Hermitian symmetry. The equations can be easily generalized to all other cases including multipoles and those in the rest of this subsections.

Angular Power

We can generalize the formalism developed in Sec 4.2 to the angular power spectrum $C(\ell)$ which are generally used to measure the correlations in observables from CMB, weak gravitational lensing as well as the clustering of photometric samples of galaxies. Compared to the multipole case, the angular window effect is captured by some $\mathcal{Q}$ windows similar to (4.28) but defined on the sky, and the evaluation involves double Bessel integral but with the zeroth-order Bessel function J_0 in place of the spherical Bessel functions. The curved sky window can be incorporated simply by using the $\mathcal{Q}$ window measured from the random catalog as a function of angular separation on the sky similar to Sec. 4.2. Our method can be a fast alternative to the calculation on the sphere as is generally done for the CMB measurements [207].

Projected Power and Projected Cross Multipoles

When cross correlating a field with 3-dimensional information, such as spectroscopic galaxies, with a field with poor radial information, e.g. weak lensing shear or convergence, one can estimate their correlation projected at fixed transverse separation using the radial information of the 3D field. This is commonly done in galaxy-shear cross correlations measurements using spectroscopic galaxies for which covariance was derived by [242]. Because a projected power spectrum is indeed the transverse part of its 3D counterpart, and the transverse power can be approximated by a series of multipoles, we can adapt the multipole formalism to work for this cross covariance.

Correlation Functions

In configuration space, the Landy-Szalay estimator [156] is typically employed to measure the correlation function, and it can be written in our notation as

$$\hat{\xi}(\boldsymbol{s}) = \frac{1}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{x}} \delta_W(\boldsymbol{x})\delta_W(\boldsymbol{x}+\boldsymbol{s}), \tag{4.32}$$

where we have defined the normalization factor analogous to (4.16)

$$\mathcal{Q}_W(\boldsymbol{s}) \equiv \int_{\boldsymbol{x}} W(\boldsymbol{x}+\boldsymbol{s})W(\boldsymbol{x}). \tag{4.33}$$

(4.32) is free from the window effect, that can be shown easily in flat sky by taking its expectation

$$\begin{aligned}\left\langle\hat{\xi}(\boldsymbol{s})\right\rangle &= \frac{1}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{x}} W(\boldsymbol{x})W(\boldsymbol{x}+\boldsymbol{s})\left\langle\delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{s})\right\rangle \\ &\equiv \frac{\xi(\boldsymbol{s})}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{x}} W(\boldsymbol{x})W(\boldsymbol{x}+\boldsymbol{s}) = \xi(\boldsymbol{s}),\end{aligned} \tag{4.34}$$

where we have used the translation invariance of the correlation function $\xi(\boldsymbol{s}) \equiv \left\langle\delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{s})\right\rangle$.

The 3D correlation function and power spectrum are simply related by the Fourier transform, again assuming translation invariance. However, the relation between their estimators is more complex due to the difference in normalizations:

$$\begin{aligned}\xi(\boldsymbol{s}) &= \int_{\boldsymbol{k}} P(\boldsymbol{k})e^{i\boldsymbol{k}\cdot\boldsymbol{s}}, \\ \hat{\xi}(\boldsymbol{s}) &= \frac{\mathcal{W}_0}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{k}} \hat{P}(\boldsymbol{k})e^{i\boldsymbol{k}\cdot\boldsymbol{s}}.\end{aligned} \tag{4.35}$$

Here $\mathcal{W}_0$ is equal to the value of $\mathcal{Q}_W(\boldsymbol{s}=0)$ according to (4.14) and (4.33). So the covariance of the two 3D estimators are related by

$$\mathrm{Cov}\left[\hat{\xi}(\boldsymbol{s}),\hat{\xi}(\boldsymbol{s}')\right] = \frac{\mathcal{W}_0^2}{\mathcal{Q}_W(\boldsymbol{s})\mathcal{Q}_W(\boldsymbol{s}')} \int_{\boldsymbol{k},\boldsymbol{k}'} \mathrm{Cov}\left[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\right] e^{i\boldsymbol{k}\cdot\boldsymbol{s}} e^{i\boldsymbol{k}'\cdot\boldsymbol{s}'}. \tag{4.36}$$

We are more interested in the correlation function covariances cast in the multipole, angular, and projected form.

Wedges

In this chapter and especially Sec. 4.2, we focus on using the multipole moments to label the angular variation in 2-point functions. Another popular choice is the so-called wedges, which are averages of the 2-point functions within bins of polar angle cosine, μ_i, between

bin edges $(\mu_{i-\frac{1}{2}}, \mu_{i+\frac{1}{2}})$ for $i = 1, \ldots, N_{\rm wedge}$. The power spectrum wedges are related to the multipoles by

$$\hat{P}(k, \mu_i) = \int_{\mu_{i-\frac{1}{2}}}^{\mu_{i+\frac{1}{2}}} \frac{\mathrm{d}\mu}{\mu_{i+\frac{1}{2}} - \mu_{i-\frac{1}{2}}} \Big[\sum_{\ell=0}^{\ell_{\rm max}} \hat{P}_\ell(k)\, \mathcal{L}_\ell(\mu)\Big] = \sum_{\ell=0}^{\ell_{\rm max}} \hat{P}_\ell(k) \bar{\mathcal{L}}_\ell(\mu_i), \tag{4.37}$$

where we have truncated the multipoles at $\ell_{\rm max}$ and denoted the mean Legendre polynomial across a wedge by $\bar{\mathcal{L}}$

$$\bar{\mathcal{L}}_\ell(\mu_i) \equiv \int_{\mu_{i-\frac{1}{2}}}^{\mu_{i+\frac{1}{2}}} \frac{\mathcal{L}_\ell(\mu)\, \mathrm{d}\mu}{\mu_{i+\frac{1}{2}} - \mu_{i-\frac{1}{2}}}. \tag{4.38}$$

Therefore

$$\mathrm{Cov}^{\rm disc}\big[\hat{P}(k, \mu_i), \hat{P}(k', \mu_j)\big] = \sum_{\ell=0}^{\ell_{\rm max}} \sum_{\ell'=0}^{\ell_{\rm max}} \mathrm{Cov}^{\rm disc}\big[\hat{P}_\ell(k), \hat{P}_{\ell'}(k')\big] \bar{\mathcal{L}}_\ell(\mu_i) \bar{\mathcal{L}}_{\ell'}(\mu_j). \tag{4.39}$$

The same relation applies to the correlation functions[5] by replacing $P \to \xi$, $k \to s$.

Binning

All previous results in this section assume the power spectrum covariance as a function of continuous k and k'. When numerically evaluating the equations we then discretize k and s on logarithmic grids (more details in Sec. 4.3). In practice, the estimated power spectra are averaged in bins k_i separated by bin edges $(k_{i-\frac{1}{2}}, k_{i+\frac{1}{2}})$ for $i = 1, \ldots, N_{\rm bin}$. In accordance, we integrate the analytic prediction within spherical k-shells

$$\mathrm{Cov}\big[\hat{P}(k_i), \hat{P}(k_j)\big] = \int_{k_{i-\frac{1}{2}}}^{k_{i+\frac{1}{2}}} \frac{4\pi k^2 \mathrm{d}k}{V_{k_i}} \int_{k_{j-\frac{1}{2}}}^{k_{j+\frac{1}{2}}} \frac{4\pi k'^2 \mathrm{d}k'}{V_{k_j}} \mathrm{Cov}\big[\hat{P}(k), \hat{P}(k')\big], \tag{4.40}$$

where $V_{k_i} = 4\pi(k^3_{i+\frac{1}{2}} - k^3_{i-\frac{1}{2}})/3$ is the volume of the ith k-shell. The numerical integration is achieved by interpolating the covariance function along each k at a time with B-splines before integrating the piecewise polynomials.[6]

In the case of correlation functions, because their denominators are scale dependent, we need to average separately the numerator and denominator in s bins before dividing them. The binning integrals are similar to the above equation.

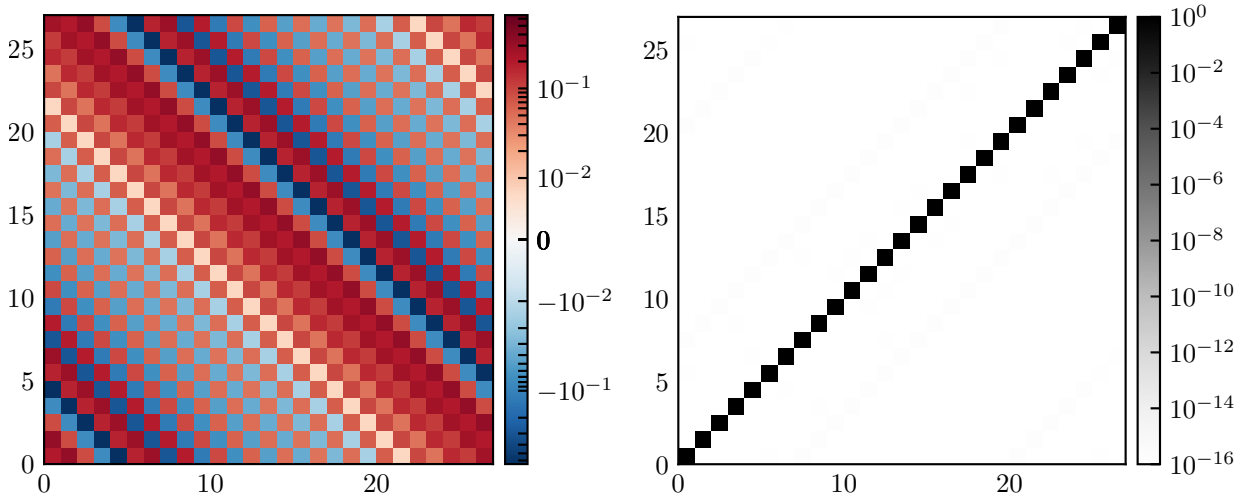


Figure 4.1: An example 27×27 Hankel matrix H_0, for the J_0 Hankel transform. We visualize it in the left panel, and demonstrate in the right panel that the squared Hankel matrix is an identity matrix. As a circulant matrix, a Hankel matrix carries out the convolution operation in the FFTLog algorithm (Eq. 4.43 and 4.45). It is also the building block of the double Bessel quadrature method (Eq. 4.41 and 4.50).

4.3 Double Bessel Quadrature

Eq. (4.25), (4.26), and (4.27) all have integrals involving two spherical Bessel functions. More generally, we are interested in evaluating the following integral with Bessel functions

$$G(y, y') = \int_0^\infty x\mathrm{d}x\ F(x) J_\nu(xy) J_{\nu'}(xy'), \tag{4.41}$$

which proves useful also for the applications in Sec.4.2. This applies to (4.25), (4.26), and (4.27) given

$$j_n(t) = \sqrt{\frac{\pi}{2t}} J_{n+\frac{1}{2}}(t). \tag{4.42}$$

Usual quadrature methods struggle to converge for (4.41) because the Bessel kernels are highly oscillatory and damp slowly. Recent solutions proposed by Ref. [21, 95] generalized the FFTLog algorithm [264, 101] to perform the integral transform from $F(x)$ to $G(y, y')$ for each fixed ratio y'/y.

[5]Note that the wedges are different from the μ bins used for estimating $\hat{\xi}_\ell(s)$. While one can directly estimate $\hat{P}_\ell(k)$, due to its normalization $\hat{\xi}_\ell(s)$ is usually converted from $\hat{\xi}(s, \mu_i)$ which is estimated in very fine μ bins. Here the width of the wedges is assumed to be much wider than the width used for such conversion.

[6]We package and release this simple and general-purpose utility at `https://github.com/eelregit/avgem`.

Here we present a fast and simple algorithm that potentially suits the covariance calculation better than that in Ref. [21, 95]. First let's look at a simpler integral with only one Bessel function

$$B(x) = \int_0^\infty y\mathrm{d}y\ A(y)J_\nu(xy), \tag{4.43}$$

known as the Hankel transform. We can perform integral transforms like this one using the `mcfit` [165] package, which implements and generalizes the FFTLog algorithm. This method exploits the convolution theorem in terms of $\ln x$ and $\ln y$. It approximates $A(e^{\ln y})$ with truncated Fourier series over one period of the periodic approximant, and Fourier-transforms the kernel $J_\nu(e^{\ln x+\ln y})$ analytically. Because of the exact treatment of the kernel function, this algorithm is ideal for oscillatory kernels like the Bessel functions.

Specifically, we compute an equivalent form of (4.43)

$$xB(x) = \int_0^\infty \frac{\mathrm{d}y}{y}\ \big[yA(y)\big]\big[xyJ_\nu(xy)\big], \tag{4.44}$$

with the input, output, and kernel functions rescaled by corresponding linear powers: $A(y) \to yA(y)$, $B(x) \to xB(x)$, and $J_\nu(t) \to tJ_\nu(t)$. `mcfit` evaluates (4.44) by discretizing x and y to grids of equal logarithmic intervals $\Delta \equiv \delta \ln x = \delta \ln y$, on which the linear transform essentially takes the following matrix form

$$B_i = x_i^{-1} \sum_{j=1}^{N} H_{\nu,ij} y_j A_j, \tag{4.45}$$

where the H_ν is an $N \times N$ real matrix carrying out the convolution, and N is the number of grid points. H_ν is a Hankel circulant matrix, i.e. $H_{\nu,ij} = h_{i+j}$ with h being periodic ($h_{i+N} = h_i$). The explicit form of h can be found in Eq.(B26) of Ref. [101]. As an example we show in Fig. 4.1 a 27×27 H_0 matrix with $\Delta = 0.34$. The above rescaling in (4.44) is important because it (together with the FFTLog low-ringing condition when N is even) makes H_ν a unitary matrix and involutory (being its own inverse), leading to the most numerically stable results as demonstrated in the right panel of Fig. 4.1 by the nearly perfect agreement between H_0^2 and the identity matrix to machine precision.

Notice that formally (4.45) can be obtained by applying the following replacement rule on (4.44) or (4.43):

$$xyJ_\nu(xy) \longrightarrow \frac{H_{\nu,ij}}{\Delta}, \qquad \int \frac{\mathrm{d}y}{y} \longrightarrow \Delta \sum_j . \tag{4.46}$$

While we emphasize that this rule is not rigorously true, i.e. $H_{\nu,ij} \not\propto J_\nu(x_i y_j)$, coincidentally, our final recipe for the double Bessel integral (4.41) also follows from this rule.

Now we are ready to derive our double Bessel quadrature algorithm. Consider the following integral that further integrates (4.41) with some arbitrary auxiliary functions $A(y)$ and $A'(y')$ over y and y'

$$I = \int_0^\infty y\mathrm{d}y \int_0^\infty y'\mathrm{d}y'\ A(y)G(y,y')A'(y'). \tag{4.47}$$

We can change the order of integration to first integrate out y and y' to get $B(x)$ and $B'(x)$, respectively, as shown from (4.43) to (4.45), so that

$$I = \int_0^\infty x\mathrm{d}x\ F(x)B(x)B'(x) \approx \Delta \sum_i F_i \sum_j H_{\nu,ij} y_j A_j \sum_{j'} H_{\nu',ij'} y_{j'} A'_{j'}. \tag{4.48}$$

Alternatively, a direct discretization of (4.47) reads

$$I \approx \Delta^2 \sum_{jj'} y_j^2 A_j G_{j,j'} y_{j'}^2 A'_{j'}. \tag{4.49}$$

Since A and A' are arbitrary, by comparing the above two equations we derive the formula to compute $G(y, y')$ on the logarithmic grid

$$G_{j,j'} = \frac{1}{\Delta} \sum_i y_j^{-1} H_{\nu,ij} F_i H_{\nu',ij'} y_{j'}^{-1}. \tag{4.50}$$

This result can also be obtained, formally, by applying the same replacement rule (4.46) to (4.41).

Evaluation of (4.50) is simple and fast, as the formula can be optimized to $\mathcal{O}(N^2 \log N)$ time complexity: constructing the H_ν matrix requires one Fast Fourier Transform per ν; the summation over i is a convolution therefore can be done efficiently with FFT by calling `mcfit`; and our algorithm does not require any additional special function implementations as do Ref. [21, 95].

Furthermore, (4.49) implies that $G_{j,j'}$ approximates $G(y, y')$ as average values in the logarithmic intervals. Generally $G(y, y')$ peaks sharply on the $y = y'$ diagonal and almost vanishes elsewhere, with roughly constant peak width. Our algorithm does not intend to compute the exact value of G at point (y, y') but a smoothed version over the $(y_j, y_{j'})$ grid. This turns out to be beneficial as we are only interested in its coarse-grained values (see Sec. 4.2) rather than the finer structure, and (4.50) allows us to perform the numerical integration without scanning a very fine grid. On the other hand, the method in Ref. [21, 95] evaluates $G(y, y')$ itself, thus need to sample densely around the peak width to arrive at the same coarse-grained result. And because of the logarithmic grid and constant peak width, the sampling rate is determined by the large y end, and resolving the peak at large y leads to waste of computation at small y.

4.4 Simulations

To verify our analytic covariances, we compare them to the empirical results measured from $N_{\mathrm{mock}} = 1000$ realizations of the Multidark-Patchy galaxy mocks [146], produced for BOSS data release 12 (DR12) [15]. The Patchy algorithm are based on the augmented Lagrangian perturbation theory and a stochastic halo biasing scheme calibrated on high-resolution simulations. It then uses halo occupation distribution to construct catalogs to

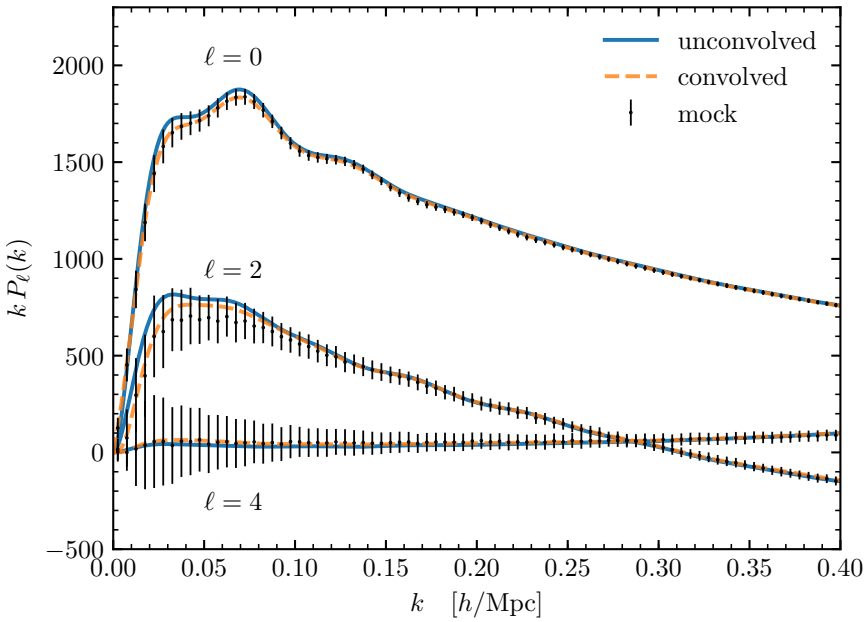


Figure 4.2: Power spectrum multipoles, of 1000 Patchy mocks for NGC of BOSS DR12 in $0.5 < z < 0.75$. The points with errobars show the sample mean and variance of the power spectrum multipoles of the galaxy mocks. The "convolved" curve shows the best-fit RSD model from [105], including the convolution with the appropriate window function [277] to accounts for mixing of power by the window. We also show the "unconvolved" model prior to the convolution. We will use this "unconvolved" model along with the window factors $\mathcal{Q}$ shown in Fig. 4.3 to compute the covariance matrices.

match the observed galaxy clustering and its redshift evolution. In this chapter we use the North Galactic Cap (NGC) region covering 6800 deg^2 of the sky, and two redshift ranges: the first ($0.2 < z < 0.5$) and the third ($0.5 < z < 0.75$) redshift bins of BOSS DR12. The cosmology assumed by the Patchy mocks is $\Omega_\mathrm{m} = 0.307$, $\Omega_\mathrm{b} = 0.048$, $h = 0.678$, $\sigma_8 = 0.829$.

We estimate the redshift-space power spectrum multipoles with the endpoint Yamamoto estimator [279] in the third redshift bin. We employ the FFT-accelerated algorithm [104] enabled by multipole decomposition [248], implemented in the large-scale structure toolkit `nbodykit` [106]. We use the standard Landy-Szalay [156, 242] estimator to estimate the correlation functions in the first redshift bin. Correlation function multipoles are transformed from the correlation function measured in 50 μ bins and the projected correlation function is computed using the ratio of pair counts integrated over the line of sight. Note that our projected correlation function estimator is dimensionless and different from the projected correlation function which has dimensions of length (see e.g. [242] for correlation function with dimensions of length).

For a pair of 2-point observables, denoted by O and O', the unbiased sample covariance

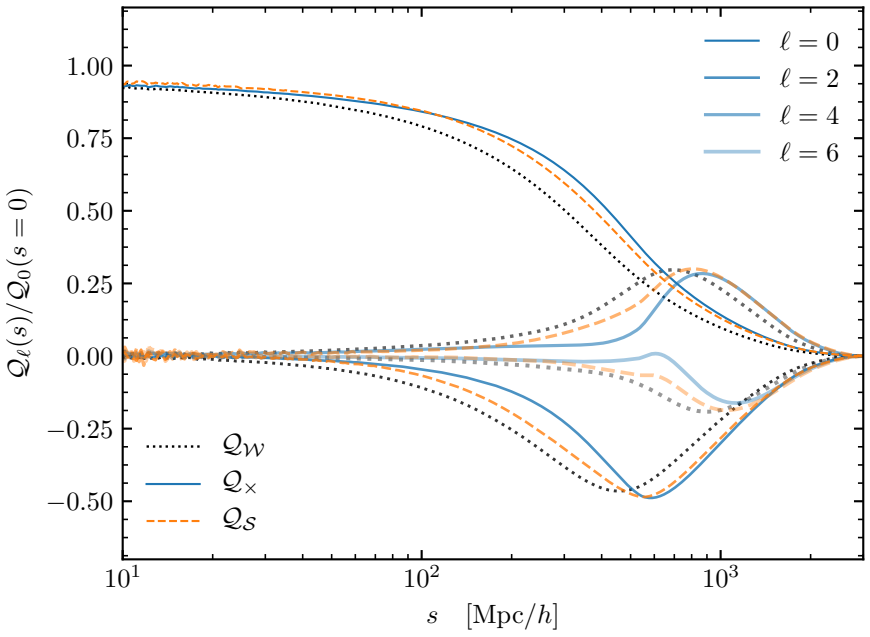


Figure 4.3: Window function multipoles, of 1000 Patchy mocks for NGC of BOSS DR12 in $0.5 < z < 0.75$. The $\mathcal{Q}$ window functions are directly related to the covariance, defined in (6.11), (4.16) and (4.28). We normalize the window multipoles by their zero-lag monopoles to compare their shapes. As the covariance is $\sim (P + P_{\rm shot})^2$, three $\mathcal{Q}$ factors in different colors describe the windows for different pieces: $\mathcal{Q}_\mathcal{W}$ for P^2, $\mathcal{Q}_\times$ for $P \times P_{\rm shot}$, and $\mathcal{Q}_\mathcal{S}$ for $P^2_{\rm shot}$. Their normalized monopoles start from 1 on small scales and vanish on large scales beyond the size of the window. The higher order multipoles are only non-vanishing around the window scale and capture the anisotropy of the window. Combining the unconvolved P_ℓ in Fig. 4.2 and the $\mathcal{Q}_\ell$ here, we can compute the disconnected covariance of $\hat{P}_\ell$ following Sec. 4.2.

matrix from the $N_{\rm mock}$ realizations is

$$\widehat{\rm Cov}(O, O') = \frac{1}{N_{\rm mock} - 1} \sum_{r=1}^{N_{\rm mock}} (\hat{O}_r - \bar{O})(\hat{O}'_r - \bar{O}'), \tag{4.51}$$

where $\bar{O}$ ($\bar{O}'$) is the sample mean of $\hat{O}_r$ ($\hat{O}'_r$) for r over all $N_{\rm mock}$ realizations.

For all the test cases in this chapter, we use the following weight function

$$w = w_{\rm FKP} \times w_{\rm veto}, \tag{4.52}$$

in which $w_{\rm FKP} \equiv 1/(1 + \bar{n} \times 10^4\,{\rm Mpc}^3/h^3)$, and the veto mask excludes certain types of unobservable regions, e.g. near bright stars. To estimate the power spectrum, we use synthetic random catalogs that have 50 times as many points as the mock galaxies, i.e. $\alpha = 0.02$. For correlation functions we use 5 times as many randoms, i.e. $\alpha = 0.2$. And we estimate the window functions using pair counting with $\alpha = 0.2$.

4.5 Results

In this section, we validate our analytic disconnected covariance against the sample covariance measured from the mock simulations. We compare them in two test cases: the power spectrum multipole covariance, and the joint covariance of projected correlation function and correlation function multipoles.

Power Spectrum Multipole Covariance Matrix

We compute the analytic disconnected covariance of $P_\ell(k)$ using the equations in Sec. 4.2, that requires the unconvolved P_ℓ and the window function multipoles $\mathcal{Q}_\ell$ as inputs. To obtain the former, we measure the sample mean and variance of the power spectrum multipoles from the 1000 Patchy mocks, shown in Fig. 4.2. With all model parameters initialized to their fiducial values, we use the RSD model from [105] to compute a diagonal covariance matrix [97] and fit this model to the mock sample mean over the k range from $0.02\,h/\mathrm{Mpc}$ to $0.4\,h/\mathrm{Mpc}$. We obtain the maximum a posteriori (MAP) estimate of the model parameters using the L-BFGS algorithm, and subsequently the diagonal covariance is re-computed using the best-fit model. We then repeat the fitting process until convergence is reached. This analysis is performed with the `pyrsd` package. To account for the window convolution in (6.9), we convolve the model with the appropriate window [277] before fitting it to the mocks. We show both the unconvolved and convolved model curves in Fig. 4.2.

To measure the window functions represented by their multipole moments, as defined in (6.11), (4.16), and (4.28), we use the random catalog of Patchy mocks, and follow the procedure described in Sec. 4.2. As shown from (4.13) to (6.12), because the covariance has a quadratic form in power spectrum and its shot noise, i.e. $\sim (P + P_{\rm shot})^2$, the three different window factors describe the covariance shape of different pieces: $\mathcal{Q}_\mathcal{W}$ for P^2, $\mathcal{Q}_\times$ for $P \times P_{\rm shot}$, and $\mathcal{Q}_\mathcal{S}$ for $P^2_{\rm shot}$. They capture the same survey geometry but differ in the weights, and therefore have similar shapes and different normalizations. Our analytic $\mathrm{Cov}^{\rm disc}$ is derived in the flat-sky limit, so only depends on the even order $\mathcal{Q}_\ell$'s. We find that our results have converged when truncating the multipoles at $\ell_{\rm max} = 10$. In Fig. 4.3, we show the shapes of the first 4 $\mathcal{Q}_\ell$'s after normalizing them by their monopoles at zero lag. Starting from the small scales, the window monopoles are $\lesssim 1$ while the other multipoles are negligible, reflecting the fact that the windows are homogeneous and isotropic in the small scale limit. Moving to larger scales, the monopoles start to decrease, whereas the higher order multipoles rise due to the anisotropy of the window on those scales. And finally all the multipoles vanish beyond the size of the window.

With the unconvolved P_ℓ the $\mathcal{Q}_\ell$, we can evaluate the analytic disconnected covariance of $\hat{P}_\ell$ using the equations in Sec. 4.2. The equations involve double Bessel integrals that we compute using the quadrature method introduced in Sec. 4.3. And finally we average the covariance function in linear k bins of width $0.005\,h/\mathrm{Mpc}$ to obtain the covariance matrix, as described in Sec. 4.2. The computation takes 2 minutes on a single CPU.

In Fig. 4.4, we compare our analytic covariance for power spectrum multipoles, P_0, P_2, P_4, to the mock sample covariance computed using eq. (4.51). Out of the six combinations, we focus on the auto covariance of the P_0, P_2, P_4, and the cross covariance between P_0 and P_2, in different panels. All panels are normalized by P_0 (without shot noise), to present the relative error with respect to the monopole. Because the covariance matrices peak sharply near the k-k' diagonal where the disconnected piece dominates, and are smooth otherwise (as better illustrated by Fig. 4.5), for each 2D matrix we plot a few slices across fixed k near the diagonal ($|k' - k| \lesssim 0.0025\,h/\mathrm{Mpc}$).

Recall that our analytic covariance matrix only has the disconnected contribution, whereas the mock result also contains the connected piece. By subtracting the analytic matrices from the mock ones, we can obtain an estimate of the connected covariance matrices:

$$\widehat{\mathrm{Cov}}^{\mathrm{conn}} = \widehat{\mathrm{Cov}} - \mathrm{Cov}^{\mathrm{disc}}\,. \tag{4.53}$$

We also show the differences in mock and analytic covariance in Fig. 4.4 and find them smooth as expected, implying that the analytic result has accounted for most, if not all, of the disconnected covariance of the mocks.

Having shown the amplitude and shape of the covariance in a few slices near the diagonal, in Fig. 4.5 we present the shape of the full covariance matrix by normalizing it as the linear correlation coefficients given by

$$\mathrm{Corr}(O, O') = \frac{\mathrm{Cov}(O, O')}{\sqrt{\mathrm{Cov}(O, O)\,\mathrm{Cov}(O', O')}}, \tag{4.54}$$

where $\mathrm{Cov} \in \{\mathrm{Cov}^{\mathrm{disc}}, \widehat{\mathrm{Cov}}\}$ for analytic and mock covariances, respectively. The visualized matrix has 3×3 blocks, with the horizontal and vertical blocks corresponding to $O \in \{P_0(k), P_2(k), P_4(k)\}$ and $O' \in \{P_0(k'), P_2(k'), P_4(k')\}$, respectively. Because the covariance is symmetric, we combine the analytic and mock covariances in the top panel, with the upper triangular part showing the analytic covariance and the lower triangular part showing the mock covariance. The diagonals of corresponding blocks in the upper and lower triangles have very similar shapes and scale dependence. We again observe that far from the block diagonals the analytic blocks have vanishing elements, while the mock blocks vary smoothly with some noise.

To visualize the full connected covariance, we normalize the (4.53) by the diagonal of the analytic disconnected covariance

$$\mathrm{Corr}^{\mathrm{conn}}(O, O') = \frac{\widehat{\mathrm{Cov}}^{\mathrm{conn}}(O, O')}{\sqrt{\mathrm{Cov}^{\mathrm{disc}}(O, O)\,\mathrm{Cov}^{\mathrm{disc}}(O', O')}}, \tag{4.55}$$

and show it in the lower panel of Fig. 4.5. Note that it is not normalized properly as a correlation matrix, but correspond to the ratio of connected to disconnected covariance. As was shown in Fig. 4.4, this residual component is smooth, suggesting that the disconnected

part has been cleanly removed. There are some notable remaining low k diagonal features for $\ell = 4$ and $\ell = 2$, and they are likely due to the bias in mock $\hat{P}_\ell$ caused by the sparse angular sampling of the Fourier grid.

Note that the connected covariance estimated with (4.53) also includes possible error in our analytic $\mathrm{Cov}^{\mathrm{disc}}$ as well as the noise in the mock sample covariance. From $\widehat{\mathrm{Cov}}^{\mathrm{conn}}$ we can extract the connected part alone, by exploiting the fact that it is smooth and has a low-rank approximation [185, 111]. We perform principal component analysis on $\mathrm{Corr}^{\mathrm{conn}}$ in (4.55), and find that only the first 4 principal components have eigenvectors with broadband features, while the rest are consistent with noise. Therefore we can replace $\mathrm{Corr}^{\mathrm{conn}}$ with this rank 4 substitute, and rescale it by the denominator in (4.55) to obtain a smooth estimate of $\widehat{\mathrm{Cov}}^{\mathrm{conn}}$, which we denote by $\widehat{\mathrm{Cov}}^{\mathrm{conn,sm}}$. Now if we subtract that from the mock sample covariance, the residual is an estimate of its disconnected covariance from the mocks

$$\widehat{\mathrm{Cov}}^{\mathrm{disc}} = \widehat{\mathrm{Cov}} - \widehat{\mathrm{Cov}}^{\mathrm{conn,sm}}. \tag{4.56}$$

This mock $\widehat{\mathrm{Cov}}^{\mathrm{disc}}$ can be compared directly to the analytic $\mathrm{Cov}^{\mathrm{disc}}$, which we present in Fig. 4.6 as slices across the correlation matrix. The slices are arranged in the same way as in Fig. 4.4. Also shown is the difference between the analytic and mock results, with everything normalized by the analytic $\mathrm{Cov}^{\mathrm{disc}}$ as in (4.55) so that the comparison is fair. This residual contains possible flaw in the analytic $\mathrm{Cov}^{\mathrm{disc}}$ and the error on mock sample covariance. We can quantify the level of the latter with the bootstrapping method, and find it comparable to the residual. Therefore our analytic disconnected covariance is accurate to the extent of the errors on the sample covariance from 1000 mocks.

Before moving on, we examine the dependence of analytic $\mathrm{Cov}^{\mathrm{disc}}$ on the $\mathcal{Q}$ windows. In this chapter we have been modeling and measuring three distinct covariance windows, namely $\mathcal{Q}_\mathcal{W}$, $\mathcal{Q}_\times$, and $\mathcal{Q}_\mathcal{S}$. Given their similarity in shapes (see Fig. 4.3), one possible simplification is to approximate all $\mathcal{Q}$ windows with the same shape. In practice this can be achieved in (4.13) by adding the shot noise to the power spectrum and multiplying the squared sum by a single $\mathcal{Q}$ window, for which we use $\mathcal{Q}_\mathcal{W}$. We compare this simplified treatment with our previous result in Fig. 4.7. Since all results are analytic, the unbinned curves are shown for better visualization. We find that the approximation leads to 20% error in the disconnected covariance. And this error depends on the shot noise magnitude since we have changed the shot noise window, and is larger on small scales where shot noise is more important. Therefore, it is necessary to use three separate $\mathcal{Q}$ windows for accurate evaluation of $\mathrm{Cov}^{\mathrm{disc}}$, especially when shot noise is significant.

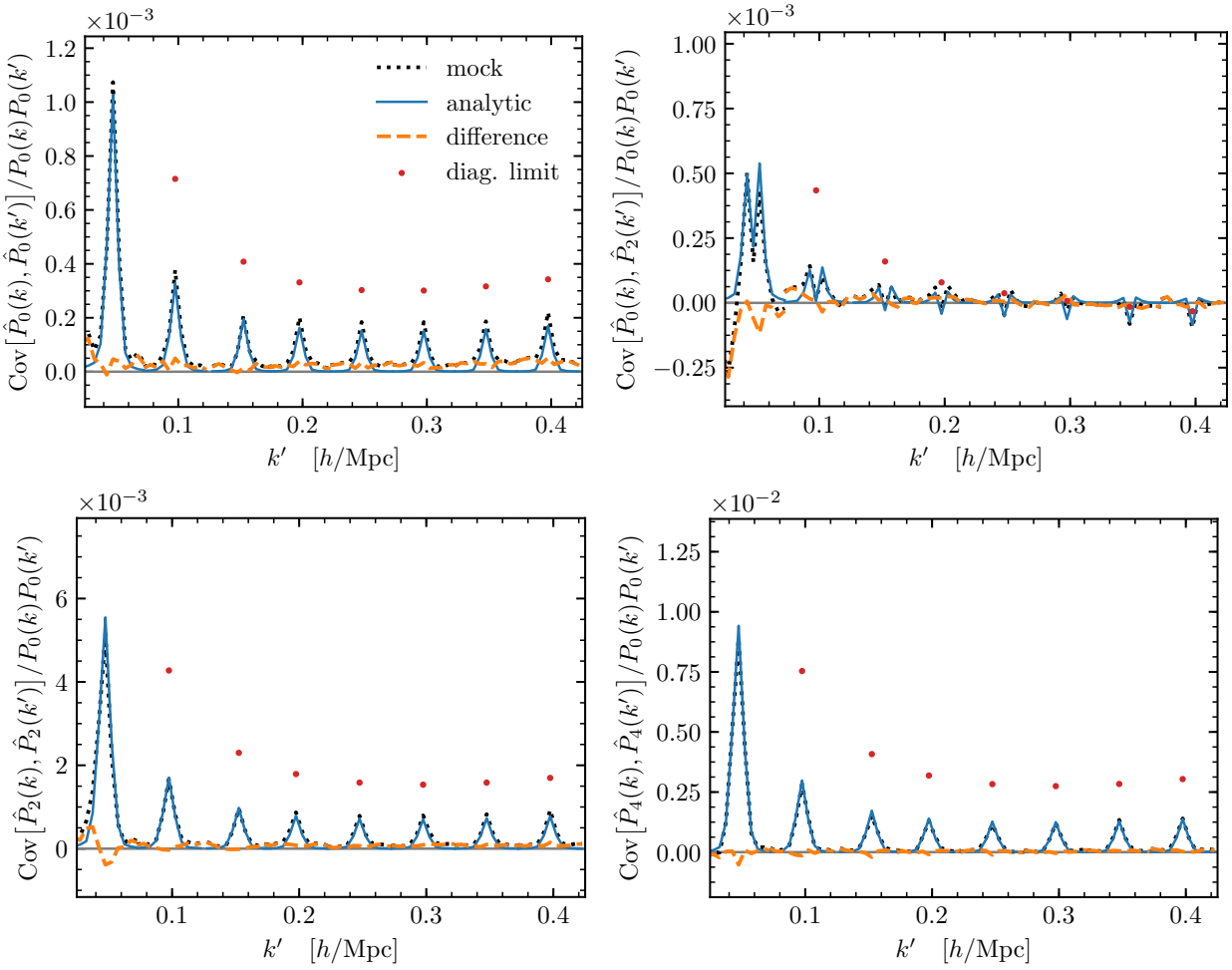


Figure 4.4: Slices of covariance matrices of power spectrum multipoles, of bin width 0.005 h/Mpc, normalized by the monopole. Different panels shows the auto and cross covariance matrices of different multipoles, as labeled next to the vertical axes. Every spike corresponds to a slice of covariance matrix near the diagonal ($k' \approx k$) at fixed k (marked by the position of the peak). The difference (dashed orange) between the mock (dotted black) and analytic (solid blue) covariance is consistent with a smooth component, as expected from the connected covariance in the mocks. This demonstrates that the analytic result has accounted for most, if not all, of the disconnected covariance of the mocks. For comparison we also show the diagonal limit of the analytic covariances from (4.20) in red dots, that ignores the inhomogeneity and anisotropy of the survey window and are only nonzero at the peak of each spike with very different values.

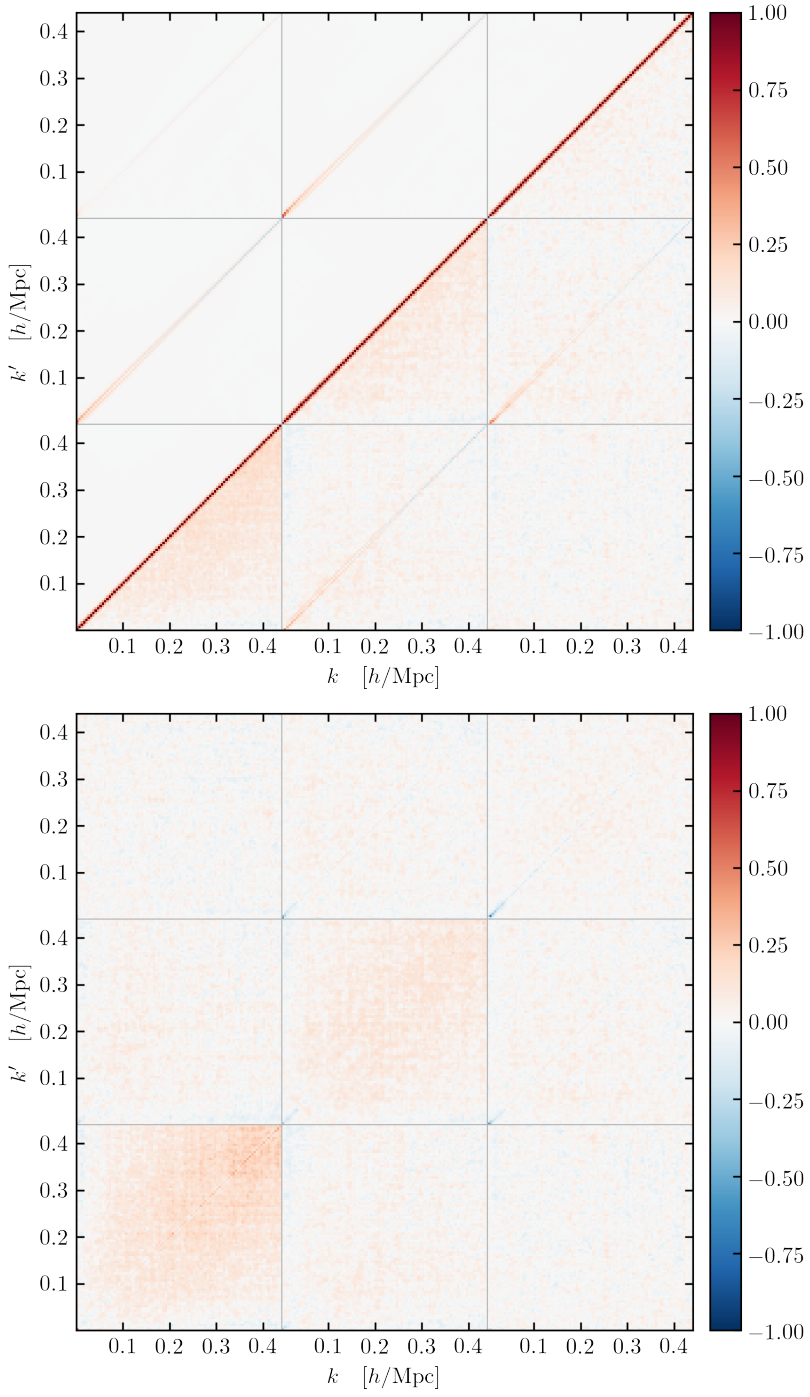


Figure 4.5: Correlation matrix of the power spectrum multipoles. In the 3×3 blocks, from bottom to top and from left to right, we visualize the auto and cross correlations of $\hat{P}_0$, $\hat{P}_2$, $\hat{P}_4$. The top panel compares the analytic result in its upper triangular block with the mocks covariance in the lower triangular block. The bottom panel shows the difference between the mock and analytic covariance matrices, normalized by the diagonal of the latter. As was shown in Fig. 4.4, this residual component is smooth, and captures the connected part of the covariance matrix. There are some remaining low k diagonal features that are more prominent for larger ℓ, and are likely due to the bias in mock $\hat{P}_\ell$ for $\ell > 0$ caused by the sparse angular sampling of the Fourier grid.

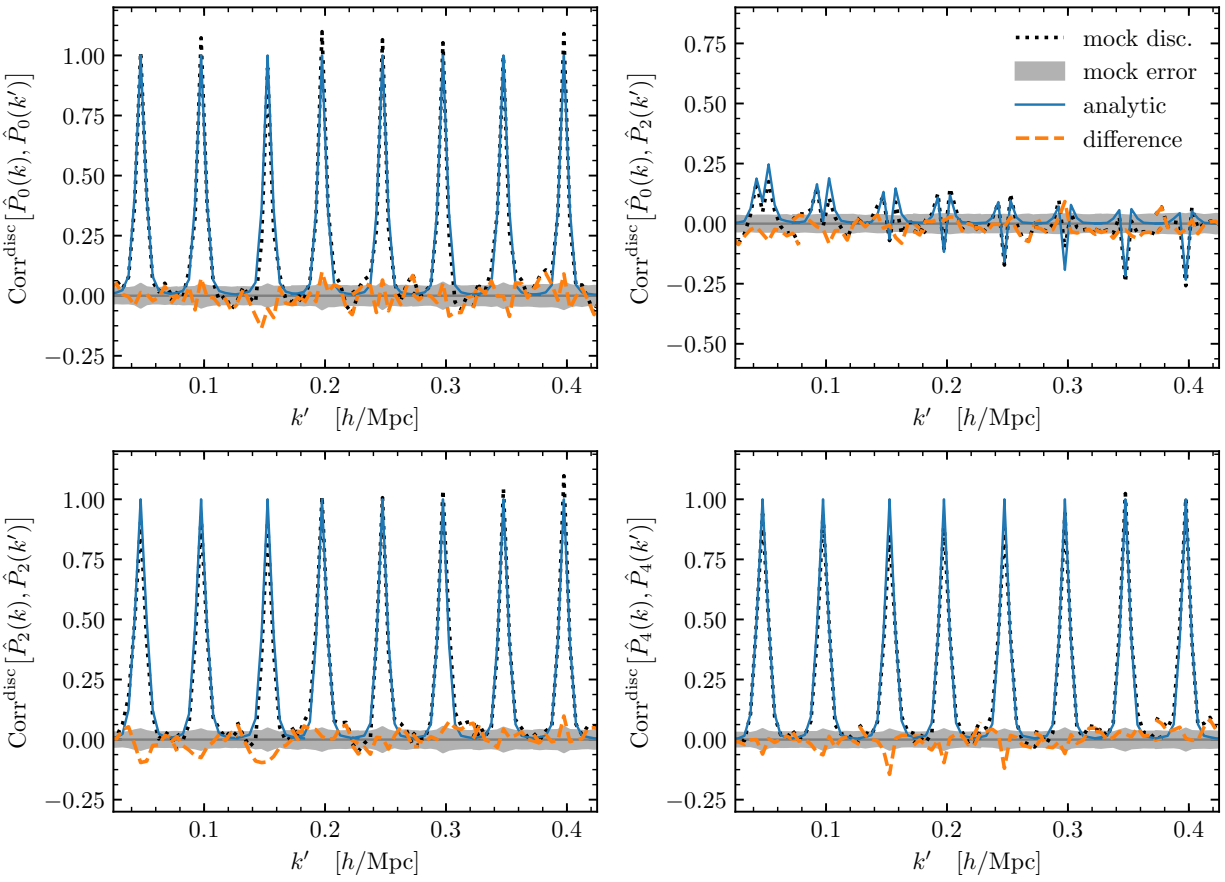


Figure 4.6: Slices of disconnected correlation matrix of power spectrum multipoles. Every spike corresponds to a slice of covariance matrix near the diagonal ($k' \approx k$) at fixed k (marked by the position of the peak). For fair comparison, all covariance matrices are normalized by the diagonal of the analytic one. The mock disconnected covariance is obtained by subtracting the connected part, approximated with the first 4 principal components of the lower panel of Fig. 4.5, from its full covariance. Its difference from the analytic result reflects latter's accuracy and this residual is comparable to the bootstrapping errors on the mock sample covariance (grey band around zero).

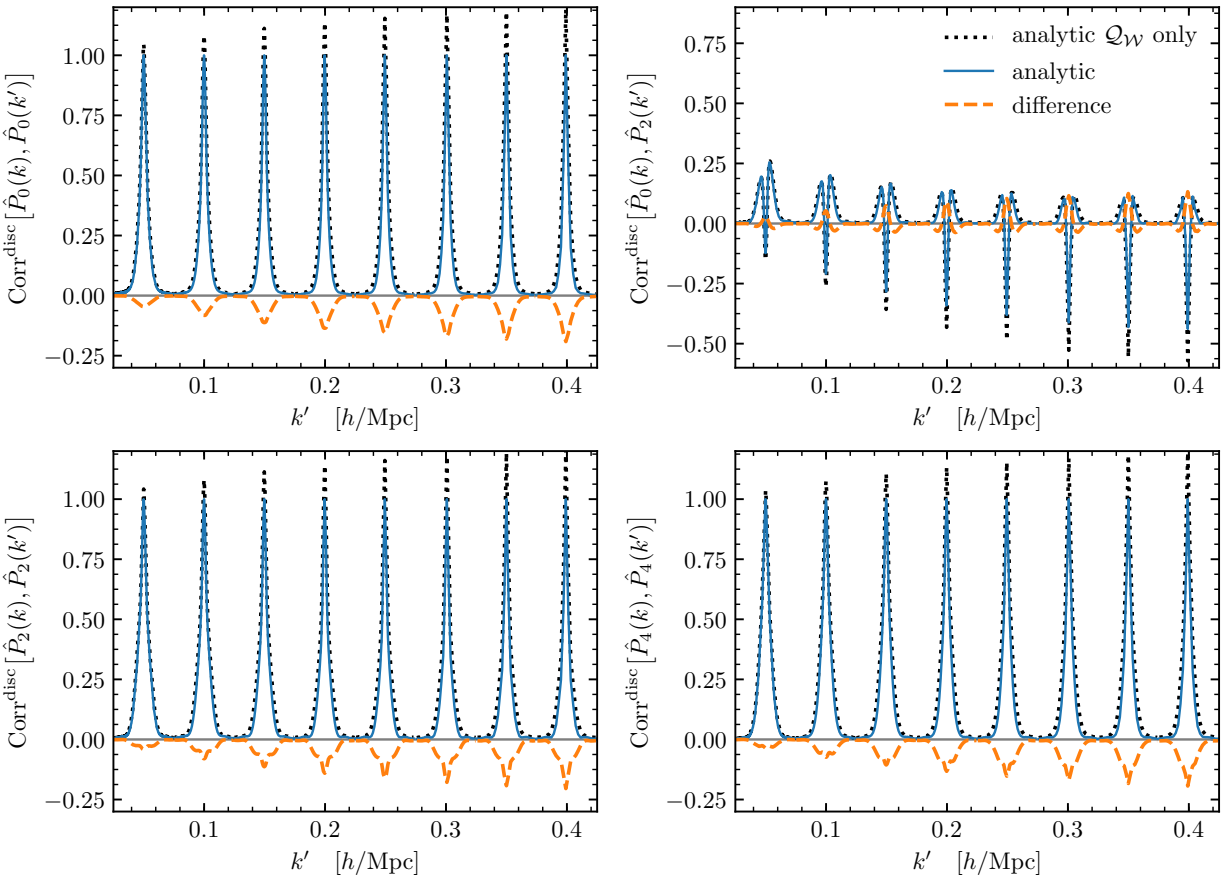


Figure 4.7: Slices of analytic disconnected correlation matrix of power spectrum multipoles, using different $\mathcal{Q}$ functions. The solid blue lines show the same but unbinned result as in previous figures, with which we normalize all the covariance in this figure for fair comparison. It was computed with the window functions $\mathcal{Q}_\mathcal{W}$, $\mathcal{Q}_\times$, and $\mathcal{Q}_\mathcal{S}$, which are shown in Fig. 4.3. To illustrate the necessity of modeling three separate windows, in dotted black line we show an approximation computed by assigning the same shape to all $\mathcal{Q}$'s: $\mathcal{Q}_\times \leftarrow \mathcal{Q}_\mathcal{W}$ and $\mathcal{Q}_\mathcal{S} \leftarrow \mathcal{Q}_\mathcal{W}$. The dashed orange lines give the difference between the two approaches. Although three $\mathcal{Q}$ functions have similar shapes, their differences can leads to 20% difference in the disconnected covariance. This difference depends on the shot noise magnitude, and is larger on small scales where shot noise is more important.

Signal-to-Noise Ratio

Our analytic disconnected covariance $\mathrm{Cov}^{\mathrm{disc}}$ captures the correlation between neighboring k bins and multipoles due to the window effect. In Sec. 4.2 we have derived its diagonal limit $\mathrm{Cov}^{\mathrm{diag}}$, that only captures the survey size but neglect its inhomogeneity and anisotropy. Fig. 4.4 compares the two and shows they can have very different shapes and values. However such comparison is sensitive to binning: coarser bins would suppress the off-diagonal elements of $\mathrm{Cov}^{\mathrm{disc}}$ and reduce the difference between $\mathrm{Cov}^{\mathrm{disc}}$ and $\mathrm{Cov}^{\mathrm{diag}}$ on their diagonals. Here we compare their signal-to-noise ratios, which are independent of binning, thereby study more carefully the impact of the survey window on the information content.

The signal-to-noise ratio of each power spectrum multipole is defined as

$$\left(\frac{S}{N}\right)^2_{\mathrm{disc}} = \sum_{k_i,k_j<k_{\mathrm{max}}} P_\ell(k_i)\,\mathrm{Cov}^{\mathrm{disc}}\left[\hat{P}(k_i),\hat{P}(k_j)\right]^{-1} P_\ell(k_j), \tag{4.57}$$

as a function of the maximum wavenumber k_{max}. The covariance is limited to the disconnected part to compare $\mathrm{Cov}^{\mathrm{disc}}$ and $\mathrm{Cov}^{\mathrm{diag}}$, and P_ℓ are the convolved power spectrum multipoles. Fig. 4.8 shows the results for P_0 and P_2. We find that with the diagonal covariance S/N is underestimated. The difference is more significant on large scales: $\gtrsim 10\%$ below $0.1\,h$/Mpc, and $\approx 5\%$ even between $0.2\,h$/Mpc and $0.4\,h$/Mpc. This trend is expected because the window affect mostly the largest scales in the survey, and is roughly homogeneous and isotropic on small scales as shown in Fig. 4.3. We also show that it further biases S/N on small scales if one ignores the differences among the three effective volumes in (4.18) and uses only $V_{\mathcal{W}}$, due to incorrect normalizations of the shot noise terms.

We warn however that this figure cannot be used for a quantitative estimate of S/N of cosmological parameters, and a more complete analysis that marginalizes over all nuisance parameters, and includes also the connected covariance, is needed. This analysis will be presented in a future work.

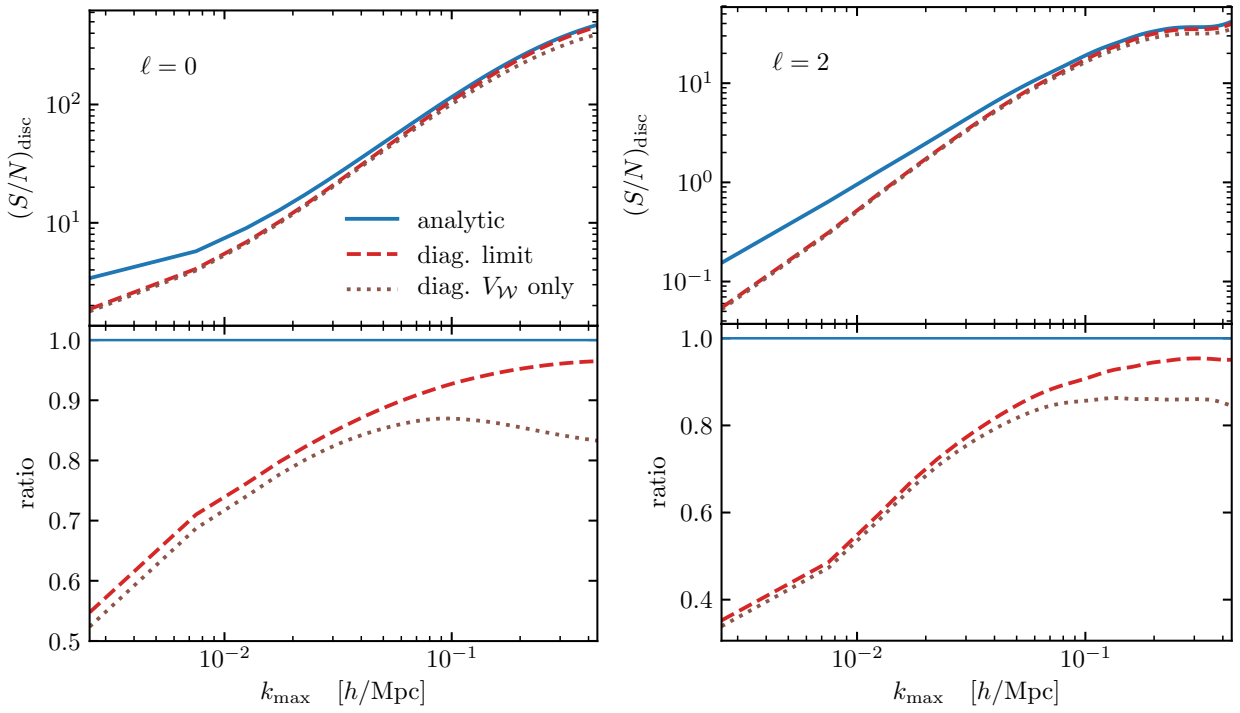


Figure 4.8: The signal-to-noise ratio of the power spectrum monopole and quadrupole with the disconnected covariance. We compare the signal-to-noise ratios using our analytic $\mathrm{Cov}^{\rm disc}$ and its diagonal limit (Sec. 4.2). The latter captures the window size with three effective volumes in (4.18), but ignores the shape of the window, resulting in very different covariance matrices (see Fig. 4.4). It underestimates the signal-to-noise ratio, and the difference is significant on large scales. This trend is expected because the window affects mostly the largest scales in the survey, and is roughly homogeneous and isotropic on small scales (see Fig. 4.3). Also shown is the approximation that uses only $V_{\mathcal{W}}$ for all three effective volumes, which causes large bias on small scales where the shot noise terms are incorrectly normalized.

Projected-Multipole Correlation Function Covariance Matrix

Using the equations derived in Sec. 4.2, we can transform the power spectrum covariance to that of the correlation functions. Here we test this formalism on the joint covariance of the projected correlation function $\xi_\perp$ and correlation function multipoles ξ_ℓ. Note that here $\xi_\perp$ is dimensionless unlike the w_p estimator which is more commonly used and has the dimension of length. While it's possible to predict the covariance of w_p with our formalism, we choose $\xi_\perp$ as the test case as it is simpler to compute (see [242] for derivation of covariance for w_p). $\xi_\perp$ and w_p have similar signals and correlation matrices, as $\xi_\perp$ is effectively w_p with one very large bin of LOS separation, although the amplitudes of the two estimators are different due to the factor of length in w_p. However, the correlation matrices for two estimators can be different for very large LOS integration: the w_p estimator gets larger contribution from noisier

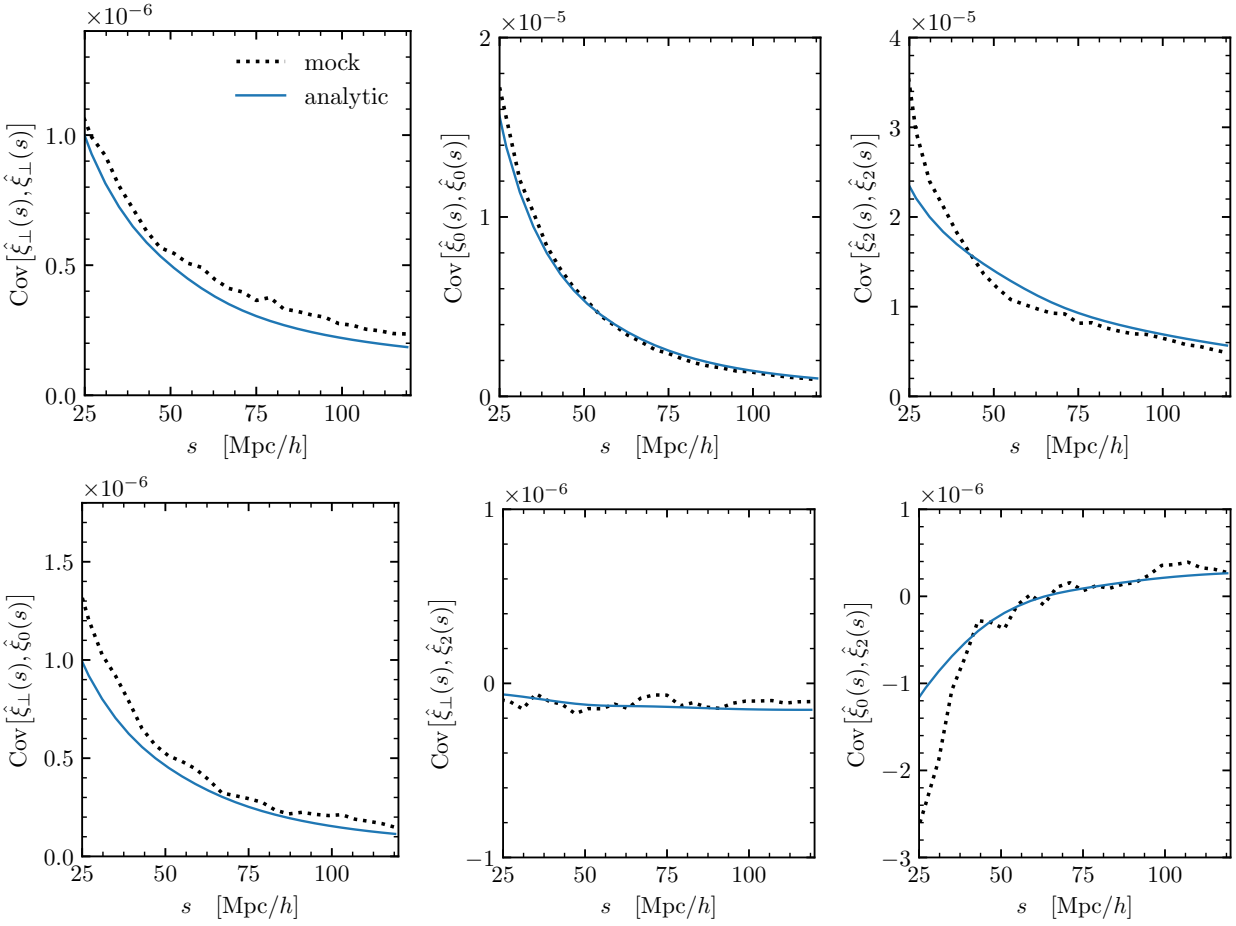


Figure 4.9: The diagonal of the joint covariance matrices of the projected correlation function and correlation function multipoles. We verify that the small-scale deviations in the ξ_ℓ covariances are due to the connected covariance.

bins from large LOS separation, while in $\hat{\xi}_\perp$ those bins are damped by the LOS window.

First, we use the same power spectrum model P_ℓ and window function multipoles $\mathcal{Q}_\ell$ as in Sec. 4.5, to compute the auto covariance of the projected power spectrum $P_\perp$, and the cross covariance between $P_\perp$ and P_ℓ's. Together with the auto covariance of P_ℓ's obtained in Sec. 4.5, we have the full joint covariance of $P_\perp(k)$ and $P_\ell(k)$'s. Note that this is before the binning operation and k is sampled on a logarithmic grid. Then we can perform Hankel transforms on this covariance twice with `mcfit` to replace k with s, and k' with s'. The result is almost the correlation function covariance, except for the difference in the normalization factors arising from the $\hat{P}$ and $\hat{\xi}$ estimators. We account for this normalization difference, before binning the covariance into matrices with s bins of width $4\,\mathrm{Mpc}/h$. The whole process takes less than 30 minutes on a single CPU, and is more than ten times as long as the computation for the power spectrum. Most of the time is spent on the $\xi_\perp$ cross ξ_ℓ blocks

where we approximate a Dirac delta function with multipole expansion truncated at $\ell_{\rm max} = 16$, much higher than the power spectrum case. The computation can be significantly faster when the cross covariance is not needed.

Again we compare the analytic correlation function $\mathrm{Cov}^{\rm disc}$ to the sample covariance from mocks. Fig. 4.9 shows this comparison on the $s = s'$ diagonal of the auto covariance matrices. We find the analytic $\mathrm{Cov}^{\rm disc}$ reproduce well the scale dependences. On small scales $\lesssim 50\,\mathrm{Mpc}/h$, the mock covariances of ξ_0 and ξ_2 are larger than the analytic ones, and we find these excesses are due to the connected covariance by transforming the power spectrum $\widehat{\mathrm{Cov}}^{\rm conn,sm}$ from Sec. 4.5. Apart from that, the analytic and mock covariance still differ by an offset for $\xi_\perp$ and a 15% deficit near $50\,\mathrm{Mpc}/h$ for ξ_2. This is probably due to the fact that the analytic $\mathrm{Cov}^{\rm disc}$ is based on the best-fit P_ℓ to the mock power spectrum, that in general can be somewhat different if fitted against the correlation function measurement.

We also compare the shapes of the analytic and mock covariance by showing their correlation matrices in Fig. 4.10. The top panel shows the two have very similar shapes, and their difference (normalized by the diagonal of the analytic $\mathrm{Cov}^{\rm disc}$ similar to Fig. 4.5) is $\lesssim 10\%$ except in the regions noted above.

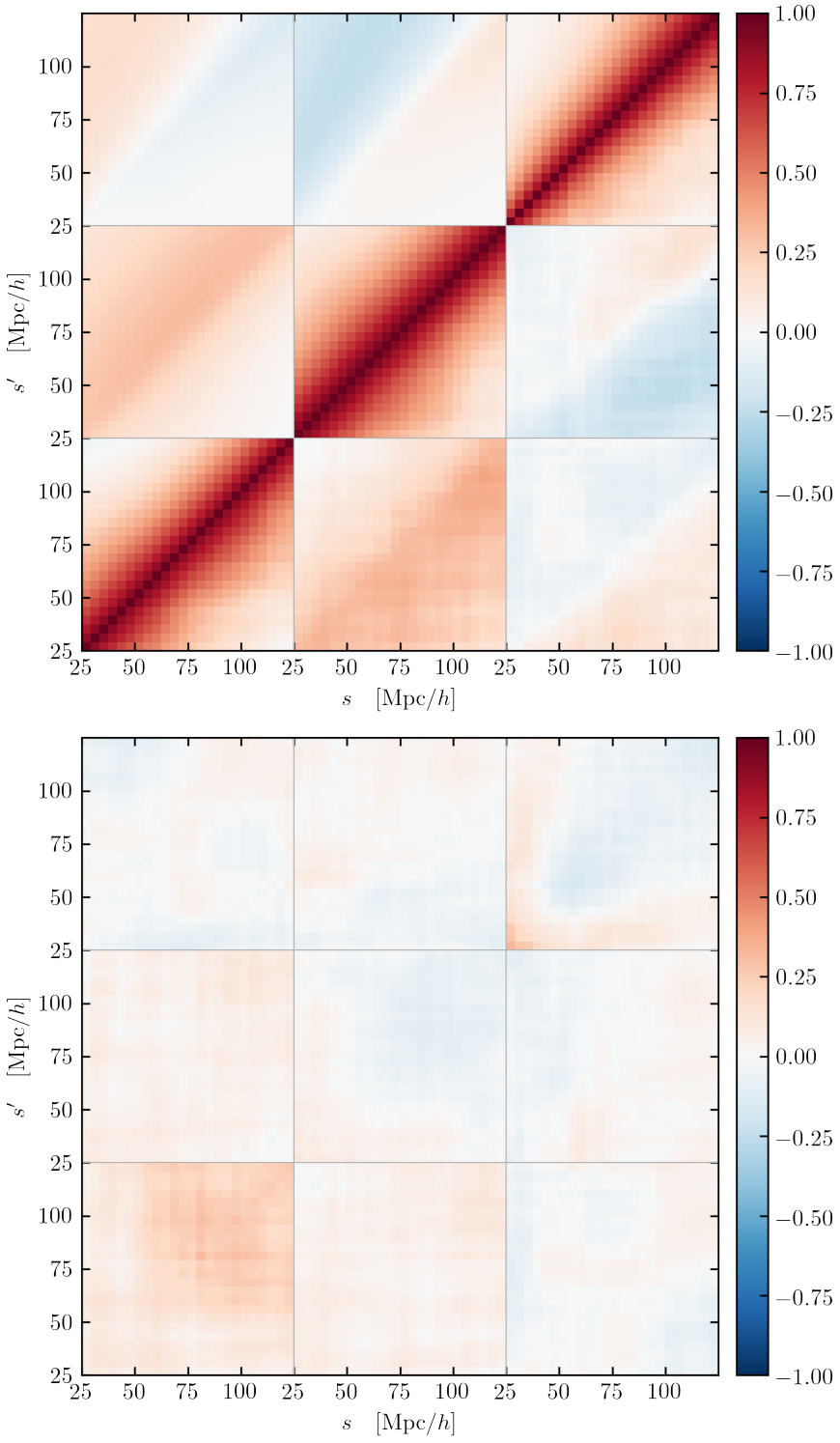


Figure 4.10: Correlation matrix of of the projected correlation function and correlation function multipoles. In the 3×3 blocks, from bottom to top and from left to right, we visualize the auto and cross correlations of $\hat{\xi}_\perp$, $\hat{\xi}_0$, $\hat{\xi}_2$. The upper figure compares the analytic result in its upper left corner with the mocks in the lower right corner. And the lower figure shows the difference between the mock and analytic covariance matrices, normalized by the diagonal of the latter.

4.6 Discussion

We presented an analytic method to compute the disconnected part of the covariance matrix of 2-point functions in large-scale structure studies, accounting for the survey window effect. This method works for both power spectrum (Sec. 4.2) and correlation function (Sec. 4.2), and applies to the covariances for various probes including the multipoles (Sec. 4.2) and the wedges (Sec. 4.2) of 3D clustering, the angular and the projected statistics of clustering and lensing, as well as their cross covariances.

We verify our analytic covariance against the covariance from 1000 galaxy mock simulations for BOSS DR12. We compare the analytic and mock covariance matrices on power spectrum multipoles, and demonstrate that we obtain an excellent agreement without the sampling noise associated with numerical covariance matrices. As a consequence, our method does not require various corrections and inflated errors that have been developed for this purpose [112, 238].

Another advantage of our method is that its predictions use the best-fit power spectrum model to the data, and does not assume a fiducial model that may not fit the data well. The method is thus best viewed as part of a full data analysis pipeline, where we determine both the power spectrum and the covariance matrix within the same (iterative) procedure. We also tested the joint covariance matrices of the projected correlation function and correlation function multipoles, and find the accuracy of the analytic prediction is satisfactory, given that our model is fitted to the power spectrum instead of the correlation function. The analytic computation is efficient and costs negligible CPU time compared to the mocks.

In contrast to previous work [97] our method includes the window effect on the covariance. This commonly adopted approximation to the disconnected covariance only captures the size of the survey window but ignores its shape (inhomogeneity and anisotropy), resulting in a diagonal power spectrum covariance matrix. We show the proper diagonal limit arises as the homogeneous and isotropic limit of our analytic covariance, and it generalizes the usual expression by accounting for the inhomogeneous number density. We show that such a diagonal covariance underestimates the signal-to-noise ratio compared to our analytic covariance. Other previous works modeled the window effect to different extents. In Ref. [122], the authors accounted for the window function in the diagonal elements of the power spectrum covariance matrix in the flat-sky limit, by using FFT and pair summation over Fourier modes, but they needed an ansatz to approximate the off-diagonal elements that again ignored the window. Ref. [42] presented an analytic calculation of the disconnected covariance of the galaxy power spectrum multipoles with variable LOS, under some approximations that simplify the coupling between the power and the window. Our formalism assumes flat sky and similar simplified power-window coupling, but uses the curved-sky window to account for the spherical geometry. Compared to our method, the equations in ref. [42] are more complicated and need Monte-Carlo integration to evaluate. Ref. [242] also derived the expression for the covariance and cross covariance for 3D clustering and projected correlation functions, but assumed uniform window when computing the covariance matrices. Our method generalizes the disconnected covariance for arbitrary windows which are normally encountered in the

large-scale structure surveys.

Being able to compute the accurate disconnected covariance analytically opens the possibility of calibrating the connected part using only small-volume mocks or internal covariance estimators from the data, thereby substantially reduces the computational cost required for estimating the full covariance matrices. We will address this connected part of covariance matrix in the future work. Having the complete noiseless covariance will allow for the optimal analyses of the 2-point functions from the large-scale structure data.

Chapter 5

Variational Inference with L_2 optimization

Standard Variational Inference (VI) optimizes the expectation lower bound objective (ELBO). In higher dimensions it must be evaluated using stochastic integration of KL divergence, which has large sampling noise. In this chapter, we propose instead to minimize the expectation of L_2 distance squared between the gradient of approximating log posterior $q_\theta(\boldsymbol{z})$ and the log of the joint distribution $p(\boldsymbol{x}, \boldsymbol{z})$. This expectation of L_2 objective (EL_2O) reduces the sampling noise, and as a result has better convergence properties. We apply the method to a full rank Gaussian ansatz for $q_\theta(\boldsymbol{z})$, and further extend it to point-wise nonlinear transformations. We present several examples, which show that EL_2O converges faster than ELBO based Stochastic Variational Inference such as ADVI. These include multi-variate Gaussian of different dimensionality, as well as a scientific inverse problem example of a 13 dimensional galaxy clustering analysis with an expensive nonlinear forward model, in both cases EL_2O being orders of magnitude faster than ADVI and MCMC. EL_2O residual also provides an estimate of the error, and can be used as an initialization of MCMC.[1]

5.1 Introduction

A general statistical inference of an inverse problem is how to infer parameters from the data: we have some data $\boldsymbol{x} = \{x_i\}_{i=1}^N$ and some parameters the data depend on, $\boldsymbol{z} = \{z_j\}_{j=1}^M$, with a forward model $\boldsymbol{m}(\boldsymbol{z})$ relating the two. In Bayesian methodology we describe the uncertainty quantification as a posterior of $\boldsymbol{z}$ given data $\boldsymbol{x}$. We can define the posterior $p(\boldsymbol{z}|\boldsymbol{x})$ as

$$p(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{z})}{p(\boldsymbol{x})}, \quad \text{where} \quad p(\boldsymbol{x}) = \int p(\boldsymbol{x}, \boldsymbol{z}) d\boldsymbol{z}. \tag{5.1}$$

[1] This chapter is taken from "Posterior inference unchained with EL_2O," *Seljak U. and Yu B.* (arXiv: 1901.04454).

Here the joint is $p(\boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})$, where $p(\boldsymbol{x}|\boldsymbol{z})$ is the likelihood of the data, $p(\boldsymbol{z})$ is the prior of $\boldsymbol{z}$ and $p(\boldsymbol{x})$ is the normalization or evidence. In general we have access to the joint, i.e. the prior and the likelihood, but not the normalization $p(\boldsymbol{x})$. In many inverse problems the likelihood is often written as a Gaussian noise probability model between the data and the forward model $\boldsymbol{m}(\boldsymbol{z})$, which maps from the parameters $\boldsymbol{z}$ to the data space $\boldsymbol{x}$, so $p(\boldsymbol{x}|\boldsymbol{z}) = N(\boldsymbol{m}(\boldsymbol{z}) - \boldsymbol{x}, \boldsymbol{\sigma^2})$, where $\boldsymbol{\sigma^2}$ is a vector of noise variances σ_i^2, which in general can depend on data index i.

For many scientific applications the model $\boldsymbol{m}(\boldsymbol{z})$ is expensive to evaluate, and often we do not have the gradient information with respect to all parameters $\boldsymbol{z}$. For example, the model could be the outcome of a complicated ODE or PDE numerical solution. Note that the gradient of the log likelihood is easy to evaluate if we have access to the gradient of the model, $\nabla_{\boldsymbol{z}} \ln p(\boldsymbol{x}|\boldsymbol{z}) = -\nabla_{\boldsymbol{z}} \boldsymbol{m}(\boldsymbol{z}) \cdot [(\boldsymbol{m}(\boldsymbol{z}) - \boldsymbol{x})/\boldsymbol{\sigma}^2]$, while the likelihood component of the Hessian can often be well approximated with the Gauss-Newton method $-\nabla_{\boldsymbol{z}} \nabla_{\boldsymbol{z}} \ln p(\boldsymbol{x}|\boldsymbol{z}) \approx \nabla_{\boldsymbol{z}} \boldsymbol{m}(\boldsymbol{z})^T \boldsymbol{\sigma}^{-2} \nabla_{\boldsymbol{z}} \boldsymbol{m}(\boldsymbol{z})$. To this one must also add the prior component of the Hessian $-\nabla_{\boldsymbol{z}} \nabla_{\boldsymbol{z}} \ln p(\boldsymbol{z})$, which vanishes for flat priors.

The final result of a scientific inverse problem analysis is a posterior of individual parameters marginalized over all other parameters. The posteriors are best parametrized in terms of 1D probability density distributions. Occasionally we also want to examine higher dimensional posteriors, such as 2D posteriors, but we rarely go to higher dimensions due to the difficulty of visualizing it. While we want to summarize the results in a series of 1D and 2D posterior plots, the actual problem can have a large number of dimensions M, many of which we may not care about, but which are correlated with the ones we do care about. The main difficulty in obtaining reliable lower dimensional posteriors lies in the marginalization part: marginals, i.e., averaging over the probability distribution of other parameters, can change the answer significantly relative to the answer for the unmarginalized posterior where those parameters are assumed to be known.

A standard approach to posterior estimation is Monte Carlo Markov Chain (MCMC) sampling, which gives samples in M-dimensional space, which can directly give 1D and 2D marginal density estimates with a simple application of a kernel density estimation. However, in many applications the cost of evaluating the likelihood can be very high, and one typically requires 10^5 or more likelihood evaluations, making MCMC often too expensive to be feasible, even in its more efficient forms such as Hamiltonian Monte Carlo. Alternatives to MCMC are approximate methods such as Maximum A Posteriori (MAP) estimation, where we search for the peak posterior. This can be supplemented by the Laplace approximation, where we evaluate the Hessian at MAP, to give a Gaussian approximation to the posterior. While this is often sufficient, per central limit theorem argument, there are situations where the posteriors are non-Gaussian. Moreover, the first order Taylor expansion at the peak may be inaccurate, and can lead to unreasonable results, if for example the curvature (Hessian) changes a lot from the peak to the nearby regions.

An alternative is KL divergence minimization based Variational Inference (VI) [275, 48]. In variational approaches we approximate the posterior $p(\boldsymbol{z}|\boldsymbol{x})$ with $q_\theta(\boldsymbol{z})$ parametrized by parameters θ we optimize for via KL divergence optimization. VI can be less expensive than

MCMC, but can also give inaccurate results and must be used with care: for example, a mean field VI can be inferior to the MAP + Laplace where Hessian is a full rank matrix, since the 1D marginals will ignore the correlations between the parameters and lead to too narrow posteriors. Increasing the expressivity of the posterior approximation $q_\theta(\boldsymbol{z})$ in VI makes it more accurate, but also increases the computational cost.

A recent development is stochastic VI such as Automatic Differentiation Variational Inference (ADVI) [153], which uses the evidence lower bound (ELBO) to optimize on parameters θ using gradient optimization. As the authors acknowledge, full rank ADVI can be prohibitively slow in high dimensions. As we show in this chapter, a property of ELBO stochastic VI optimization is an inherent sampling noise present in KL divergence integration. The integrand of a KL divergence can be positive or negative, such that the integral can be zero, but for a finite number of sampling points there is integration noise scaling as number of samples $N_k^{-1/2}$.

In this chapter we propose is to replace KL divergence optimization (ELBO) with optimization of L_2 norm between the logs of the two distributions evaluated at the same sampling point: this is only zero if the two distributions are equal (up to normalization) at each sample. If gradient is available we can minimize the L_2 norm between the gradients of the two distributions, in which case the normalization drops out. This L_2 norm based divergence is less noisy and we show it converges significantly faster than KL divergence based standard VI. An interesting aspect of L_2 norm is that it corresponds to the second cumulant of the importance weighting cumulant expansion. However, we generalize this concept to the common situation where gradients are available, in which case it becomes a Fisher divergence [103].

From a practical perspective, our main motivation is to develop a method that starts close to MAP + Laplace, and when that is the correct solution, it remains there, only updating it if needed. We will show that stochastic VI such as ADVI [153] is noisy and often fails to converge even when it starts from the correct solution, while our proposed EL_2O approach converges rapidly. EL_2O also provides an estimate of the error. If large at the solution EL_2O can be used as a starting point for MCMC. For example, it can provide a mass matrix for Hamiltonian MC [34], or a sampling proposal matrix for Metropolis-Hastings MC.

5.2 KL divergence versus Fisher divergence

We can define $V(\boldsymbol{x}, \boldsymbol{z}) = \ln q_\theta(\boldsymbol{z}) - \ln p(\boldsymbol{x}, \boldsymbol{z})$ and use the importance sampling identity to derive the following cumulant expansion [196]

$$\ln p(\boldsymbol{x}) = \ln \int d\boldsymbol{z} p(\boldsymbol{x}, \boldsymbol{z}) = \ln \int d\boldsymbol{z} q_\theta(\boldsymbol{z}) \frac{p(\boldsymbol{x}, \boldsymbol{z})}{q_\theta(\boldsymbol{z})} \equiv \ln \mathbb{E}_q \frac{p(\boldsymbol{x}, \boldsymbol{z})}{q_\theta} =$$
$$\ln \mathbb{E}_q \exp[-V(\boldsymbol{x}, \boldsymbol{z})] = -\mathbb{E}_q V(\boldsymbol{x}, \boldsymbol{z}) + \frac{1}{2} \mathbb{E}_q [V(\boldsymbol{x}, \boldsymbol{z}) - \mathbb{E}_q V(\boldsymbol{x}, \boldsymbol{z})]^2 + ... \quad (5.2)$$

where in the last expression we only explicitly wrote the first two cumulants. The first cumulant

$$\text{ELBO} \equiv -\mathbb{E}_q V(\boldsymbol{x}, \boldsymbol{z}) = \ln p(\boldsymbol{x}, \boldsymbol{z}) - \ln q_\theta(\boldsymbol{z}), \tag{5.3}$$

is the evidence lower bound, which is optimized in the standard VI, and is the result of the minimization of KL divergence between the approximate posterior $q(\boldsymbol{z})$ and true, but unknown, posterior $p(\boldsymbol{z}|\boldsymbol{x})$.

The second cumulant in equation 5.2 is an L_2 norm between $\ln q_\theta(\boldsymbol{z})$ and $\ln p(\boldsymbol{x}, \boldsymbol{z})$. This can be optimized too, but depends on the first cumulant ELBO, which can be optimized together with the parameters θ,

$$\text{EL}_2\text{O} = \min_\theta \mathbb{E}_q[(\ln p(\boldsymbol{x}, \boldsymbol{z}) - \ln q_\theta(\boldsymbol{z}) - \text{ELBO})^2]. \tag{5.4}$$

This loss has an L_2 form. The proposal of this chapter is to replace the first cumulant ELBO optimization with the second cumulant optimization.

Often however we also have available gradients $\nabla_{z^i} \ln p(\boldsymbol{x}, \boldsymbol{z})$. At each sample $\boldsymbol{z}$ this gives an additional M constraints, and when M is large it provides a lot more information than the joint $\ln p(\boldsymbol{x}, \boldsymbol{z})$ itself. When gradients are available the proposal of this chapter is to replace the first cumulant ELBO optimization with the expectation of L_2 optimization (EL_2O) of gradients between $\ln q(\boldsymbol{z})$ and $\ln p(\boldsymbol{x}, \boldsymbol{z})$. This is independent of the ELBO normalizing constant, which does not dependent on $\boldsymbol{z}$. It takes advantage of the gradient information as a pointwise sum over L_2 norm of all the gradients with respect to z^i,

$$\text{EL}_2\text{O} = M^{-1} \min_\theta \mathbb{E}_{\tilde{q}} \left\{ \sum_{i=1}^{M} [\nabla_{z^i} \ln q_\theta(\boldsymbol{z}) - \nabla_{z^i} \ln p(\boldsymbol{x}, \boldsymbol{z})]^2 \right\}, \tag{5.5}$$

where the averaging is done over the samples $\boldsymbol{z}_k$ drawn from a previous iteration of q_θ which we denote $\tilde{q}$, and the sum is over all M gradient terms. The corresponding divergence is called Fisher divergence [103]. If we denote each sample as $\boldsymbol{z}_k$ then the above expression is

$$\text{EL}_2\text{O} = (N_k M)^{-1} \min_\theta \left\{ \sum_{k=1}^{N_k} \sum_{i=1}^{M} [\nabla_{z^i} \ln q_\theta(\boldsymbol{z}_k) - \nabla_{z^i} \ln p(\boldsymbol{x}, \boldsymbol{z}_k)]^2 \right\}. \tag{5.6}$$

If $q_\theta(\boldsymbol{z})$ is modeled as a full rank Gaussian $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the optimization with respect to its parameters can be evaluated analytically, and the resulting algorithm is given in algorithm 1.

Expectation of these equations has been derived in the context of variational methods [197], showing that the solution to EL_2O is the same as ELBO optimization in the high sample limit. However, optimizing EL_2O differs from optimizing ELBO with respect to the parameters θ of $q_\theta(\boldsymbol{z})$. One difference is that unlike in ELBO optimization we do not need to propagate the gradients with respect to the samples z_k using the reparametrization trick: we can view the samples $\boldsymbol{z}_k$ from $q_\theta(\boldsymbol{z})$ as being samples from a previous iteration of q_θ, which we denote $\tilde{q}$. One then minimizes the L_2 loss without having to propagate the gradient with respect to θ through the samples $\boldsymbol{z}_k$. The lower the EL_2O value is the better approximation q_θ

Algorithm 1 Full Rank EL_2O

Input: data x_i, size N
Find a solution near MAP. Initialize $\tilde{q}$ with Laplace $\mathbf{\Sigma}$. Draw $N_k = M$ samples $\boldsymbol{z}_k$.
while EL_2O value has not converged **do**
  Draw a new sample z_{N_k+1} from $\tilde{q}$. Increase N_k by 1.
  $\mathcal{H}_{ij} = \sum_{k=1}^{\mathrm{N_k}}(z_{k,i} - \mu_i)(z_{k,j} - \mu_j)$
  $\Sigma_{ij}^{-1} = -\sum_{k=1}^{\mathrm{N_k}}[(z_{k,i} - \mu_i)\nabla_{z_j} \ln p(\boldsymbol{x}, \boldsymbol{z}_k) + (z_{k,j} - \mu_j)\nabla_{z_i} \ln p(\boldsymbol{x}, \boldsymbol{z}_k)]/2\mathcal{H}_{ij}$
  $\boldsymbol{\mu} = N_{\mathrm{k}}^{-1} \sum_{k=1}^{\mathrm{N_k}}[\boldsymbol{z}_k + \mathbf{\Sigma}\nabla_{\boldsymbol{z}} \ln p(\boldsymbol{x}, \boldsymbol{z}_k)$
  Compute EL_2O; update $\tilde{q} = N(\boldsymbol{\mu}, \mathbf{\Sigma})$
end while

is to the posterior. Another important difference is that both $q_\theta(\boldsymbol{z})$ and $p(\boldsymbol{x}, \boldsymbol{z})$ are evaluated at the same position of a given sample $\boldsymbol{z}_k$. We will show that this eliminates the sampling noise inherent in MC integration of KL divergence that underlies ELBO optimization. For example, the presence of $\boldsymbol{z}_k$ at the end of algorithm 1 guarantees there is no sampling noise, while replacing it with its expectation $\boldsymbol{\mu}$ there is sampling noise. We thus obtained a set of equations that use gradient information, but these are different from stochastic VI equations of ADVI [153]. When optimizing for a large number of parameters, such as full rank $q_\theta(\boldsymbol{z})$, the difference in convergence between the two becomes significant, as we show in the examples. EL_2O optimization converges fast even in high dimensions, in contrast to the full rank ELBO optimization in ADVI.

If we have no access to gradients we can still apply the method using a finite difference version of the gradient evaluated between two nearby points $\boldsymbol{z}_k$ and $\boldsymbol{z}_k + \delta\boldsymbol{z}_k$. The accuracy of the finite difference gradient in approximating the actual gradient is not important, since the same two points are chosen to evaluate $\ln q$ and $\ln p$.

A Mean Field Gaussian example

To explain the motivation and the differences between ELBO and EL_2O we will for simplicity assume we only have a single parameter z given the data $\boldsymbol{x}$. We would like to fit the posterior of z to a simple form such as a Gaussian $q_\theta(z) = N(\mu, \Sigma)$. We will assume the true posterior is a Gaussian, $p(z|\boldsymbol{x}) = N(\mu_t, \Sigma_t)$, where the subscripts t denotes the true value, but as this is something we do not know in advance we perform the estimation of parameters of μ and Σ and study their convergence to the true values as a function of number of samples.

In stochastic VI ELBO optimization, as for example automatic differentiation variational inference (ADVI) [153], we draw samples from the current etimate of $q_\theta(z)$ and use these to update the ELBO gradient. One can define the samples as $z_k = \Sigma^{1/2}\epsilon_k + \mu$, where ϵ_k is a

random number drawn from a unit variance zero mean Gaussian $N(0,1)$. With this we find

$$\text{ELBO} \equiv -\mathbb{E}_q V(\boldsymbol{x},\boldsymbol{z}) \approx N_k^{-1} \sum_k \left[\frac{\epsilon_k^2}{2} + \frac{\ln(2\pi\Sigma)}{2} + \ln p(\boldsymbol{x}, z_k) \right], \tag{5.7}$$

where N_k is the number of samples. We want to use gradient of ELBO with respect to μ to update the information on μ, but it only enters via z_k inside $\ln p(\boldsymbol{x}, z_k)$. So if we want to optimize ELBO with respect to μ by making its gradient equal zero we have to propagate its derivative through z_k, the so called reparametrization trick [144, 218], which gives for this example $\nabla_\mu \ln p(\boldsymbol{x}, z) = \nabla_z \ln p(\boldsymbol{x}, z) = -(z_k - \mu_t)/\Sigma_t$. With $z_k = \Sigma^{1/2}\epsilon_k + \mu$ we find the mean μ where the gradient is zero is

$$\sum_k \frac{(z_k - \mu_t)}{\Sigma_t} = 0 \;\rightarrow \mu = \mu_t - N_k^{-1} \sum_k \Sigma^{1/2}\epsilon_k. \tag{5.8}$$

Since the mean of ϵ_k is zero this will converge to the correct answer μ_t, but the convergence will be stochastic, converging as $N_k^{-1/2}$. Even if we start at the correct solution provided by MAP+Laplace, the stochastic nature of KL divergence leads to fluctuations in the subsequent iterations.

To solve for the variance we similarly take a gradient of ELBO with respect to Σ and set it to zero, with solution

$$\Sigma = \frac{N_k \Sigma_t}{\sum_k [\epsilon_k^2 + (\mu - \mu_t)\Sigma^{-1/2}\epsilon_k]}. \tag{5.9}$$

This also leads to a stochastic convergence to the true answer as $N_k^{-1/2}$, but with a larger prefactor, as shown in figure 5.1.

To understand the advantages of EL_2O let us look at its solution of the same 1D Gaussian case example above. We again assume we have access to the gradient $\nabla_z \ln p(\boldsymbol{x}, z)$, as we assumed for ELBO. The EL_2O objective is equation 5.5

$$\text{EL}_2\text{O} = \arg\min_{\mu,\Sigma} N_k^{-1} \sum_k \left[\frac{\mu - z_k}{\Sigma} - \nabla_z \ln p(\boldsymbol{x}, z_k) \right]^2. \tag{5.10}$$

Since for the toy example $\nabla_z \ln p(\boldsymbol{x}, z) = -(z_k - \mu_t)/\Sigma_t$ using gradient of EL_2O with respect to μ equal zero gives

$$\mu = \frac{\Sigma}{\Sigma_t}\mu_t + \left(1 - \frac{\Sigma}{\Sigma_t}\right) N_k^{-1} \sum_k z_k. \tag{5.11}$$

A similar equation is obtained for Σ,

$$\Sigma^{-1} = \Sigma_t^{-1} \frac{\sum_k (\mu - z_k)(\mu_t - z_k)}{\sum_k (\mu - z_k)^2}. \tag{5.12}$$

These solutions also converge to the correct values, but they differ from ELBO optimization. These equations show that as we approach the correct solution $\Sigma = \Sigma_t$ and $\mu = \mu_t$, the

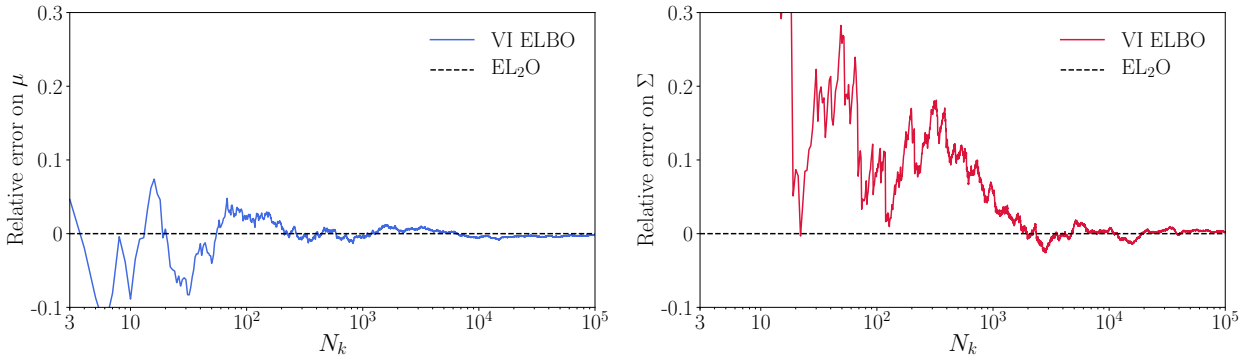


Figure 5.1: Relative errors on the mean μ and variance Σ for the Gaussian ansatz of q_θ in a setting where $p(z|\boldsymbol{x})$ is a Gaussian. We find that the stochastic VI ELBO optimization is noisy and only slowly converges to the correct answer, while EL$_2$O gives the exact solution after 3 evaluations for gradient free case and 2 evaluations for gradient case

.

sampling noise from z_k vanishes, in contrast to ELBO in equations 5.8 and 5.9. A second order Newton update can solve this system even more efficiently. Moreover, in this gradient case we only need 2 samples z_1 and z_2 to solve the problem exactly, and these two samples can be drawn anywhere if q_θ family covers the posterior. If $N_k > 2$ the problem is over-constrained: we are not gaining any additional information and EL$_2$O remains zero, as there is no sampling noise in minimizing EL$_2$O if q_θ covers the true posterior. Finally, if we start at the correct solution (for example, when MAP + Laplace is exact) we stay there, as z_k dependence vanish. If q_θ does not cover the posterior it is beneficial for the samples to be drawn as close as possible to the true posterior, so we iterate, using samples of q_θ from the previous iteration(s). A generalization of these equations to the multi-variate full rank Gaussian for q_θ is given in algorithm 1.

These properties of EL$_2$O optimization differ from the stochastic ELBO optimization and incentivize use of expressive $q_\theta(\boldsymbol{z})$, even if they contain many more parameters. We will show that a full rank Gaussian $q_\theta(\boldsymbol{z})$ converges rapidly, in contrast to ADVI. This in turn leads to a higher quality marginal posterior, which is our main goal. As we add more expressivity to $q_\theta(\boldsymbol{z})$ beyond the mean field or full rank Gaussian approximation we not only improve the posterior, we also improve the convergence because the sampling noise is reduced. If $q_\theta(\boldsymbol{z})$ covers the posterior we only need as many samples as the number of parameters θ we are optimizing for. Moreover, a better q_θ also leads to a tighter ELBO, as we show in numerical examples.

Can ELBO be modified in the similar way? If we treat the samples z_k as from a previous iteration unrelated to the current optimization of θ then

$$\text{ELBO} \approx N_k^{-1} \sum_k \left[\frac{(z_k - \mu)^2}{2\Sigma} + \frac{\ln(2\pi\Sigma)}{2} + \ln p(\boldsymbol{x}, z_k) \right]. \tag{5.13}$$

Taking a gradient with respect to μ equal zero gives $\mu = N_k^{-1} \sum_k z_k$. We see that the joint $\ln p(\boldsymbol{x}, z_k)$ does not enter in the optimization at all, so ELBO in this form cannot learn from it, in contrast to EL_2O. So reparametrization trick is necessary for ELBO optimization, while it is not required for EL_2O optimization.

The difference can be explained by the fact that the KL divergence is evaluated with stochastic integration $-\text{ELBO} = \mathbb{E}_q V(\boldsymbol{z}, \boldsymbol{x}) = \int dz q_\theta [\ln q(\boldsymbol{z}) - \ln p(\boldsymbol{z}|\boldsymbol{x})]$. The KL divergence is only positive after the integration, while the integrand can be positive or negative, but the area under the positive integrands is larger. Deterministic integration is only feasible in very low dimensions, while stochastic integration via Monte Carlo converges slowly, as $N_k^{-1/2}$. If we wanted to evaluate KL divergence with a single sample we cannot guarantee that the result is positive. In contrast, minimizing EL_2O is based on comparing gradients of $\ln q(\boldsymbol{z}_k)$ and $\ln p(\boldsymbol{z}_k, \boldsymbol{x})$ at the same sampling points $\boldsymbol{z}_k$: if the two distributions are to be equal their gradients should agree at every sampling point individually, so the L_2 norm is always positive or zero, even if evaluated on a single sample. There is no need to perform the integral to obtain a positive quantity, which is the fundamental difference between the L_2 norm of Fisher divergence and the KL divergence. Thus there is no stochastic integration noise when using Fisher divergence. It is however still a divergence, since the samples are drawn from $q(\boldsymbol{z})$.

Posterior expansion beyond the full rank Gaussian

So far we presented an algorithm with a full rank Gaussian solution as an iterative process which rapidly converges. If there is a strong variation of the Hessian elements evaluated at different sampling points then we know the posterior is not well described by a multi-variate Gaussian. In this case we may want to consider proposal functions beyond the full rank Gaussian. A full rank Gaussian is the only correlated multi-variate distribution where analytic marginalization can be done by simply inverting the Hessian matrix, which we want to keep. For this reason we will only consider one-dimensional transformations of the original variables $\boldsymbol{z}$, for which this property is still preserved. Such 1D variable transformations need to be bijective so that we can easily go from one set of the variables to the other and back (e.g. [217]). Here we will use a very simple family of models that give rise to skewness and curtosis, which are the one-dimensional versions of the gradient expansion at third and fourth order.

Specifically, we will consider nonlinear (NL) bijective transformations of the form $y_i(z_i)$ such that

$$q_\theta(\boldsymbol{z}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Pi_i |J_i|, \;\; J_i = \frac{dy_i}{dz_i}, \tag{5.14}$$

with $\boldsymbol{y}$ distributed as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Under this form the marginalization over the variables is trivial:marginalized posterior distribution of z_i is $q[z_i(y_i)] = N(\mu_i, \Sigma_{ii}) |dy_i/dz_i|$, where Σ_{ii} is the diagonal component of the covariance matrix, obtained by inverting the Hessian matrix $\boldsymbol{\Sigma}^{-1}$.

Generalizing skewness and curtosis to bijective NL transforms gives

$$y_i(z_i) = \sinh_\eta \left[\frac{\exp(\epsilon_i z_i) - 1}{\epsilon_i} \right], \tag{5.15}$$

where for $\epsilon_i = 0$ the above is just $y_i(z_i) = \sinh_\eta z_i$ [232]

$$\sinh_\eta(x) = \begin{cases} \eta^{-1}\sinh(\eta x) & (\eta > 0) \\ x & (\eta = 0) \\ \eta^{-1}\text{arcsinh}(\eta x) & (\eta < 0). \end{cases} \tag{5.16}$$

Range constraints

If a variable has a boundary the posterior distribution can abruptly changed at the boundary, which is difficult to handle with Gaussians. The most common case is that a given variable is bounded to a one sided interval, so will only consider this case, but generalization to a two-sided boundary is straight-forward.

There are two methods one can adopt, first one is a transformation to an unconstrained variable [153], and second one is a reflective boundary condition. For the latter, given a boundary $z'_i > a_i$, the method extends the range to $z'_i < a_i$ using a reflective (or mirror) boundary condition across $z'_i = a_i$, such that if $z'_i < a_i$ then $\tilde{\mathcal{L}}_p(z'_i - a_i) = \tilde{\mathcal{L}}_p(a_i - z'_i)$. This leads to a simple but non-bijective transformation: we have z'_i defined on entire range and we model it with a sum of two mirrored Gaussians or their NL generalizations. Effectively this is equivalent to an unconstrained posterior analysis, where we take the posterior at $z'_i < a_i$ and add it to $z'_i > a_i$. It solves the boundary problem of the unconstrained transformation, as the posterior at the boundary is not forced to zero, since it can be continuous and non-zero across the boundary a_i. The marginalization over this parameter remains trivial, since it is as if the parameter is not constrained at all. For the purpose of the marginalized posterior for the parameter with the range constraint itself, we must add the $z'_i < a_i$ posterior to $z'_i > a_i$ posterior. If the posterior mass is non-zero at $z'_i = a_i$ this will result in the posterior abruptly transitioning from a finite value to 0 at the boundary, as desired.

Related Work

Our proposed divergence is in the family of f-divergences, specifically, it is the Fisher divergence [103]. Other divergences have been introduced in recent literature (e.g. [73]) to counter the claimed problems of KL divergence such as its asymmetry and exclusivity of q_θ. Here we do not try to do that: our solution is the same as that of KL divergence in the limit of a large number of samples.

Stochastic VI has been explored for posteriors in several papers, including ADVI [153]. In direct comparison test presented below we find it has a slower convergence than EL_2O. Fisher divergence minimization has been proposed by [129] as a score matching statistic, but was rewritten through integration by parts into a form that does not cancel sampling variance and has similar convergence properties as stochastic VI. Reducing sampling noise has also been explored more recently in [222] in a different context and with a different approach. Quantifying the error of the VI approximation has been explored in [280].

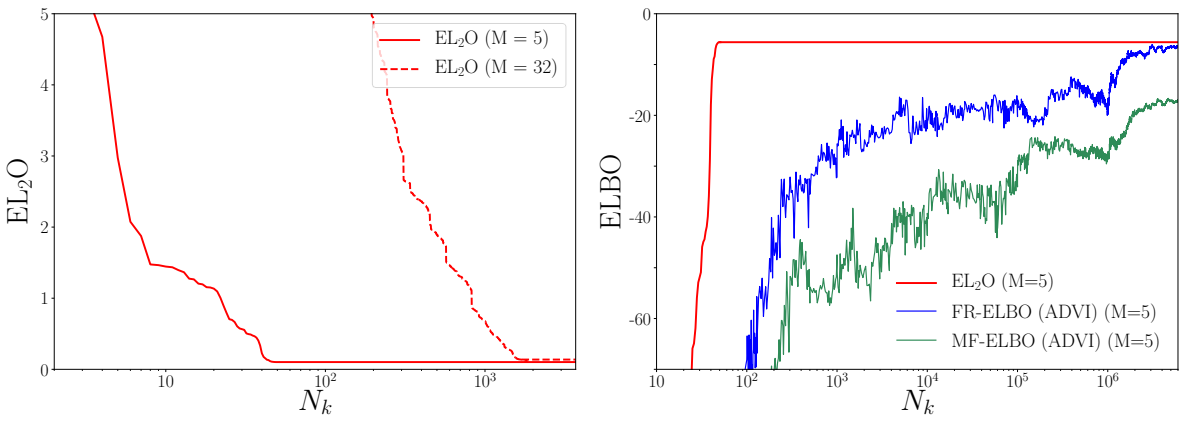


Figure 5.2: EL$_2$O and ELBO values as a function of the number of samples, for the multi-variate Gaussian example in section 5.3. *Left*: EL$_2$O converges rapidly for both $M = 5$ and $M = 32$. Most of the iterations shown are during burn-in phase. The value of EL$_2$O estimates the quality of the fit; typically values $\lesssim 0.2$ indicate that we have obtained a satisfactory posterior. *Right*: Convergence of stochastic VI ELBO optimization (ADVI) for $M = 5$ versus EL$_2$O. Full rank (FR) ADVI (blue) provides a tighter ELBO bound than mean field (MF) ADVI (green) because of correlations, but converges more slowly. We do not show $M = 32$ since ADVI does not converge.

NL transformations have been explored in terms of boundary effects in [153]. Our NL transformations correspond more explicitly to generalized skewness and curtosis parameters, and as such are useful for general description of probability distributions. We employ analytic marginals to obtain posteriors and for this reason we only employ a single layer point-wise NL transformations, instead of the more powerful normalizing flows [217]. More recently, [169] also adopt GM and NL for similar purposes, also using Hessian based second order optimization.

5.3 Numerical experiments

Here we present examples of the method, starting with two toy problems followed by a real scientific inverse problem analysis. For all numerical experiments, we use the stopping criterion such that when $\Delta\text{EL}_2\text{O} < 10^{-3}$ the algorithm terminates. Similarly for stochastic VI ELBO optimization, we use ADVI implementation in PyTorch: the algorithm terminates when the fractional change of ΔELBO, averaged over the past five iterations, is less than 10^{-3}.

High dimensional multi-variate Gaussian

The first example is a simple correlated Gaussian example of increasing dimensionality M. We model it as a full rank Gaussian without the point-wise non-Gaussian extensions.

In previous section we presented as a toy model results for $M = 1$. As we increase the dimensionality M, the number of parameters for a full rank approximation grows as M^2. As a result, the stochastic nature of ELBO optimization leads to slower and slower convergence. In our experiments we find that ELBO based ADVI converges, albeit slowly, for $M < 15$, but fails to converge above that, even if the starting point is MAP + Laplace, which is the correct solution in this example.

In contrast, EL_2O optimization converges rapidly for all the dimensions. In the left-hand side of figure 5.2, we show EL_2O optimization results for $M = 5$ and 32. Here we start at random and include the burn-in, and EL_2O requires about 50 and 2000 samples to fully converge, for $M = 5$ and 32 respectively. At each iteration we can also evaluate ELBO, and results for $M = 5$ are shown on the right-hand side of figure 5.2. In contrast, we see that ELBO optimization in the full rank form requires nearly 10^7 iterations to converge to the ELBO bound that EL_2O has reached nearly instantly! The mean field ADVI converges to an ELBO value that is considerably worse because of correlations between the variables. Moreover, full rank ADVI fails to converge for $M = 32$ even if the starting point is MAP + Laplace.

Non-Gaussian correlated 2D posterior

In this example we have a 2-dimensional problem modeled as two Gaussian distributed and correlated variables z_1 and z_2, but the second one is nonlinearly transformed using $\exp(z_2)$ mapping. We model it as a full rank $M = 2$ Gaussian plus nonlinear (NL) point-wise extension for both parameters.

This transformation is not in the family of skewness and curtosis transformations proposed in section 5.2. Here we model the posterior using ϵ and η in addition to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We determine the marginal posterior of z_2 using EL_2O with the NL transform (NL-EL_2O), and compare to MAP, mean field (MF) and full rank (FR) EL_2O (which equals MFVI and FRVI in the large sampling limit).

The results are shown in figure 5.3. Left panel shows the 2D contours, which open up towards larger values of z_2 and as a result the MAP is away from the mean. Right panel shows the resulting 1D marginalized posterior of z_2. MAP gets the peak posterior correct but not the mean. MF improves on the mean, and FR improves it further. None of these, however, get the full posterior. Only NL-EL_2O gets the full posterior in nearly perfect agreement with the correct distribution, with EL_2O value of 0.13, versus 0.5 or 0.7 for FR and MF, respectively, as shown in figure 5.4.

We have argued that sampling noise free nature of EL_2O can lead to faster convergence even if $q_\theta(\boldsymbol{z})$ has more freedom. This is seen in figure 5.4. The convergence of NL-EL_2O is faster, despite having more parameters: the convergence has been reached after 8 iterations. We started with drawing 1 sample per iteration ($N_k = 1$) and ended with $N_k = 5$ for this example, and we reused samples from previous iterations, with 25 total number of likelihood evaluations ($N_{\text{total}} = 25$). As shown in the right-hand side of figure 5.4, VI ELBO optimization

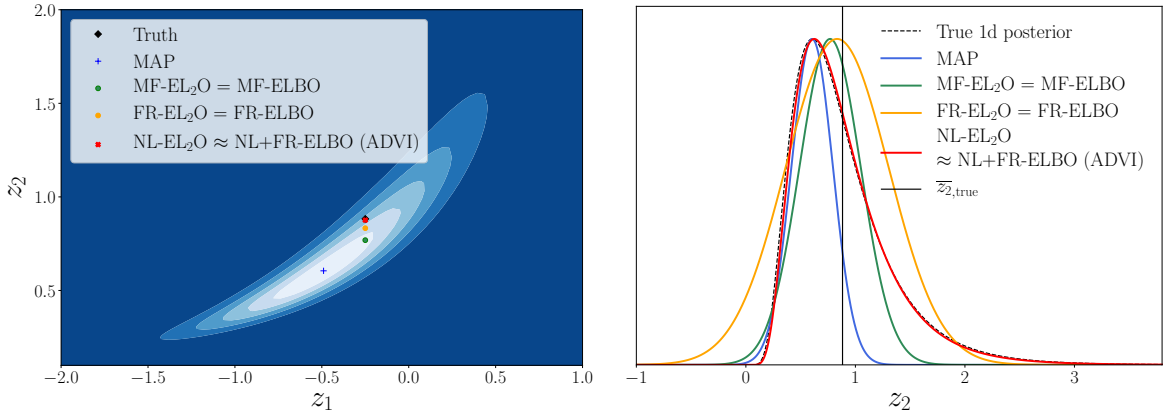


Figure 5.3: Example of a non-Gaussian correlated posterior problem in section 5.3. *Left*: The 2D posterior and the means estimated by various methods. *Right*: 1D marginalized posterior of z_2, with the black vertical line marking its true mean. MAP (blue) finds the mode, and MFVI (green) and FRVI (yellow) estimate the mean relatively well. All of them, however, fail to capture the correct shape of the posterior and its variance. EL$_2$O with the NL transform (NL-EL$_2$O, red) accurately models the posterior. Full rank VI ELBO with the NL transform provides a equivalently good fit, but its convergence is slow and noisy, as shown in figure 5.4.

All curves have been normalized to the same value at the peak to reduce their dynamical ranges.

is slow and noisy, requiring roughly fifteen times as many iterations for convergence than EL$_2$O.

An application to a scientific inverse problem: galaxy clustering analysis

In scientific inverse problems, we typically have some data and a likelihood with respect to the model parameters, but the model and the likelihood may be very expensive to evaluate. Furthermore, we may have many model parameters we need to vary, but some of the gradients may not be available. As an example of such analysis, we apply EL$_2$O to a data analysis of galaxy clustering data of the SDSS BOSS survey using the forward model of [105]. We observe about 10^6 galaxy positions (redshift and angular position), measured out to about half of the lookback time of the universe, and distributed over a quarter of the sky. The radial positions are determined by the redshift, as extracted from the galactic emission lines in spectroscopic data. Galaxy clustering is anisotropic because of the redshift space distortions (RSD), generated by the Doppler shifts, which are proportional to the galaxy velocities.

We can summarize the anisotropic clustering by measuring the multipole moments of the power spectrum $P_l(k)$.

In this specific case we are given measured summary statistics of galaxy clustering $\hat{P}_l(k)$,

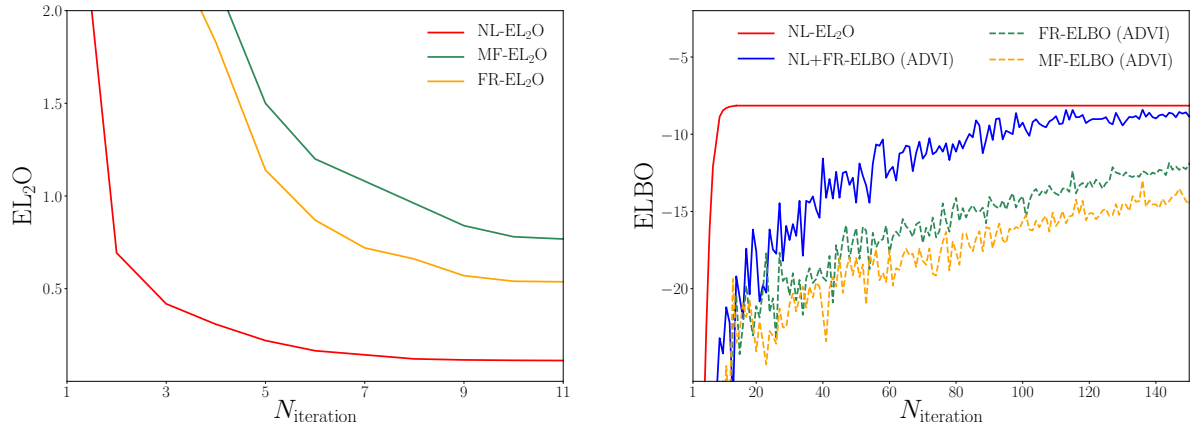


Figure 5.4: EL$_2$O and ELBO values as a function of the number of iterations, for the 2D non-Gaussian correlated posterior example in section 5.3. *Left*: Convergence is faster for NL-EL$_2$O than for MF or FR EL$_2$O despite having more parameters, a consequence of reduced sampling noise of EL$_2$O if q_θ covers the true posterior better. In this example each iteration draws 5 samples, and we average over the past samples after the burn-in. *Right*: VI ELBO requires roughly fifteen times as many iterations for convergence than EL$_2$O. Similar to figure 5.2, FR-ADVI finds a tighter ELBO bound than MF-ADVI, while NL extension of FR further improves on ELBO.

where $l = 0, 2, 4$ are the angular multipoles (Legendre transforming the angular dependence on μ) of the power spectrum and k is the wavevector amplitude, as well as their errors. Figure 5.5 presents the measured values and errors of the monopole $P_0(k)$, quadrupole $P_2(k)$, and hexadecapole $P_4(k)$ extracted from the SDSS BOSS galaxies at redshifts between 0.4 and 0.6 in the North Galactic Cap (NGC) [105].

The model of [105] predicts these power spectra $P_l(k)$. The details of the model are not relevant for the purpose of this work and we refer the reader to [105]. The galaxy power spectrum model depends on 13 physically-motivated parameters. The likelihood is formed assuming Gaussian errors, which are given as part of the galaxy clustering analysis, so it is in the form $p(\boldsymbol{x}|\boldsymbol{z}) = N(\boldsymbol{m}(\boldsymbol{z}) - \boldsymbol{x}, \boldsymbol{\sigma^2})$, where $\boldsymbol{x}$ is the data $P_l(k)$, $\boldsymbol{m}(\boldsymbol{z})$ is the model for the data, $\boldsymbol{z}$ is the 13 parameters, and $\boldsymbol{\sigma^2}$ is the noise (which we assume to be diagonal).

The model is nonlinear in its parameters and expensive to evaluate: the cost of each model evaluation is of order 10 seconds, and gradients are available for 9 out of 13 parameters. We use a finite difference version of EL$_2$O equation for the 4 parameters without the gradients and we can use algorithm 1 directly. The finite difference gradient is evaluated using $\Delta z_i = 0.1\Sigma_{ii}^{1/2}$, and we verified that the results are insensitive to this choice: the finite difference gradient does not need to be accurate compared to the true gradient, what matters is that we can learn $q_\theta(\boldsymbol{z})$ using this finite difference comparison between $q_\theta(\boldsymbol{z})$ and $p(\boldsymbol{x}, \boldsymbol{z})$. Each iteration therefore requires five model evaluations, one at the sample $\boldsymbol{z}_k$ and four at the distance Δz_i from it. We assume flat priors on all the parameters, but we have range constraints on some parameters that are positive definite, which together with the nonlinearity of the

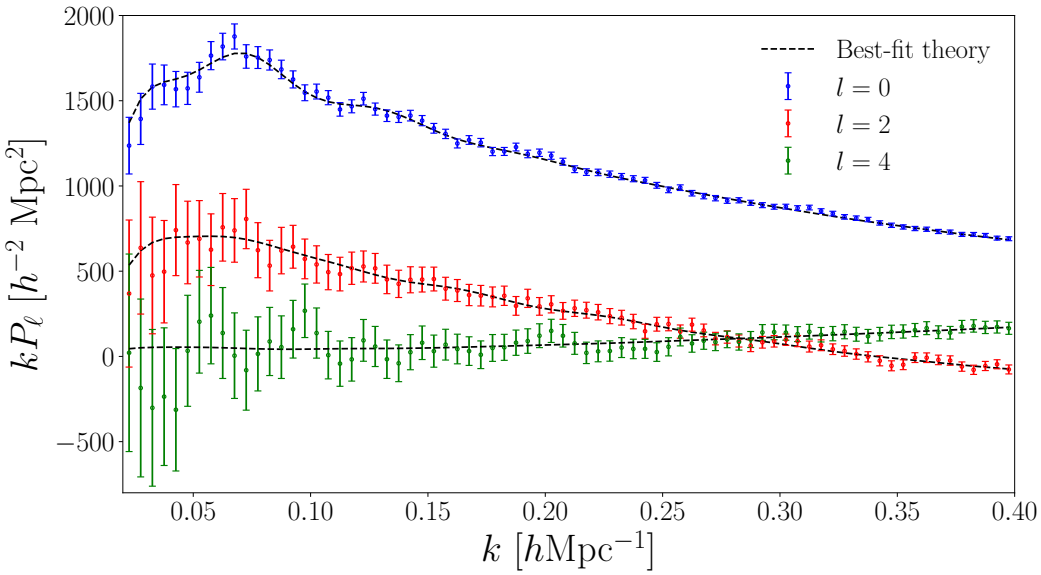


Figure 5.5: An inverse problem example: we fit the model to the galaxy power spectrum multipoles (monopole P_0, quadrupole P_2, and hexadecapole P_4). We show the best-fit (MAP) theory model (solid lines) and data measurements (points with errors). Fitting the model to data over the wavenumber range $k = 0.02 - 0.4h\text{Mpc}^{-1}$, we find a good agreement between the best fit model and the data.

forward model leads to non-Gaussian posteriors. While the model is 13-dimensional, we only care for a few parameters which represent the information about our universe, while the remaining parameters describe the galaxy clustering parameters unrelated to the cosmological information. Since the relevant cosmological parameters are correlated with all the others we need to evaluate a 12-dimensional marginal of each parameter of interest. For this problem we use full rank algorithm 1, but we go beyond the full rank by adopting 1D NL transformations to model q_θ.

We start with an optimization assuming q_θ is a delta function (MAP approach), which defines the burn-in phase of finding MAP. MAP gives the best-fit model predictions in figure 5.5 (black dotted lines). We then create the first q_θ from the Laplace approximation evaluated close to the MAP, and switch to sampling from q_θ as we approach the MAP, and gradually increase the number of samples N_k once we are past the burn-in, reusing samples from the previous iterations after the burn-in. Overall for EL_2O it took 25 iteration steps to converge to the full non-Gaussian posterior solution. Because gradients were not available for 4 parameters finite difference gradient has been used instead, leading to $N_{\text{total}} = 125$.

Results are shown in figure 5.6 and compared to MCMC and ADVI. In the top panel the parameters are $f\sigma_8$ (product of the growth rate f and the amplitude of matter fluctuations σ_8, which summarizes the cosmological information), b_1 (linear bias), σ_c (velocity dispersion for central galaxies), and $f_{1h,s_B s_B}$ (normalization parameter of the 1-halo amplitude), which

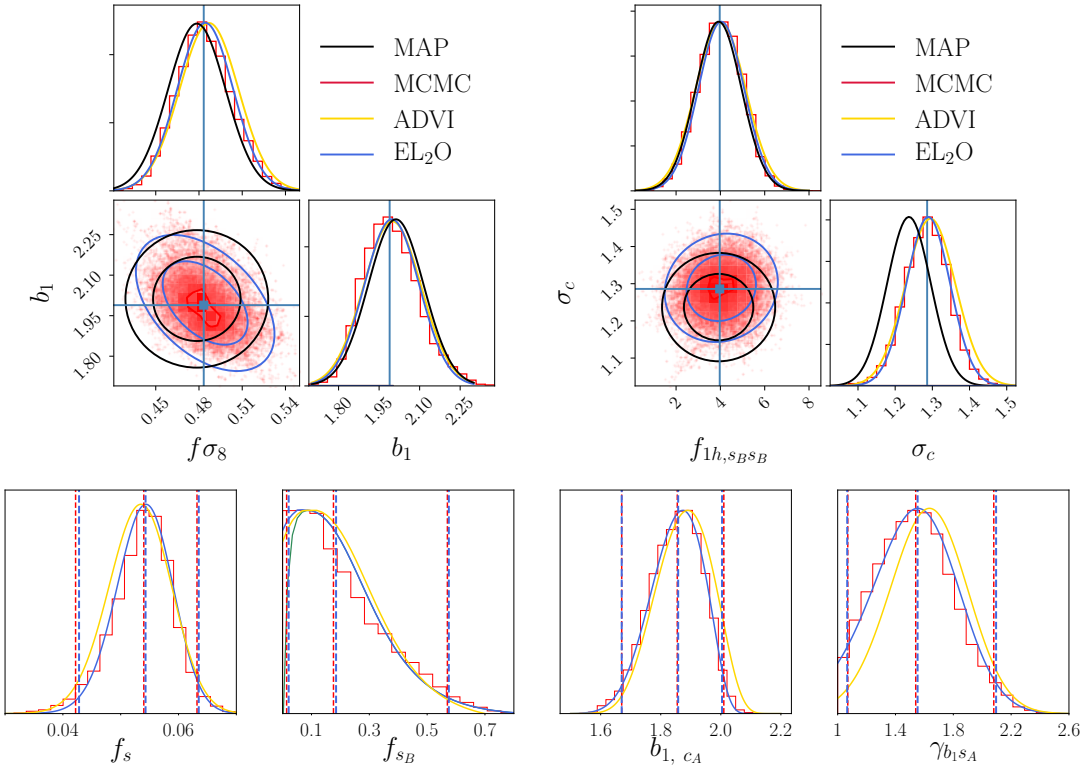


Figure 5.6: *Top*: 1D and 2D posterior distributions of four selected model parameters whose posteriors are close to Gaussian. *Top left panel*: MAP+Laplace gives inaccurate 2D posterior relative to EL_2O, even if 1D projections are accurate. *Top right panel*: MAP can be displaced in the mean, while EL_2O and ADVI results agree very well with MCMC samples. *Bottom*: 1D posteriors for parameters which are most non-Gaussian. Together with the NL transform (blue solid curves), EL_2O results closely match the MCMC posterior (red solid). Also shown are 2.5%, 50%, and 97.5% intervals (dotted lines), for MCMC and EL_2O. 125 likelihood evaluations were used for EL_2O, compared to 10^5 for MCMC, and 2.3×10^4 for ADVI. Despite taking almost 200 times more steps than EL_2O, ADVI posteriors are considerably worse. For f_{sB} parameter we have a boundary $f_{sB} > 0$, and we model it with the unconstrained transformation method (green solid) and adding the reflective boundary method to it, the latter allowing the posterior density at the boundary to be non-zero (blue solid).

are all galaxy clustering parameters unrelated to cosmological information. It is of interest to explore how the results compare to MAP+Laplace (using Hessian at MAP to determine the inverse covariance matrix) and we show these results as well in the top panel of figure 5.6. We see that MAP+Laplace can fail in the mean, or in the covariance matrix. This could be caused by the marginalization over non-Gaussian probability distributions of other parameters, or caused by small scale noise in the log posterior close to the MAP, which EL_2O improves on by averaging over several samples. The results have converged to the correct

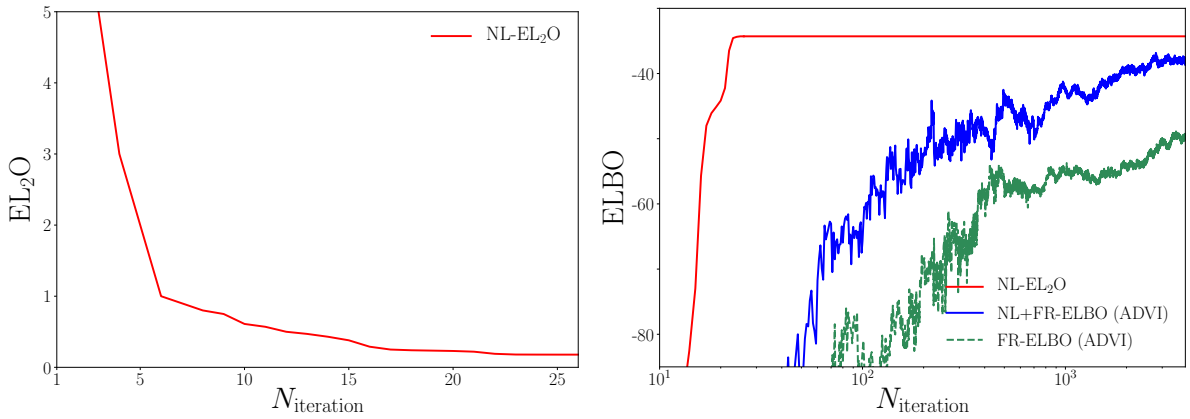


Figure 5.7: EL$_2$O and ELBO values as a function of the number of iterations, for the galaxy clustering example in section 5.3. We draw 5 samples each iteration and average over the past samples after the burn-in. EL$_2$O is 2 orders of magnitude faster than VI ELBO, providing a tighter ELBO bound. As in figure 5.2 and 5.4, NL extension of FR improves on ELBO, but ADVI has difficulty converging to the solution found by EL$_2$O.

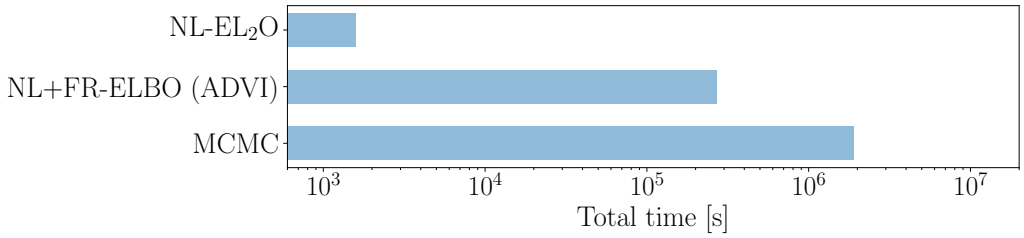


Figure 5.8: Comparison of computational cost for the galaxy clustering example in section 5.3. EL$_2$O is roughly 10^3 times faster than MCMC and 170 times faster than ADVI using the same parametrization, but note that ADVI did not converge, resulting in a worse posterior and lower ELBO, as seen in figures 5.6 and 5.7.

posterior after 25 iterations, at which point the EL$_2$O value is stable and around 0.18, which is low enough for the posteriors to be accurate. Here we compare to MCMC emcee package [87], which initially did not converge, so we restarted it at the EL$_2$O best fit parameters (results are shown with 10^5 samples after burn-in).

In the bottom panel we explore parameters that have the most non-Gaussian posteriors. In all cases the EL$_2$O posteriors agree remarkably well with MCMC. This is even the case for the parameter f_{sB}, which has a positivity constraint $f_{sB} > 0$, but is poorly constrained, with a very non-Gaussian posterior that peaks at 0. Even for this parameter the median and 2.5%, 97.5% lower and upper limits agree with MCMC. When we model this parameter with the unconstrained transformation method, we see that the probability rapidly descends to

zero at the boundary $f_{sB} > 0$.

The reflective boundary method corrects this and gives a better result at the boundary, as also shown in the same figure. In this example, with reflective boundary, we allowed the posterior to go to -0.2 on this parameter. We do not show MAP+Laplace results since they poorly match these non-Gaussian posteriors. In figure 5.8 we show the timing results, showing that EL_2O is 3 orders of magnitude faster than MCMC and 2 orders of magnitude than ADVI, but it should be noted that ADVI has not converged to EL_2O solution at this number of evaluations and its ELBO value is lower than the corresponding solution found by EL_2O, as shown in figure 5.7.

5.4 Conclusions

Computationally expensive inverse problems require minimizing the number of forward model evaluations needed for the log joint probability $\ln p(\boldsymbol{x}, \boldsymbol{z})$. In many settings, specially in scientific inverse problems, an evaluation of the forward model can be extremely costly. At the same time, scientific discovery demands high quality uncertainty quantification: we want to know for example not just one sigma error, but the full posterior, such that we can quantify the probability of excluding certain parameter values. The current gold standard for uncertainty quantification with a posterior analysis are MCMC methods, which asymptotically converge to the correct answer, but require a very large number of likelihood evaluations, often exceeding 10^5 or more. Using brute force MCMC sampling methods in these situations is often impossible.

The main motivation of this work is to develop a reliable and fast uncertainty quantification for inverse problems that improves upon MAP + Laplace and is faster than MCMC. Stochastic VI approach such as ADVI [153] is a popular alternative, but converges very slowly because of a large stochastic noise that is only slowly reduced with number of samples, and often fails to converge at all. This can be traced to the feature of KL divergence that its integrand does not have to be positive, even if the final KL divergence integral is: if sampled with a few points only there is no guarantee that the KL integral will be positive, which leads to an integration noise in the ELBO value. The optimization objective ELBO is the first cumulant of variational expansion of the normalizing constant approximated with $q_\theta(\boldsymbol{z})$ using importance sampling identity. In this chapter we propose instead to minimize the second cumulant, which we call EL_2O. It is based on comparing $\ln q_\theta(\boldsymbol{z}_k)$ and $\ln p(\boldsymbol{x}, \boldsymbol{z}_k)$ square distance at the same sampling points $\boldsymbol{z}_k$: if the two distributions are to be equal they should agree at every sampling point, up to the normalization constant. Similarly, if gradients are available, they should agree at every sampling point. Thus even a few sampling points can inform us of the quality of the approximation, in contrast to ELBO optimization. If $q_\theta(\boldsymbol{z}_k)$ covers the true posterior there is in fact no stochastic sampling noise in its optimization. This thus motivates using more expressive forms of $q_\theta(\boldsymbol{z}_k)$, and we show that this can actually lead to a faster convergence. If available we can easily add higher order derivatives to the method, or mix and match gradient and gradient free versions for different parameters θ,

which further reduces the required number of model evaluations for convergence.

In our applications we start with the MAP + Laplace approximation for the burn-in, then slowly increase the number of samples we average over by reusing samples from past iterations, typically to about 10-15 samples if gradient and Hessian are available. We first solve for the full rank Gaussian and if EL_2O residual is high generalize it to point-wise non-Gaussianity (NL). There is no guarantee that this approach will result in a good posterior, a feature in common with all other VI approaches. However, EL_2O residual can inform us when the method fails by EL_2O value being too high (higher than 0.2 in our experiments), in which case EL_2O informs us that we need to resort to MCMC. In this case it can also give a useful starting point for MCMC methods such as HMC, which requires a good initial mass matrix.

We present an application of the method to a realistic scientific application of an inverse problem in the field of cosmology, with 13 parameters and no analytic derivatives for 4 of them. We obtain good posteriors with about 25 iterations, with 5 calls each to obtain the finite difference gradients for a total of 125, of which about half were needed to find MAP. This can be compared to 10^5 likelihood calls for MCMC. This is a particularly difficult problem for MCMC, which did not converge until we restarted it at MAP solution. For many of the parameters the posteriors are very non-Gaussian, yet we found a remarkable agreement between our EL_2O posteriors and full MCMC, suggesting pointwise NL can be effective in modeling non-Gaussian distributions. Using the same parametrization $\boldsymbol{\theta}$ with ADVI required 2.3×10^4 calls, but the posteriors are worse and ELBO lower, despite about 200 times higher computational cost, a consequence of sampling noise of ELBO based optimization.

In this inverse problem example evaluation of 10^5 MCMC samples was feasible (although expensive) and we were able to compare the results of EL_2O to MCMC, but in many realistic situations MCMC would not be feasible, and methods such as EL_2O may be one of the few possible alternatives. If a high quality posteriors are required for the inverse problem solution, but we can only afford a small number of forward model evaluations, then EL_2O VI optimization provides an attractive alternative that can deliver realistic posteriors at a low computational cost.

Chapter 6

RSD measurements from BOSS galaxy power spectrum using the halo perturbation theory model

In this chapter, we present growth of structure constraints from the cosmological analysis of the power spectrum multipoles of SDSS-III BOSS DR12 galaxies. We use the galaxy power spectrum model of [105], which decomposes the galaxies into halo mass bins, each of which is modeled separately using the relations between halo biases and halo mass. The model combines Eulerian perturbation theory and halo model calibrated on N-body simulations to model the halo clustering. In this work, we also generate the covariance matrix by combining the analytic disconnected part with the empirical connected part: we smooth the connected component by selecting a few principal components and show that it achieves good agreement with the mock covariance. We find tight constraints on $f\sigma_8$: $f\sigma_8(z_{\rm eff} = 0.38) = 0.489 \pm 0.036$ and $f\sigma_8(z_{\rm eff} = 0.61) = 0.455 \pm 0.026$ at $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, in good agreement with Planck amplitude. This corresponds to $S_8 = 0.821 \pm 0.037$ or an overall amplitude error of 4%, within 0.3 sigma of Planck's $S_8 = 0.832 \pm 0.013$. We discuss the sensitivity of cosmological parameter estimation to the choice of scale cuts, covariance matrix, and the inclusion of hexadecapole $P_4(k)$. We show that with $k_{\rm max} = 0.4\ h{\rm Mpc}^{-1}$ the constraint improves considerably to an overall 2.7% amplitude error (with $S_8 = 0.786 \pm 0.021$), but there is some evidence of model misspecification on MultiDark-PATCHY mocks. Choosing $k_{\rm max}$ consistently and reliably remains the main challenge of RSD analysis methods. [1]

6.1 Introduction

Large-scale clustering of the galaxies in redshift surveys is one of the major cosmological probes which gives us insight into gravity, dark energy, and primordial non-Gaussianities. We

[1]This chapter is taken from "RSD measurements from BOSS galaxy power spectrum using the halo perturbation theory model," *Yu B., Seljak U., Li Y., Singh S.* (arXiv: 2211.16794).

can quantify this structure using the 2-point correlation function or the power spectrum. The 2-point analyses have made accurate measurements of baryon acoustic oscillations (BAO), caused by sound waves in the pre-decoupling Universe [80]. The BAO data have both isotropic and anisotropic components, and with galaxy samples from Baryon Oscillation Spectroscopic Survey (BOSS), a part of Sloan Digital Sky Survey (SDSS)-III, it provides constraints on the distance scale with a percent-level precision [15].

Galaxy clustering amplitude cannot be directly related to the dark matter amplitude due to galaxy biasing. However, we can consider another kind of anisotropy in the clustering of galaxies caused by the redshift-space distortions (RSD). It is created by peculiar velocities of galaxies, affecting the measured clustering signal in redshift space along the line-of-sight, but not transverse to it. Such distortions depend on the underlying matter density field, which are correlated with the velocity field. In the linear regime we parametrize RSD with the parameter $\beta = f/b$, where f is the linear growth rate, and b is the galaxy bias [138]. On small scales the linear theory breaks down, and non-linear distortions, such as the Finger-of-God (FoG) effect, clustering dilution in redshift space along the line of sight due to the motion of galaxies within virialized dark matter halos, need to be accounted for.

RSD has become one of the most powerful cosmological probes by measuring the growth of structure via the parameter combination $f(z)\sigma_8(z)$, thereby testing dark energy and different gravity models. [215] provides the 2.5% constraint on $f\sigma_8$ on the BOSS CMASS galaxies using a simulation-based analysis, but [215] does not employ an analytic approach to model the small-scale clustering and uses only a single simulation box. [15] presents the Data Release 12 (DR12) final consensus results on the BOSS galaxies, over the redshift range $0.2 < z < 0.75$, and provides 9.3 and 8.0% $f\sigma_8$ constraints on low-redshift ($z_{\rm eff} = 0.38$) and high-redshift galaxies ($z_{\rm eff} = 0.61$). More recently, there are BOSS DR12 measurements of the growth of structure from PT-based models [131, 68, 291, 58] and from simulation-based models [157, 148, 285, 284]. In particular, [285] proposes a Gaussian Process emulator and provides 7.4, 5.6, and 4.9% constraints on $f\sigma_8$ at $z_{\rm eff} = 0.25, 0.4$ and 0.55, and [284] develops a hybrid emulator which combines emulator with Markov chain Monte Carlo (MCMC) sampling, giving the 3.6% constraint on the BOSS CMASS galaxies. In addition, some of the recent works measure the clustering of the DR16 extended BOSS (eBOSS) samples and provides the growth of structure measurements [130, 57].

This work applies the redshift-space galaxy power spectrum model of [105] to the BOSS DR12 galaxy samples. [105] proposes an approach which combines perturbation theory(PT)-based modeling techniques and simulation-based analyses. Following the halo model formalism in [193], this model decomposes a galaxy sample into centrals and satellites and separately model the 1-halo and 2-halo—correlations of 2 galaxies in the same and different halos, respectively—contributions to the clustering of galaxies. The dark matter halo power spectrum model in redshift space is based on the distribution function approach [236, 191, 192, 272, 273, 45], and Eulerian PT and halo biasing model are used to model the underlying dark matter correlator terms [273]. Then, some of the key terms in the models are calibrated from the results of N-body simulations. [105] tests and validates this power spectrum model by performing independent tests using high-fidelity, periodic N-body simulations and realistic

BOSS CMASS mocks, showing that the recovered values of $f\sigma_8$ has only small bias.

In this work, we not only extend the work from [105] by applying its power spectrum model to the BOSS DR12 galaxies but also develop the hybrid covariance matrix, which combines the analytic disconnected part [166] and empirical connected part, including up to four principal components. Such covariance matrix can be especially useful for the analysis of next-generation redshift surveys. We also discuss how the choice of scale cuts, covariance matrix, and the inclusion of hexadecapole $P_4(k)$ affect our clustering analysis and show that removing BAO information from the multipole measurements only affects cosmological parameter estimation in a negligible way. We also compare our growth of structure constraints with other BOSS DR12 measurements in the literature, based on both PT-based models and simulation-based models.

The remainder of this chapter is organized as follows. In section 6.2, we describe the mock simulations used to validate our model and the actual galaxy sample from BOSS DR12 used for the main analysis. Section 6.3 presents the galaxy power spectrum estimator and model parameters, as well as the survey window function convolved with the model. In section 6.4, we outline analysis methods for the cosmological parameter estimation and introduce the hybrid covariance matrix, demonstrating its accuracy compared to the mock covariance matrix. In section 6.5, we validate the model performance by providing test results on the mock catalogues which mimic the BOSS DR12 target selection. In section 6.6, we discuss the main results of this chapter and conclude in section 6.7.

6.2 Data

SDSS-III BOSS DR12

In this work, we use the spectroscopic galaxy samples from SDSS-III BOSS DR12 [44, 50, 12, 70, 249, 16], selected using the imaging data from earlier SDSS-I and SDSS-II surveys. The BOSS DR12 samples are divided into the three redshift bins - z1 ($0.2 < z < 0.5$), z2 ($0.4 < z < 0.6$), and and z3 ($0.5 < z < 0.75$), following [15]. Because z2 overlaps with the other two samples and thus gives results correlated with others, we only consider z1 and z3, two non-overlapping BOSS DR12 samples. Each sample is observed in two different patches on the sky: North Galactic Cap (NGC) and South Galactic Cap (SGC). The BOSS DR12 sample, covering the redshift range $0.2 < z < 0.75$ over the area of 10,252 deg^2, contains 1,198,006 galaxies - 864,924 in NGC and 333,082 in SGC, and due to the difference in the imaging of the NGC and SGC samples, they have different characteristics, such as the bias parameter, and therefore we run an independent analysis on each sky region.

To address the problems arising from incompleteness of the BOSS survey, we apply weights to the galaxies, where the weights are given by

$$w = w_{\rm sys}(w_{\rm no-z} + w_{\rm cp} - 1), \tag{6.1}$$

where $w_{\rm sys}$ is a systematic weight. w_{no-z} and $w_{\rm cp}$ correct for missing redshifts due to failure to obtain redshift (no-z) and fiber collisions for close pairs (cp) [223].

The effective redshift for the z1 and z3 samples can be obtained as

$$z_{\rm eff} = \frac{\sum_i^{N_{\rm gal}} w_{{\rm fkp},i} \cdot w_i \cdot z_i}{\sum_i^{N_{\rm gal}} w_{{\rm fkp},i} \cdot w_i}, \tag{6.2}$$

where $w_{\rm fkp} = (1 + \bar{n}(z)P_0)^{-1}$ with $P_0 = 10^4 h^{-3}{\rm Mpc}^3$. We find that $z_{\rm eff} = 0.38$ and 0.61, respectively for z1 and z3.

MultiDark-PATCHY mock catalogues

We use the MultiDark(MD)-PATCHY mock catalogues [146] for the BOSS DR12 dataset, produced using approximate gravity solvers and galaxy biasing models calibrated to the BigMultiDark simulations, which use 3840^3 particles on a volume of $(2.5h^{-1}Mpc)^3$, and it reproduces the observed evolution of the clustering of the BOSS DR12. All quantities in these catalogues assume Planck13 cosmology: $\Omega_m = 0.307115, \Omega_L = 0.692885, \Omega_b = 0.048, \sigma_8 = 0.8288$ and $h = 0.6777$. We have 2048 mock catalogues available for both NGC and SGC hemispheres.

In section 6.4, we use Version 6C (V6C) catalogues, which is adjusted to reproduce the observed clustering measurements of the BOSS DR12, to evaluate the mock covariance matrix. Figure 6.2 shows that the power spectrum multipoles of V6C catalogues (colored dotted curves) match well with the BOSS DR12 measurements (circular data points). In section 6.3, we use Version 6S (V6S) catalogues to confirm that our theoretical model is accurate enough to obtain the cosmological parameter constraints. However, the difference between V6C and V6S catalogues are only subtle [38].

6.3 Redshift-space galaxy power spectrum

Model: perturbation theory

In this work, we use the galaxy power spectrum model of [105], which is based on perturbation theory combined with simulation-based calibration of halo model terms. We only briefly summarize the model here and refer the reader to [105] for more details.

This model follows the halo model formalism in [193], separately modeling the 1-halo and 2-halo contributions to the correlation of central and satellite galaxies. For this modeling, we decompose the galaxy sample into four sub-samples, based on whether there exists at least one other neighboring satellite in the given halo: isolated centrals without satellites ("type A" centrals), centrals with one or more satellites ("type B" centrals), isolated satellites ("type A" satellites), and non-isolated satellites ("type B" satellites). Such sub-sampling helps us separate 1-halo and 2-halo terms when modeling the total galaxy power spectrum in redshift space:

$$P_{gg}(\mathbf{k}) = (1 - f_s)^2 P_{cc}(\mathbf{k}) + 2f_s(1 - f_s)P_{cs}(\mathbf{k}) + f_s^2 P_{ss}(\mathbf{k}), \tag{6.3}$$

where P^{cc}, P^{cs}, and P^{ss} are the auto-power spectrum of centrals, the central-satellite cross-power spectrum, and the auto-power spectrum of satellites, respectively, and f_s is the satellite fraction. We also account for non-linear distortions caused by the large virial motions of satellite galaxies within their halos - known as the Finger-of-God effect. We model this effect with a Lorentzian damping factor $G(k\mu;\sigma_v) = (1 + k^2\mu^2\sigma_v^2/2)^{-2}$ applied to the redshift-space power spectrum of each sub-sample, where σ_v is the velocity dispersion of the sample. In this model, we assume a single velocity dispersion parameter for both type A and type B centrals, σ_c, and parameters for type A satellites, σ_{s_A}, and for type B satellites, σ_{s_B}. We take σ_c and σ_{s_A} as free parameters and determine σ_{s_B} from the relation between the linear bias and velocity dispersion, using the relation for the halo mass and bias and the virial theorem scaling between velocity dispersion and mass.

The resulting galaxy power spectrum model depends on the following 11 physically-motivated parameters:

$$[f(z_{\rm eff}), \sigma_8(z_{\rm eff}), b_{1,c_A}, b_{1,s_A}, b_{1,s_B}, f_s, f_{s_B}, \langle N_{>1,s}\rangle, \sigma_c, \sigma_{s_A}, f^{1h}_{s_B s_B}].$$

This includes two cosmological parameters, the growth rate f and the amplitude of matter fluctuations σ_8 evaluated at the effective redshift of the sample $z_{\rm eff}$, and linear bias parameters of type A centrals and type A and B satellites (b_{1,c_A}, b_{1,s_A}, b_{1,s_B}). We also consider the fraction of all satellites f_s, the fraction of type B satellites f_{s_B}, and the mean number of satellite galaxies in halos with more than one satellite $\langle N_{>1,s}\rangle$. The velocity dispersion parameters for some sub-samples (σ_c and σ_{s_A}) are accounted for, and we also vary the normalization nuisance parameter for the 1-halo amplitude $f^{1h}_{s_B s_B}$. Additionally, there exists the Alcock-Paczynski (AP) effect, geometric distortions of the galaxy statistics due to the mismatch between the true cosmology and the fiducial cosmology which we assumed when converting redshifts and angular positions of the observed galaxies into the three-dimensional physical positions. The AP parameter is by construction 1 at low redshifts, and in general its value is extremely insensitive to cosmological parameters at low redshifts [40, 37]. As a result we fix the scaling factors $\alpha_\parallel$ and $\alpha_\perp$ of the AP effect to their fiducial values of Planck cosmology. The primary goal of this chapter is to determine the growth rate $f\sigma_8$, and we will not attempt to separate σ_8 from f.

Measurement: power spectrum estimator

We measure the galaxy clustering signal via the multipole moments of the power spectrum $P_l(k)$, using `nbodykit`, the python software package for large-scale structure data analysis. In `nbodykit`, the FFT-based algorithm for the anisotropic power spectrum estimator, presented in [104], is implemented. This provides fast evaluation of the estimator in [279] by expanding the Legendre polynomials into spherical harmonics rather than using a Cartesian decomposition and thereby requiring only $2l + 1$ FFTs to obtain a multipole of order l.

We estimate the power spectrum multipoles as:

$$P_l(k) = \frac{2l+1}{A}\int \frac{d\Omega_k}{4\pi} F_0(\mathbf{k})F_l(-\mathbf{k}), \tag{6.4}$$

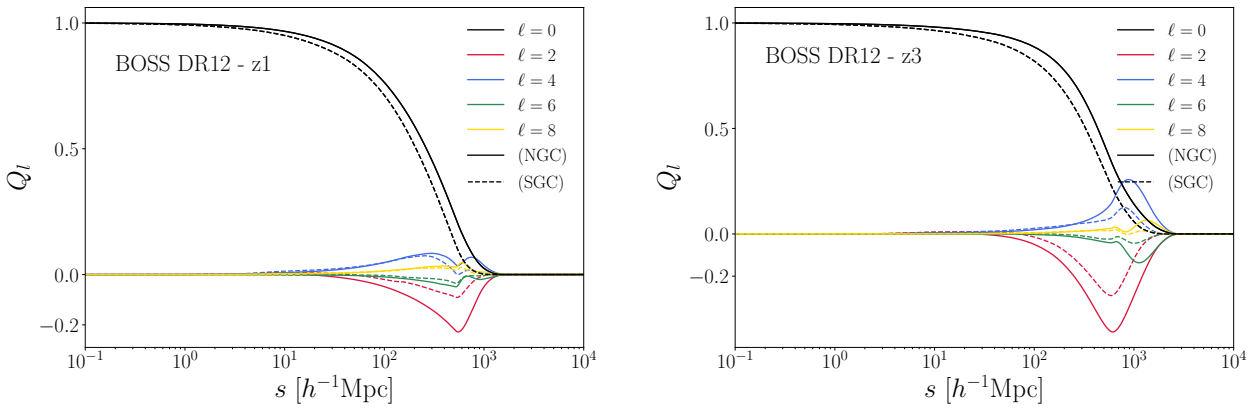


Figure 6.1: The window function multipoles in configuration space for the BOSS DR12 z1 (*Left*) and z3 (*Right*) samples. We include up to $Q_8(s)$ because the contribution of $l = 10$ or higher is negligible for the window convolution.

where Ω_k is the solid angle in Fourier space, and $\mathcal{L}_l$ is the Legendre polynomial. A is the normalization defined as $A \equiv \int d\mathbf{r}[n'_{\rm gal}(\mathbf{r})w_{\rm fkp}(\mathbf{r})]^2$, where $n'_{\rm gal}$ is the weighted galaxy number density field, and $w_{\rm fkp}$ is the FKP weight.

The weighted galaxy density field $F(\mathbf{r})$ is given by

$$F(\mathbf{r}) = \frac{w_{\rm fkp}(\mathbf{r})}{A^{1/2}}[n'_{\rm gal}(\mathbf{r}) - \alpha' n'_{\rm ran}(\mathbf{r})], \tag{6.5}$$

where $n'_{\rm gal}$ and $n'_{\rm ran}$ are the number density field for the galaxy and randoms catalogues respectively, with the factor α' normalizing $n'_{\rm ran}$ to $n'_{\rm gal}$, and

$$\begin{aligned} F_l(\mathbf{k}) &= \int d\mathbf{r} F(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{L}_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}) \\ &= \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{lm}(\hat{\mathbf{k}}) \int d\mathbf{r} F(\mathbf{r}) Y^*_{lm}(\hat{\mathbf{r}}) e^{i\mathbf{k}\cdot\mathbf{r}}. \end{aligned} \tag{6.6}$$

We compute each summation over m using a FFT, hence a total of $2l + 1$ FFTs.

However, [39] replaces the traditional definition of the normalization term A with the value enforcing the following condition on the window function multipole of order $l = 0$, $Q_0(s \to 0) = 1$, to ensure that the power spectrum and window function are normalized in a consistent way. Table 1 in [39] shows that such correction of the normalization term results in increasing the BOSS DR12 galaxy power spectrum multipole amplitudes by roughly 10%. Following [39], we corrected all galaxy power spectrum measurements presented in this work.

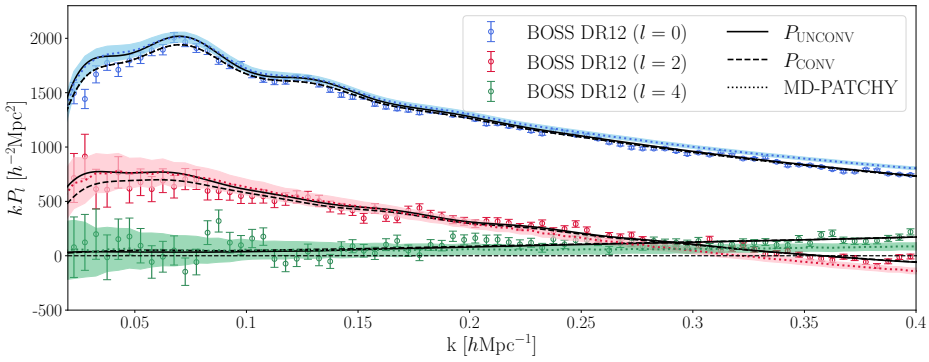


Figure 6.2: The window function effects on the power spectrum multipoles (filled circles with error bars) for the BOSS DR12 z3 NGC sample. The solid and dashed curves correspond to the unconvolved and convolved multipoles, respectively. We also show the mean of 1000 MD-PATCHY V6C mock catalogues (colored dotted curves), and shaded areas indicate 1σ deviations of 1000 mocks. Dotted curves match well with the data points, suggesting that V6C catalogues reproduce the clustering of the observed data (which makes them suitable for the covariance matrix estimation). For our main analysis, we choose the minimum wavenumber of $k_{\rm min} = 0.02h{\rm Mpc}^{-1}$ to minimize any large-scale effects of the window function.

Survey window function

We account for the window function effects by convolving the theoretical model in section 6.3 with the survey window function, which corresponds to the Fourier transform of the survey volume. We denote the resulting quantity as the "convolved" power spectrum.

We follow the method presented in [277] to compute the convolved power spectrum multipoles. First, we use the pair counting algorithm in `nbodykit` [106], which employs the `Corrfunc` package [245], to obtain the pair counts of the random catalogue. In figure 6.1, we present the resulting window function multipoles, $Q_l(s) \propto \int_{-1}^{1} d\mu RR(s,\mu)\mathcal{L}_l(\mu) \approx \sum_i RR(s_i,\mu_i)\mathcal{L}_l(\mu_i)$, of the BOSS DR12 galaxy samples in configuration space, for a set of separations s. The z1 and z3 window function multipoles vanish on scales larger than ≈ 2000 and 3000 h^{-1}Mpc, respectively, and these correspond to the largest scales in the volume of the BOSS galaxy samples. In this work, we ignore $l = 10$ or higher, as its contribution to the convolution is negligible.

Next, we convolve the correlation function multipoles $\xi_l(s)$ obtained from the theoretical

model with the survey window function to get the convolved multipoles $\hat{\xi}(s)$:

$$\begin{aligned}
\hat{\xi}_0 =&\xi_0 Q_0 + \frac{1}{5}\xi_2 Q_2 + \frac{1}{9}\xi_4 Q_4 + ... \\
\hat{\xi}_2 =&\xi_0 Q_2 + \xi_2\Big[Q_0 + \frac{2}{7}Q_2 + \frac{2}{7}Q_4\Big] \\
&+ \xi_4\Big[\frac{2}{7}Q_2 + \frac{100}{693}Q_4 + \frac{25}{143}Q_6\Big] + ... \\
\hat{\xi}_4 =&\xi_0 Q_4 + \xi_2\Big[\frac{18}{35}Q_2 + \frac{20}{77}Q_4 + \frac{45}{143}Q_6\Big] \\
&+ \xi_4\Big[Q_0 + \frac{20}{77}Q_2 + \frac{162}{1001}Q_4 + \frac{20}{143}Q_6 + \frac{490}{2431}Q_8\Big] + ...
\end{aligned} \tag{6.7}$$

Figure 6.2 demonstrates that the effects of the window function is mostly on large scales. In this work, we choose the minimum wavenumber as $k_{\rm min} = 0.02h{\rm Mpc}^{-1}$ to minimize any large-scale effects of the window function.

6.4 Analysis methods

Covariance matrices

Fitting the theoretical model to the measured data requires a covariance matrix estimate, and in this work we not only consider the covariance matrix from mock catalogues (section 6.4) but also the hybrid covariance matrix which combines the analytic disconnected component (6.4) and smoothed connected component (6.4).

Mock covariance matrix

We compute the covariance matrix from 1000 realizations of MD-PATCHY mock catalogues (section 6.2):

$${\rm Cov}\Big[P_l(k_i), P_{l'}(k_j)\Big] = \frac{1}{N-1}\sum_{\alpha=1}^{N}\Big[P_{l,\alpha}(k_i) - \overline{P}_l(k_i)\Big]\cdot\Big[P_{l',\alpha}(k_j) - \overline{P}_{l'}(k_j)\Big], \tag{6.8}$$

where N is the number of mocks, and $\overline{P}_l(k) = \frac{1}{N}\sum_{\alpha=1}^{N} P_{l,\alpha}(k)$ is the mean power spectrum. Hence, we obtain the covariances between multipoles (for $l = 0, 2, 4$) along with their uncertainties. For all multipoles, the fitting range is $0.02 < k < 0.4$ $h{\rm Mpc}^{-1}$ and $\Delta k = 0.005$ $h{\rm Mpc}^{-1}$ (corresponding to $N_{\rm bin} = 76$ for each l). We also apply the Hartlap correction [112] to get an unbiased estimate of the true inverse covariance matrix.

Analytic disconnected covariance matrix

Following [166], we compute the analytic "disconnected" covariance matrix (more conventionally, "Gaussian" covariance matrix) which takes into account the window effect. This

analytic method is free of sampling noise and therefore avoids numerical issues of the mock covariance matrix.

Assuming the flat sky approximation, we write the ensemble average of the estimated power spectrum as

$$\left\langle \hat{P}(\boldsymbol{k}) \right\rangle = \frac{1}{\mathcal{W}_0} \int_{\boldsymbol{q}} P(\boldsymbol{k}-\boldsymbol{q}) |W(\boldsymbol{q})|^2 \simeq \frac{P(\boldsymbol{k})}{\mathcal{W}_0} \int_{\boldsymbol{q}} |W(\boldsymbol{q})|^2 \text{ (for } k \gg q) = P(\boldsymbol{k}), \tag{6.9}$$

where $W(\boldsymbol{x}) \equiv \bar{n}_{\rm gal}(\boldsymbol{x}) w(\boldsymbol{x})$ denotes the windows on the fields, with $\bar{n}_{\rm gal}(\boldsymbol{x}) = \left\langle n_{\rm gal}(\boldsymbol{x}) \right\rangle$ and the weight $w(\boldsymbol{x})$ which minimizes systematic effects. This shows that we get the ensemble average of the estimator $\hat{P}$ by convolving the true power spectrum P with a window, and $\hat{P}$ is an unbiased estimate of P for scales much smaller than the window. We then split the covariance $\mathrm{Cov}\left[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\right] = \left\langle \hat{P}(\boldsymbol{k}) \hat{P}(\boldsymbol{k}') \right\rangle - \left\langle \hat{P}(\boldsymbol{k}) \right\rangle \left\langle \hat{P}(\boldsymbol{k}') \right\rangle$ into the disconnected and connected pieces:

$$\mathrm{Cov}\left[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\right] = \mathrm{Cov}^{\rm disc}\left[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\right] + \mathrm{Cov}^{\rm conn}\left[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\right]. \tag{6.10}$$

The disconnected covariance component $\mathrm{Cov}^{\rm disc}$ involves quadratic combinations of the following window factors which modulate Gaussian and Poisson parts: $\mathcal{W}(\boldsymbol{q}) = \int_{\boldsymbol{x}} \mathcal{W}(\boldsymbol{x}) e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \equiv \int_{\boldsymbol{x}} W(\boldsymbol{x})^2 e^{-i\boldsymbol{q}\cdot\boldsymbol{x}}$ and $\mathcal{S}(\boldsymbol{q}) = \int_{\boldsymbol{x}} \mathcal{S}(\boldsymbol{x}) e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \equiv (1+\alpha) \int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x}) w(\boldsymbol{x})^2 e^{-i\boldsymbol{q}\cdot\boldsymbol{x}}$. We further define the factor $\mathcal{Q}_{\mathcal{W}}, \mathcal{Q}_{\mathcal{S}}$ and $\mathcal{Q}_{\times}$ as the auto- and cross-correlations of $\mathcal{W}$ and $\mathcal{S}$,

$$\begin{aligned}
\mathcal{Q}_{\mathcal{W}}(\boldsymbol{q}) &\equiv \mathcal{W}(\boldsymbol{q}) \mathcal{W}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{W}}(\boldsymbol{s}) e^{-i\boldsymbol{q}\cdot\boldsymbol{s}}, \\
\mathcal{Q}_{\mathcal{S}}(\boldsymbol{q}) &\equiv \mathcal{S}(\boldsymbol{q}) \mathcal{S}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{S}}(\boldsymbol{s}) e^{-i\boldsymbol{q}\cdot\boldsymbol{s}}, \\
\mathcal{Q}_{\times}(\boldsymbol{q}) &\equiv \mathcal{W}(\boldsymbol{q}) \mathcal{S}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\times}(\boldsymbol{s}) e^{-i\boldsymbol{q}\cdot\boldsymbol{s}},
\end{aligned} \tag{6.11}$$

and write the disconnected covariance as

$$\begin{aligned}
\mathrm{Cov}^{\rm disc}\left[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\right] \approx \frac{1}{\mathcal{W}_0^2} \Big\{ & P(\boldsymbol{k}) P(\boldsymbol{k}') \mathcal{Q}_{\mathcal{W}}(\boldsymbol{k}-\boldsymbol{k}') \\
& + \left[P(\boldsymbol{k}) + P(\boldsymbol{k}')\right] \Re\left[\mathcal{Q}_{\times}(\boldsymbol{k}-\boldsymbol{k}')\right] + \mathcal{Q}_{\mathcal{S}}(\boldsymbol{k}-\boldsymbol{k}') \Big\} + (\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}').
\end{aligned} \tag{6.12}$$

We refer the reader to [166] for a complete and detailed derivation.

Figure 6.3 presents the 3×3 blocks of correlation matrices from mock catalogues (left panel) and from the analytic method described in this section (right panel). Correlation coefficients are defined as

$$\mathrm{Corr}(O, O') = \frac{\mathrm{Cov}(O, O')}{\sqrt{\mathrm{Cov}(O, O)\, \mathrm{Cov}(O', O')}}, \tag{6.13}$$

where $O \in \{P_0(k), P_2(k), P_4(k)\}$ and $O' \in \{P_0(k'), P_2(k'), P_4(k')\}$. [166] demonstrates that the analytic Gaussian covariance matrix is in excellent agreement with the mock covariance matrix, and this method is free of the sampling noise with much smaller computational cost, compared to the mocks.

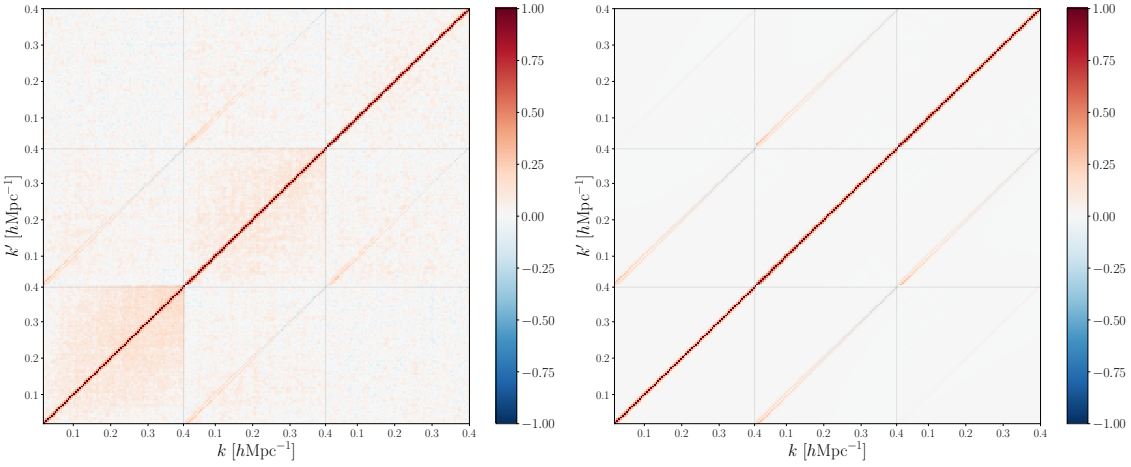


Figure 6.3: 3×3 blocks of correlation matrices of the power spectrum multipoles, which visualize the auto- and cross-correlations of P_0, P_2 and P_4. *Left*: Mock correlation matrix from 1000 MD-PATCHY z3 NGC mock simulations (section 6.4). *Right*: Analytic disconnected correlation matrix for the z3 NGC sample (section 6.4).

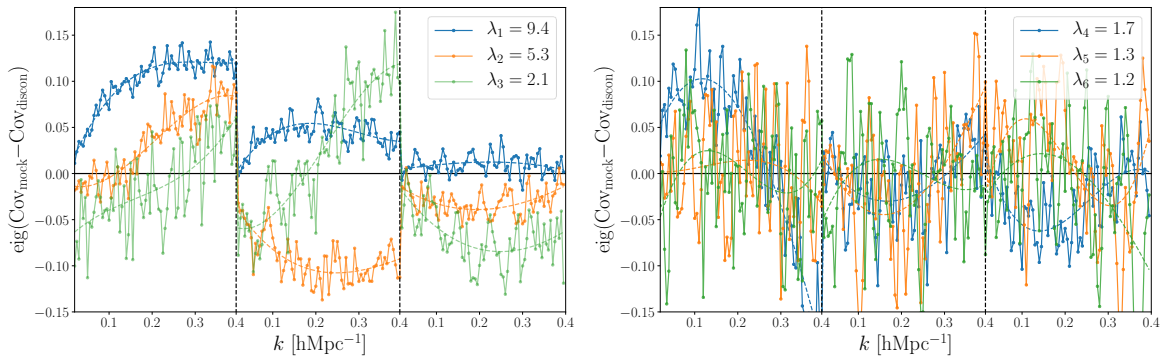


Figure 6.4: The first six principal components of the (3×3) blocks of P_l from 1000 MD-PATCHY z3 NGC mock simulations, where $l = 0, 2, 4$. We only include up to four principal components for a low-rank approximation, as the components beyond the fourth are noisy and do not contain much broadband correlations. λ_i denotes the eigenvalue of the i-th eigenvector.

Modeling the connected pieces using PCA

The remaining connected pieces of the covariance matrix, which include Poisson, non-Gaussian (trispectrum) and Gaussian (power spectrum) components, is more difficult to model analytically than the disconnected piece. However, an eigenmode decomposition of the connected components shows that it is a low-rank matrix [111, 185, 274]. In this work, we first obtain the connected parts from the MD-PATCHY mock simulations by subtracting the disconnected parts from the mock covariance matrix and show that this empirical connected

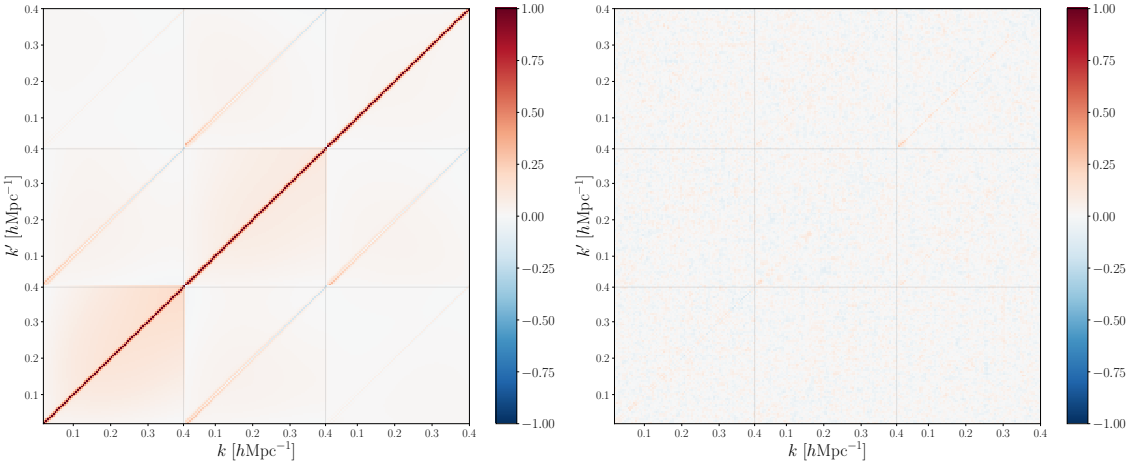


Figure 6.5: *Left*: we combine the analytic Gaussian covariance with the smoothed connected pieces obtained from low-rank components with principal component analysis. *Right*: The difference between the mock and the hybrid covariance which includes the analytic disconnected and smoothed connected covariances, normalized by the diagonal of the latter. Because the connected part is smooth and has a low-rank approximation, we find that only the first four principal components are needed to obtain a smooth estimate of the connected part, and the difference between the mock and its smooth estimate is clean.

piece is indeed a low-rank component with principal component analysis (PCA).

In Figure 6.4, we perform PCA on the (3×3) blocks of P_l, where $l = 0, 2, 4$, and show its first six principal components. The principal components beyond the fourth component are noisy and do not carry much broadband correlations, and therefore we only include up to four principal components; therefore, the connected components can be well approximated by a low-rank eigen-decomposition. The resulting smoothed connected covariance, combined with the analytic Gaussian covariance, is shown on the left panel of Figure 6.5, and the right panel shows that it achieves good agreement with the mocks.

Parameter estimation techniques

In this work, we use the following methods to obtain the parameter posterior distribution: 1) Maximum a posteriori (MAP) estimation and Laplace approximation, using the hessian of the log posterior to obtain the inverse covariance matrix of the model parameters and 2) MCMC sampling of the likelihood, assuming the hybrid covariance matrix with the smoothed connected parts. In section 6.5, we find the best-fitting model parameters for each of the 1000 mock catalogues from MAP estimation and present the 1D histograms and 2D correlations of the cosmological parameters of our interest. Section 6.6 summarizes the main results of this work: BOSS DR12 RSD measurements of the growth of structure, and we obtain the parameter posteriors using the Python module `emcee` [86]. [237] presents an optimization-

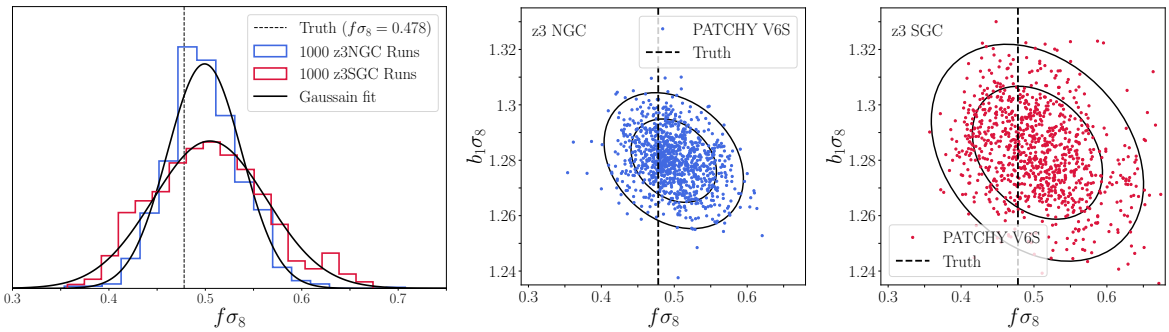


Figure 6.6: *Left*: Maximum a posteriori (MAP) results for 1000 MD-PATCHY z3 NGC (blue) and SGC (red) mock catalogues. We fit the monopole, quadrupole, and hexadecapole over the wavenumber range $0.02 < k < 0.2\ h\mathrm{Mpc}^{-1}$. 1D distribution of our $f\sigma_8$ fit results gives $f\sigma_8 = 0.499 \pm 0.038$ and $f\sigma_8 = 0.502 \pm 0.058$, for z3 NGC and SGC respectively. The true cosmology indicates $f\sigma_8 = 0.478$, which is 0.4-0.5σ away from the mean of our recovered values. *Middle, Right*: 2D correlations of $f\sigma_8$ and $b_1\sigma_8$ for 1000 MD-PATCHY z3 NGC (blue) and SGC (red) mock catalogues. Vertical dashed line indicates the true cosmology, and solid contours show 1σ and 2σ confidence regions.

based posterior inference method called $\mathrm{EL_2O}$ and shows that the posterior distribution from $\mathrm{EL_2O}$ agrees with the MCMC results. Particularly, section 4.4 in [237] discusses how $\mathrm{EL_2O}$ can be effective in galaxy clustering analyses. We refer the reader to [237] for more detailed analysis about the comparison between $\mathrm{EL_2O}$ and MCMC methods.

6.5 Model Performance

Tests on the MultiDark-PATCHY mocks

In [105], the accuracy and precision of the power spectrum model in section 6.3 are extensively assessed by performing independent tests using several sets of mocks based on high-fidelity, periodic N-body simulations and realistic BOSS DR12 CMASS mocks. To further confirm that this model is unbiased and accurate enough to provide cosmological parameter constraints of the BOSS DR12 sample, we fit our model to 1000 MD-PATCHY mock catalogues and verify that we can retrieve the true cosmology, provided by the BigMultiDark simulation.

Applying the analysis pipeline in section 6.4 to the mock catalogues, we obtain the best-fitting parameters for each of the 1000 catalogues by first measuring the power spectrum multipoles ($l = 0, 2, 4$) for each of the catalogues and obtaining the MAP estimate using the L-BFGS algorithm. Figure 6.6 presents the MAP results for 1000 MD-PATCHY z3 NGC (blue) and SGC (red) mock catalogues, and the black dotted vertical line indicates the expected parameter value from the true cosmology of the MD-PATCHY simulations.

If we include up to $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, we find that $f\sigma_8 = 0.499 \pm 0.038$ for z3 NGC and $f\sigma_8 = 0.502 \pm 0.058$ for z3 SGC, so we recover the true cosmology ($f\sigma_8 = 0.478$) within 1σ with only modest mean biases of $\Delta f\sigma_8$ of 0.5σ.

Choice of $k_{\rm max}$

We can further investigate whether the 0.4-0.5 σ bias we observe in Figure 6.6 is caused by the priors or by model misspecification, by comparing it to the analysis where we treat all 1000 mocks as a single dataset. In this case the MAP will be dominated by the likelihood and priors can be ignored. We find $f\sigma_8 = 0.478$ for z1 NGC and $f\sigma_8 = 0.490$ for z3 NGC for $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, compared to the truth ($f\sigma_8 = 0.484$ for z1 and 0.478 for z3), both with very small error of order 0.001. In both cases the mean has moved closer to the true value, suggesting the prior is driving the MAP away from the true value, even if there is also some small model misspecification for z3 NGC. If we repeat the analysis for $k_{\rm max} = 0.4\ h{\rm Mpc}^{-1}$ we find $f\sigma_8 = 0.498$ for z1 NGC and $f\sigma_8 = 0.497$ for z3 NGC. Now the bias is larger, and suggests more significant model misspecification.

To investigate this further, Figure 6.7 shows the power spectrum multipole measurements of MD-PATCHY z1 and z3 NGC mock catalogues, along with the best-fit theory lines. Measurements are averaged over 1000 realizations, and the errors are therefore significantly smaller than those of the BOSS survey. Solid and dashed curves indicate convolved and unconvolved best-fit theory curves, respectively. Comparing the upper panel figures (assuming $k_{\rm max} = 0.4\ h{\rm Mpc}^{-1}$) with the lower panel figures ($k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$), we find that extending the model to a higher $k_{\rm max}$ limit makes the model fit noticeably worse at low k: both monopole and quadrupole fits with lower $k_{\rm max}$ have better fits for z1 NGC, and similarly monopole fit has with lower $k_{\rm max}$ has a better fit for z3 NGC.

MD-PATCHY mocks are not based on a real N-body simulation, and it is unclear if the galaxy catalogs and the resulting power spectra can be realized in an actual universe. For this reason such comparisons against PATCHY have not been implemented elsewhere, and it is unclear whether we should be concerned given the model is good against real simulations. On the other hand, using 1000 mocks enables one to separate statistical fluctuations from systematics extremely well. In other tests based on one or a few simulation volumes the deviations of recovered cosmological parameters from the truth were within one statistical deviation, in which case it is unclear whether this is a purely statistical effect that can be ignored, or it is a sign of model misspecification, and one must correct for it. As we are unable to answer whether MD-PATCHY power spectra can represent an actual galaxy realization in a real universe, we present both results. However, more caution should be taken when extending to a higher $k_{\rm max}$, and we argue that a consistent, reliable choice of $k_{\rm max}$ is one of the major unresolved questions of the RSD analyses. It becomes increasingly difficult to obtain unbiased estimates as we push to higher $k_{\rm max}$, simply because the fits are dominated by the smallest errors which are close to $k_{\rm max}$, but cosmology information is entirely extracted by the low k asymptote of the fitted model: even a slight model misspecification at high k can lead to a biased answer at low k. Moreover, since we have to fit more parameters to high

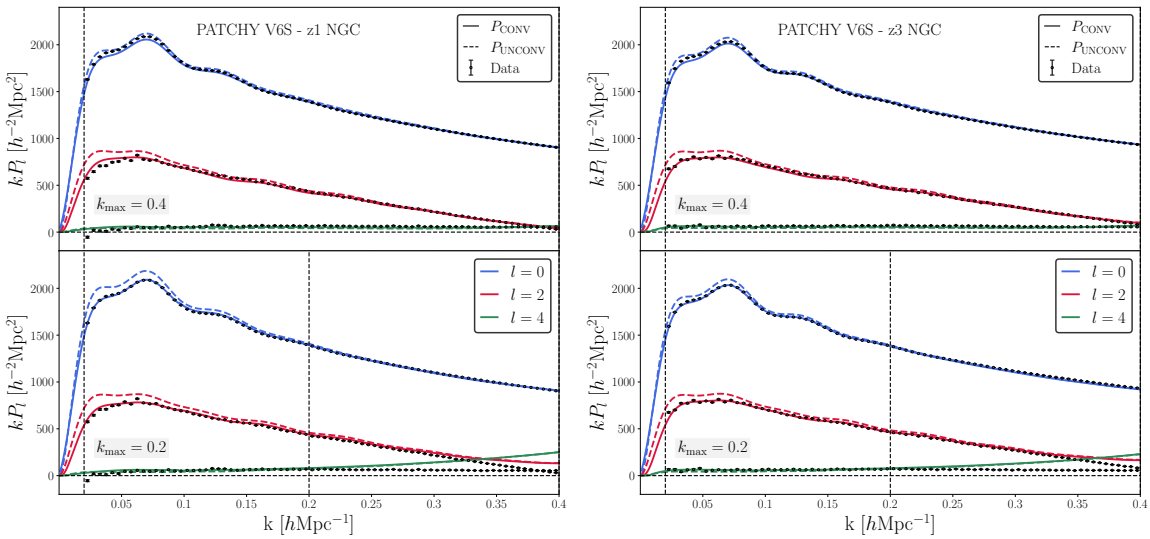


Figure 6.7: Power spectrum multipole measurements (circular points) of MD-PATCHY z1 NGC and z3 NGC mock catalogues and the best-fit theory models. We take 1000 realizations, and the errors are therefore reduced by $\sqrt{1000}$. Solid and dashed curves indicate convolved and unconvolved best-fit theory curves, respectively. The upper panel figures assume $k_{\rm max} = 0.4$ hMpc^{-1}, while the lower panel assumes a lower $k_{\rm max}$ (0.2 hMpc^{-1} with vertical lines showing the corresponding wavenumber limits to the fits. We find that extending the model to a higher $k_{\rm max} = 0.4$ hMpc^{-1} limit makes the model fit worse at low k, and as a result an incorrect cosmology may be recovered. We do not observe this issue for $k_{\rm max} = 0.2$ hMpc^{-1}.

k the choice of their prior distribution also affects the fits: even a seemingly innocent flat prior choices can project onto a significant bias in the cosmological parameters. Different works choose different values, which may partially explain why results from different studies are discrepant in terms of their $f\sigma_8$ measurements.

6.6 BOSS DR12 RSD measurements

In this section, we present the measurements of the BOSS DR12 galaxy power spectrum multipoles in Fourier space. In figure 6.8, we show the measured monopole $P_0(k)$, quadrupole $P_2(k)$, and hexadecapole $P_4(k)$ of z1 and z3 galaxies in both NGC and SGC patches, using the FFT-based galaxy power spectrum estimator described in section 6.3. We then fit the RSD model presented in section 6.3 to the measured multipoles and find that the power spectrum multipoles are accurately modeled, in agreement with [105]. In our fits, we set the minimum wavenumber $k_{\rm min}$ to 0.02 hMpc^{-1} for all samples, in order to minimize any large-scale effects of the window function. As described in section 6.3, we fix the AP distortion parameters to their fiducial values and constrain 11 model parameters $(f(z_{\rm eff}), \sigma_8(z_{\rm eff}), b_{1,c_A}, b_{1,s_A}, b_{1,s_B}, f_s, f_{s_B}, \langle N_{>1,s}\rangle, \sigma_c, \sigma_{s_A}, f^{1h}_{s_B s_B})$, of which two are primarily of

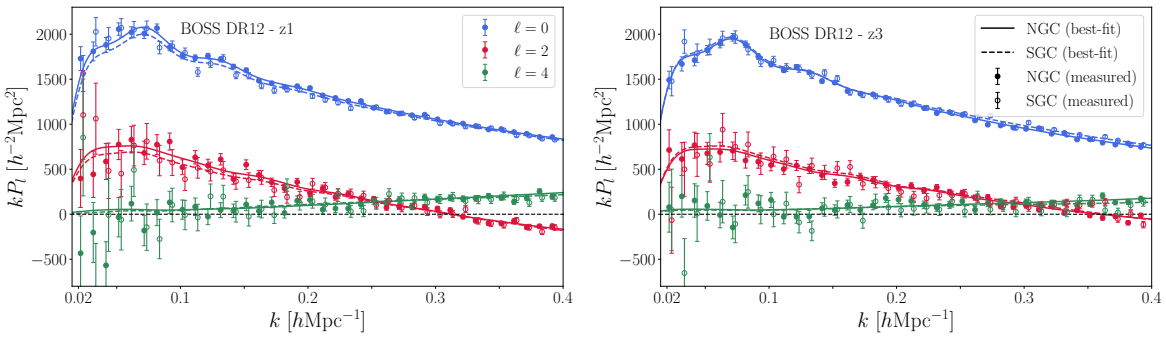
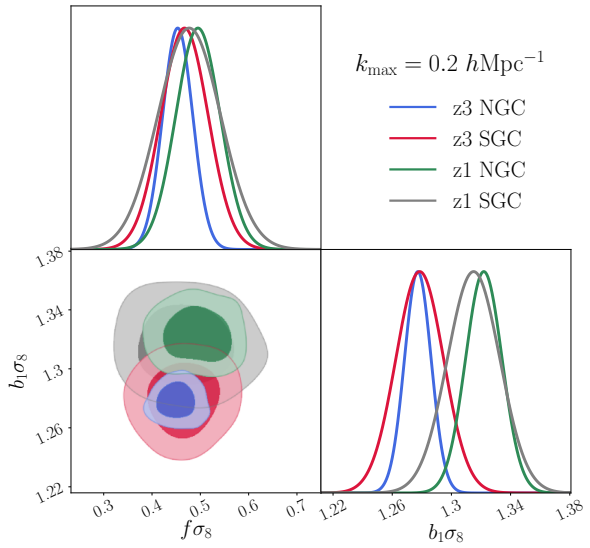


Figure 6.8: The measured galaxy power spectrum multipoles in Fourier space (circular points with error bars) and the best-fit theory curves (solid lines) for BOSS DR12 z1 (*Left* panel) and z3 (*Right* panel) samples. (We only show every other data points for simplicity.) We can fit the model to the monopole (blue), quadrupole (red), and hexadecapole (green), over the wavenumber range $0.02 - 0.4h\text{Mpc}^{-1}$, with $\Delta k = 0.005$, but the choice of $k_{\rm max}$ may affect the cosmological analysis moderately, as discussed in section 6.5. Following [39], we use a consistent definition of the normalization term for both power spectrum and window function.

our interests: the growth rate f and the amplitude of matter fluctuations σ_8. The Planck satellite provides a strong prior on f with the tight constraint of the matter density Ω_m [11], and we therefore put the Planck 2018 prior on f in all subsequent analyses.

Figure 6.9, along with the table of $f\sigma_8$ constraints with varying $k_{\rm max}$ cuts, summarizes the main results of our analysis. Fitting our RSD model to the measured BOSS DR12 multipoles (monopole, quadrupole, and hexadecapole) and marginalizing over all nuisance parameters discussed in section 6.3, we obtain the following growth of structure constraints: $f\sigma_8(z_{\rm eff} = 0.38) = 0.489 \pm 0.036$ for z1 sample and $f\sigma_8(z_{\rm eff} = 0.61) = 0.455 \pm 0.026$ for z3 sample, with $k_{\rm max} = 0.2\ h\text{Mpc}^{-1}$. In the table, we also provide the constraints as a function of $k_{\rm max}$: $f\sigma_8(z_{\rm eff} = 0.38) = 0.488 \pm 0.049$ for z1 and $f\sigma_8(z_{\rm eff} = 0.61) = 0.456 \pm 0.040$ for z3 with $k_{\rm max} = 0.1\ h\text{Mpc}^{-1}$, while $f\sigma_8(z_{\rm eff} = 0.38) = 0.445 \pm 0.022$ for z1 and $f\sigma_8(z_{\rm eff} = 0.61) = 0.447 \pm 0.014$ for z3 with $k_{\rm max} = 0.4\ h\text{Mpc}^{-1}$. For all results, we assume the full covariance with the smoothed connected parts in section 6.4. The left panel of Figure 6.9 also presents the constraints for $b_1\sigma_8$: with $k_{\rm max} = 0.2\ h\text{Mpc}^{-1}$, we obtain $b_1\sigma_8(z_{\rm eff} = 0.38) = 1.320 \pm 0.010$ for z1 and $b_1\sigma_8(z_{\rm eff} = 0.61) = 1.278 \pm 0.008$ for z3 sample.

Additionally, we present our constraints in terms of the parameter $S_8 \equiv \sigma_8(\Omega_m/0.3)^\gamma$, where $\gamma = {\rm dln}f\sigma_8/{\rm dln}\Omega_m \simeq 0.78 \cdot (1 - \Omega_m(z))$ [143]. For the galaxy samples considered in this work and for Planck fiducial value of Ω_m we have $\gamma = 0.37$ for z1 and 0.28 for z3, respectively. We find $S_8 = 0.821 \pm 0.037$ and 0.824 ± 0.054 with $k_{\rm max} = 0.2$ and $0.1\ h\text{Mpc}^{-1}$, respectively, consistent with the Planck's $S_8 = 0.832 \pm 0.013$ [11]. Extending to a higher $k_{\rm max}$ of $0.4\ h\text{Mpc}^{-1}$, we get $S_8 = 0.786 \pm 0.021$, about 2σ lower than the Planck constraint. In section 6.5 we argued that extending the model to $k_{\rm max}$ of $0.4\ h\text{Mpc}^{-1}$ may lead to model misspecification.



BOSS DR12 $f\sigma_8$ constraints (+ Planck prior on f)		
$k_{\rm max}$	sample	
(hMpc^{-1})	z1	z3
0.1	0.488 ± 0.049	0.456 ± 0.040
0.2	0.489 ± 0.036	0.455 ± 0.026
0.4	0.445 ± 0.022	0.447 ± 0.014

Figure 6.9: *Left*: 1D and 2D posterior distributions of two selected parameters $f\sigma_8$ and $b_1\sigma_8$ for all galaxy samples with $k_{\rm max}$=0.2 hMpc^{-1}. *Right*: The best-fit $f\sigma_8$ values and their 1σ uncertainties for the BOSS DR12 sample in two different redshift bins (including both NGC and SGC sky patches). We put the prior on f using the Planck 2018 prior for Ω_m [11]. For all results, we fit to the monopole, quadrupole, and hexadecapole, assuming the full covariance with the smoothed connected parts, as described in section 6.4.

Figure 6.10 compares our $f\sigma_8$ measurements to previous BOSS DR12 results in the literature [215, 15, 131, 68, 157, 58, 148, 284, 285], along with the constraint assuming the Planck 2018 ΛCDM cosmology [11]. [15] divides the BOSS galaxies into three redshift bins (z_1, z_2 and z_3) and provides the "consensus" $f\sigma_8$ constraints by combining measurements from seven companion papers. Our constraints are not in tension with the consensus analysis (with $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, 0.22σ lower for z1 and 0.73σ higher for z3) or with the Planck 2018 predictions (with $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, 0.39σ higher for z1 and 0.51σ lower for z3), while providing one of the tightest constraints on $f\sigma_8$ among recent works.

Figure 6.11 shows how different combination of dataset, wavenumber range, and covariance matrices may affect the cosmological analysis. With $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, we obtain $f\sigma_8 = 0.450 \pm 0.030$ for z3 NGC, assuming the full covariance including connected parts. Extending $k_{\rm max}$ to 0.4 hMpc^{-1} tightens the constraint significantly, while shifting the best-fitting parameter mean modestly ($f\sigma_8 = 0.458 \pm 0.018$). The dotted curves in the left panel present the constraint assuming the analytic disconnected covariance ($f\sigma_8 = 0.453 \pm 0.027$ and 0.461 ± 0.015 for $k_{\rm max} = 0.2$ and 0.4 hMpc^{-1}, respectively), and we thus find that including the connected parts inflates its standard deviation by 10-20%.

In the right panel of Figure 6.11, we show results of fitting only the monopole P_0 and quadrupole P_2 to quantify the impact of including hexadecapole P_4 on our constraints on the growth of structure. With P_0 and P_2 only, we obtain $f\sigma_8 = 0.452 \pm 0.034$ and 0.460 ± 0.023 for

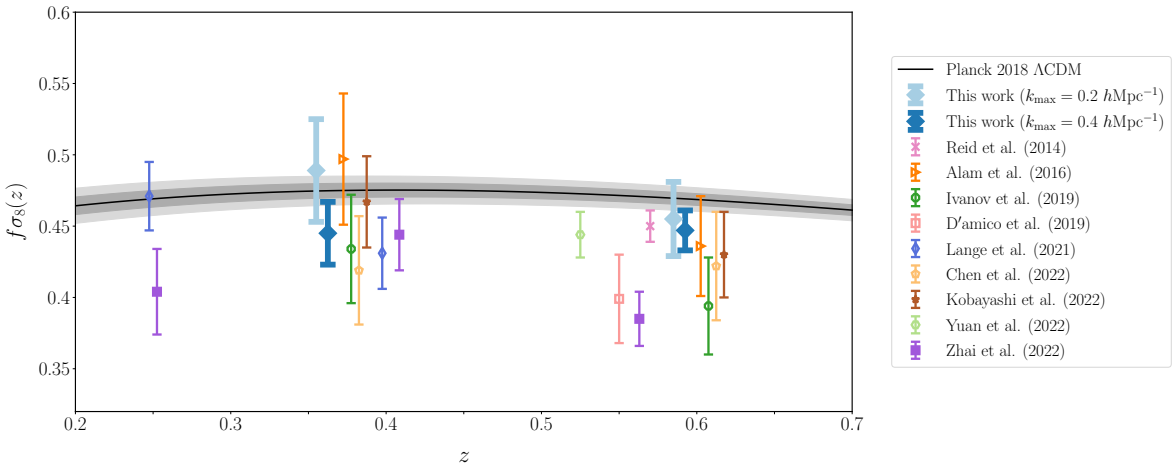


Figure 6.10: Comparison of $f\sigma_8$ constraints to previous BOSS DR12 measurements [215, 15, 131, 68, 157, 58, 148, 284, 285], along with the prediction assuming the Planck 2018 ΛCDM cosmology (black curve with grey shades) [11]. We show our main results as thick diamonds points, for two $k_{\rm max}$ limits: 0.2 (sky blue points) and 0.4 hMpc^{-1} (dark blue points). Our results, Alam et al. [15], Ivanov et al. [131], Chen et al. [58], and Kobayashi et al. [148] present measurements of $f\sigma_8$ for z1 and z3 galaxy samples ($z_{\rm eff} = 0.38$ and 0.61, respectively), but for graphical purpose they are plotted at different redshifts. Reid et al. [215], D'amico et al. [68], and Yuan et al. [284] show the constraints on the CMASS sample at $z_{\rm eff} = 0.57, 0.55$ and 0.52, respectively, Lange et al. [157] takes the galaxy samples at $z = 0.25$ and 0.4, and Zhai et al. [285] splits the galaxy sample into three redshift bins at $z_{\rm eff} = 0.25, 0.41$, and 0.56.

$k_{\rm max} = 0.2$ and 0.4 hMpc^{-1}, respectively, and find that the best-fit parameter mean remains consistent, while the standard deviation of $f\sigma_8$ increases by 15-30%, which is consistent with [37]. Therefore, we include the hexadecapole $P_4(k)$ in our main analysis (presented in Figure 6.9) because this improves RSD constraints significantly, as reported earlier in [37], [96], and [105].

In Figure 6.12 and Figure 6.13, we study the sensitivity of cosmological analysis to adding a fixed BAO template to the galaxy power spectrum model. Figure 6.12 shows the ratio of the wiggle ($P_{\rm W}$) to the no-wiggle ($P_{\rm NW}$) poles of the BOSS DR12 z3 NGC galaxy power spectrum using the best-fit cosmology; we use the fitting formula from [80] for BAO models. With this ratio, we create a fixed template for BAO wiggles and subtract it from the multipole measurements to get no-wiggle data. The left panel of Figure 6.13 shows z3 NGC multipole measurements (circular points with error bars): full data with BAO wiggles included (top) and data with BAO wiggles removed (bottom). Subsequently, we fit the theory model to both datasets, full and no-wiggle, and compare their $f\sigma_8$ constraints. The right panel of Figure 6.13 shows that the model constraints for both samples. Fitting to $k_{\rm max} = 0.2h$Mpc^{-1}, we obtain $f\sigma_8(z_{\rm eff} = 0.61) = 0.450 \pm 0.030$ and 0.459 ± 0.031 with and without BAO, respectively; we find that removing BAO information does not shift the model parameter constraints

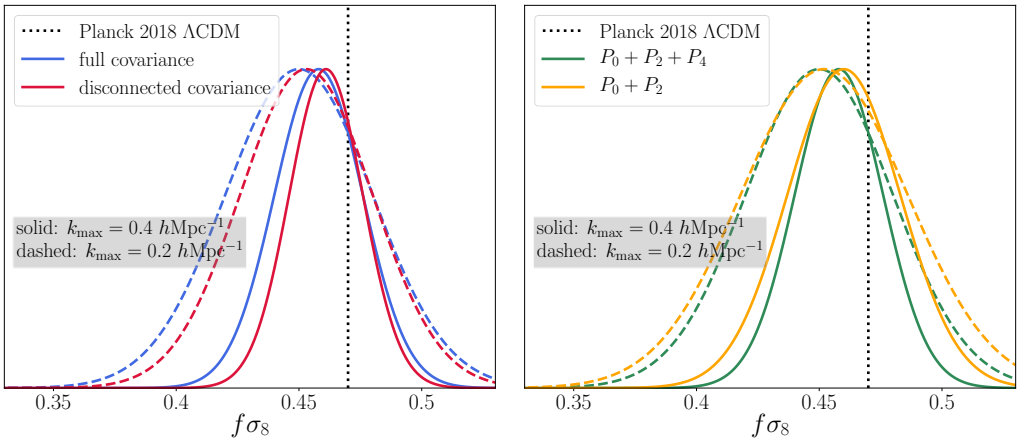


Figure 6.11: $f\sigma_8$ constraints of the BOSS DR12 z3 NGC sample, varying dataset, wavenumber range, and covariance matrices. *Left*: Constraints obtained with different covariances matrices in section 6.4. Adding the connected parts (blue) to the analytic disconnected covariance (red) weakens our constraints by 10-20%. We also show the impact of including wavenumbers in a wider range. Dashed curves assume $k_{\rm max} = 0.2\ h{\rm Mpc}^{-1}$, and extending it to $0.4\ h{\rm Mpc}^{-1}$ (solid curves) improves our constraints considerably. *Right*: Results obtained with the inclusion (green) or exclusion (orange) of the hexadecapole to quantify its impact on $f\sigma_8$ constraints. Excluding the hexadecapole inflates its standard deviation by 15-30%. For both results, we assume the full covariance matrix with the connected part.

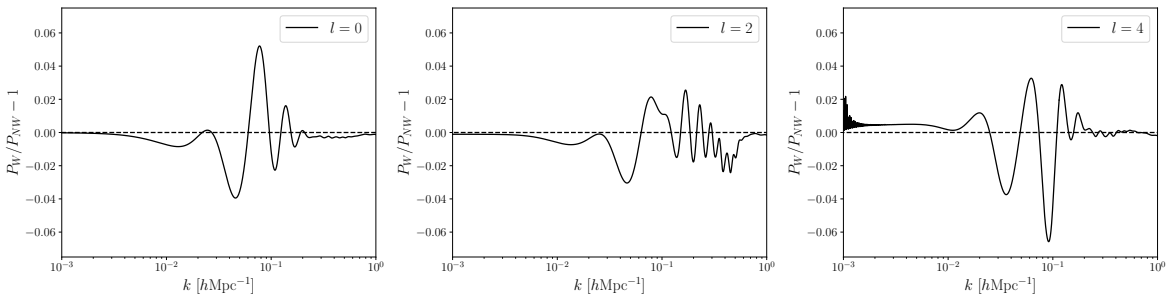


Figure 6.12: The ratio of the wiggle ($P_{\rm W}$) to the no-wiggle ($P_{\rm NW}$) poles of the BOSS DR12 z3 NGC galaxy power spectrum using the best-fit cosmology. The ratio for the monopole, quadrupole, and hexadecapoles are shown in the left, middle, and right panels, respectively.

significantly and fixing the BAO information with the fiducial template only minimally affects the cosmology. This result suggests a simplified large-scale structure analyis where the BAO information is completely independent of the de-wiggled power spectrum analysis, without a need to compute their covariance from mocks.

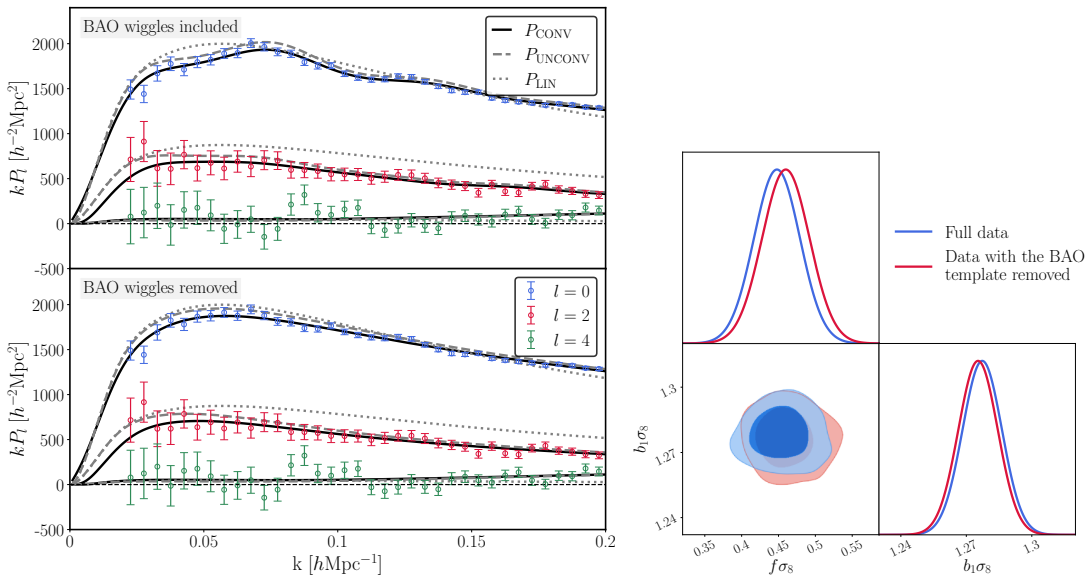


Figure 6.13: *Left*: (Top panel) Power spectrum multipole measurements of the BOSS DR12 z3 NGC sample with BAO wiggles. Solid curves indicate the best-fit theory model, convolved with the window function ($P_{\rm CONV}$). Dashed and dotted curves indicate best-fit unconvolved ($P_{\rm UNCONV}$) and linear ($P_{\rm LIN}$) theory models, respectively. (Bottom panel) We take the BAO wiggle template from figure 6.12 and subtract it from the measured data so that we can get the "no-wiggle" measurements. All best-fit theory lines also assume the models without BAO wiggles. *Right*: The model constraint with the full power spectrum and the no-wiggle fit with the fixed fiducial BAO template. Fitting to $k_{\rm max} = 0.2h$Mpc^{-1}, we obtain $f\sigma_8(z_{\rm eff} = 0.61) = 0.450 \pm 0.030$ and 0.459 ± 0.031 with and without BAO, respectively. Therefore, we argue that fixing the BAO information with the fiducial BAO template does not affect the cosmology considerably, which eliminates the need to do a joint power spectrum BAO covariance analysis on mocks.

6.7 Conclusion

We have presented the analysis of the SDSS-III BOSS DR12 galaxy sample, employing the redshift-space galaxy power spectrum model of [105], which can accurately model the monopole, quadrupole, and hexadecapole down to small scales. With the Planck 2018 prior on the growth rate f, we obtain 7.4% and 5.7% constraints on $f\sigma_8$ for BOSS DR12 low-redshift z1 ($z_{\rm eff} = 0.38$) and high-redshift z3 ($z_{\rm eff} = 0.61$) samples, respectively, fitting to $k = 0.2$ hMpc^{-1}. The combined error is 4%. Extending the wavenumber range to $k_{\rm max} = 0.4$ hMpc^{-1}, we find significant improvement in our constraint: 4.9% and 3.1% constraints on $f\sigma_8$ for z1 and z3 samples, respectively, which correspond to an overall 2.7% constraint. However, tests on MD-PATCHY mock catalogues suggest that the model fit to $k_{\rm max} > 0.2$ hMpc^{-1} may not be reliable, and more caution should be taken when extending to smaller scales. This is further supported by strong running of $f\sigma_8$ with $k_{\rm max}$ for $z1$ from 0.2 to 0.4 hMpc^{-1} seen

in Table 6.9, in contrast to little or no running from 0.1 to 0.2 hMpc^{-1}. We argue that a consistent choice of $k_{\rm max}$ is the main challenge of RSD analyses. With respect to the Planck 2018 ΛCDM cosmology predictions or the DR12 final consensus results, we find no tension in our $f\sigma_8$ constraints for $k_{\rm max} = 0.2$ hMpc^{-1}, while being 2 sigma lower for $k_{\rm max} = 0.4$ hMpc^{-1}.

Figure 6.10 presents a review of literature. Most of the measurements are below Planck, and this has been interpreted as evidence of σ_8 tension with Planck from BOSS RSD (e.g, [130, 204, 292, 59]). However, all the analyses are based on the same underlying data (galaxy positions and redshifts), and one cannot simply average these different analyses. The spread of the results is indicative of either model misspecification of some or all of the models, of the influence of the choice of priors, or of the choices made in the analysis, such as the scale cut $k_{\rm max}$ in power spectrum analysis, power spectrum versus correlation function versus wedges analysis, as well as several additional choices. It is perhaps disturbing that the spread is so large given that the underlying data are the same: at one end of the spectrum our results agree with Planck within 0.3 sigma. At the other end of the spectrum some analyses disagree with Planck at 3-4 sigma. These differences need to be understood so that we can establish the existence or absence of σ_8 tension in cosmology.

We developed the hybrid covariance matrix which combines the analytic disconnected (or more conventionally, "Gaussian") part [166], which accounts for the window function effect, and simulation-based connected part, smoothed by including up to four principal components. We demonstrate that the difference between the mock covariance and the hybrid covariance is clean, and the hybrid covariance is free from noise and hence more appropriate to be used in the likelihood analysis. Additionally, we show that using the disconnected covariance matrix underestimates the cosmological parameter constraints by 10-20%.

Furthermore, we provide growth of structure constraints without BAO information by constructing a fixed template for BAO wiggles and subtracting it from the BOSS DR12 multipole measurements. Comparing the constraints with and without BAO wiggles, we conclude that removing BAO information does not noticeably shift the cosmological parameter constraints and hence fixing BAO information with the fiducial template affects the cosmology only minimally. This method enables combining the power spectrum and BAO likelihoods independently, rather than computing their covariance matrix with mock simulations.

Finally, we note that the galaxy power spectrum model and analysis pipeline used in this work can be useful for extracting cosmological information from the next-generation galaxy redshift surveys, such as the Dark Energy Spectroscopic Instrument (DESI) [72] and the Euclid [158] and Roman [76] missions. DESI Emission Line Galaxy sample is expected to provide considerably larger constraining power, as its bias is lower and satellite fraction higher than the BOSS high-redshift z3 sample [72].

Chapter 7

Fundamental Plane of BOSS galaxies: Correlations with galaxy properties and impact on RSD measurements.

Fundamental plane of elliptical galaxies can be used to predict the intrinsic size of galaxies and has a number of plausible application to study cosmology and galaxy physics. In this chapter, we present a detailed analysis of the fundamental plane of the SDSS-III BOSS LOWZ and CMASS galaxies. For the standard fundamental plane, we find a strong redshift evolution for the mean residual and show that it is primarily driven by the redshift evolution of the surface brightness of the galaxies. After correcting for the redshift evolution, the FP residuals are strongly correlated with the galaxy properties and some observational systematics. We show that the variations in the FP between the central and satellite galaxies that has been observed in the literature can primarily be explained by the correlation of the FP with the galaxy luminosity. We also measure the cross correlations of the FP residuals with the galaxy density field. The amplitude of the cross correlations depends on the galaxy properties and environment with brighter and redder galaxies showing stronger correlation. In general, galaxies in denser environments (higher galaxy bias) show stronger correlations. We also compare FP amplitude with the amplitudes of intrinsic alignments of galaxies, finding the two to be correlated. Finally, using the FP residuals we also study the impact of intrinsic alignments on the constraint of growth rate using redshift space distortions. We do not observe any significant trends in measurements of the growth rate f as function of the amplitude of FP-density correlations, resulting in null detection of IA effects on RSD measurements. [1]

[1]This chapter is taken from "Fundamental Plane of BOSS galaxies: Correlations with galaxy properties, density field and impact on RSD measurements," *Singh S., Yu B., and Seljak B.* (arXiv: 2001.07700).

7.1 Introduction

Fundamental Plane (FP) of galaxies, an empirical relation between the size, surface brightness and the velocity dispersion of early type galaxies, has been proposed as a cosmological probe to estimate distances to galaxies, galaxy velocities [256], weak gravitational lensing magnification [32], doppler magnification of galaxies [52], impact of intrinsic alignments on galaxy selection functions [119]; in addition to its value as probe for galaxy physics.

For elliptical galaxies, a relation between the size, surface brightness and velocity dispersion can be derived from the virial theorem assuming constant mass to light ratio for the galaxies. Such relations have been observed for a long time [eg. 77, 74, 31, 227, 228], though the observed FP deviates significantly from the virial theorem predictions as the galaxies do not follow the simplified underlying assumptions. Furthermore FP has been observed to be a function of galaxy properties [eg. 235, 187, 228] and their environment [71, 135, 228].

The cosmological prowess of the FP arises from its ability to provide an estimate of the true intrinsic size of the galaxies (with some scatter). The observed galaxy size is affected by several processes, including estimates of cosmological distances, peculiar motion of galaxies (since we estimate distance through redshift), relativistic effects including the doppler shift, gravitational lensing and the effects of projecting three dimensional shapes onto the plane of the sky. Once the intrinsic size of a galaxy is known, the difference between the observed and the true size of the galaxies (hereby size residual or FP residual) can be used to study several of these effects. The size residuals can provide a (noisy) estimate of the peculiar velocity of individual galaxies which can be used to map the cosmological flows [256]. Cross correlations of the size residuals with the foreground galaxies (or clusters) can be also be used to measure the galaxy-lensing cross correlations [32, 128]. [52] also suggested measuring the dipole of the galaxy-size residual cross correlations to estimate the doppler magnification of the galaxies. [119] pointed out that due to the radial intrinsic alignment of galaxies and the projection effects, the size residuals are correlated with the local galaxy environment and if the selection function for a galaxy survey is sensitive to such residuals, it can introduce biases into measurements of redshift anisotropy of the galaxy auto correlations. Such an effect was tentatively detected by [180] using the fundamental plane of SDSS-III BOSS galaxies.

However, as pointed out earlier, the FP depends on the galaxy properties and their environment [eg. 71, 235, 187, 135, 228]. [135] detected the cross correlations between the FP residuals and the galaxy density field implying that the FP residuals are influenced by the galaxy environment. They further detected the dependence of the FP residuals on the galaxy type, with the brightest galaxies in groups having larger sizes than predicted by the FP while the satellite galaxies have smaller sizes. Such dependence of the FP residuals on the galaxy environment complicates the cosmological applications of the FP and detailed studies are required to understand such dependencies and avoid possible contamination to the cosmological inferences.

In this work, we extend the FP analysis from [135] to the BOSS LOWZ and CMASS sample galaxies. We estimate the fundamental plane for these samples as well as for subsamples using splits based on color, luminosity and environment of the galaxies. We also

study the cross correlations between FP residuals and galaxy density field and compare these cross correlations to those expected from the effects of intrinsic alignments as pointed out by [119]. Finally we also perform a detailed study of contamination to the galaxy clustering measurements from the radial alignments, similar to the analysis by [180], in an attempt to confirm their results.

Throughout this work we will use flat ΛCDM model with the Planck 2015 [208] cosmology with $h = 0.677$, $\Omega_m = 0.307$. Unless mentioned otherwise, distances are measured in unit of comoving h^{-1}Mpc. To compute the matter power spectrum, we use `CLASS` code [159] with halofit [261] prescription for the non-linear matter power spectrum. For redshift-space galaxy power spectrum calculations, we employ the FFT-based algorithm implemented in `nbodykit` [106], and use `pyRSD` [105] to compute the theoretical predictions of the redshift-space power spectrum of galaxies and run a likelihood analysis to find the best-fit theory model parameters.

7.2 Formalism

In this section we describe our formalism used in the estimation of fundamental plane and the residuals over it (size or FP residuals); the estimators and models used to study the cross-correlation between FP residuals and the density field; the power spectra and modeling of the redshift space galaxy clustering.

Fundamental Plane

To estimate the fundamental plane of galaxies (FP), we closely follow the methodology from [227] and [135]. We define the FP as

$$\log R_0 = a \log \sigma_0 + b \log I_0 + c + \sum_{i=1}^{N_z} d_i z_{cor}^i, \tag{7.1}$$

where R_0 is the physical radius of the galaxy, I_0 is the surface brightness and σ_0 is the velocity dispersion. Following [135], we also introduce polynomial terms dependent on the redshift of the galaxies. z_{cor} is the redshift of the galaxies in the CMB rest frame (correcting for the motion of the earth with rest to CMB rest frame) and is estimated as detailed in [227].

The physical size of the galaxy, R_0, is measured as (in units of kpc/h)

$$r_{cor} = r_0 \sqrt{q_{b/a}} \tag{7.2}$$

$$R_0 = D_A(z_{cor}) \tan(r_{cor}) \times 1000 \tag{7.3}$$

where r_0 is the angular galaxy size and $q_{b/a}$ is the axis ratio which is used to measure the circularized galaxy size, r_{cor} [31, 227], with both r_0 and $q_{b/a}$ measured using the de Vaucouleurs profile. $D_A(z)$ is the angular diameter distance in units of Mpc/h.

The surface brightness, I_0, is computed as

$$\log I_0 = -\frac{1}{2.5}\left[M_{ke} + 5\log\left(\frac{D_L}{D_{L0}}\right)\right] - \log(2\pi R_0) + 4\log(1+z_{cor}) \tag{7.4}$$

M_{ke} is the $k+e$ corrected absolute magnitude as defined in [241], $\frac{D_L}{D_{L0}}$ is the correction to the luminosity distance due to redshift correction (D_L is estimated using z_{cor} while D_{L0} is estimated using measured redshift in observer frame). $4\log(1+z_{cor})$ factor corrects for the cosmological dimming of the surface brightness [268].

We also correct the velocity dispersion, σ_0, for the effects of the fiber size (different correction for BOSS and SDSS spectra) as [227]

$$\sigma_0 = \sigma\left(\frac{r_{\rm fiber}}{r_{cor}/8}\right)^{0.04} \tag{7.5}$$

where $r_{\rm fiber} = 1''$ for BOSS and $r_{\rm fiber} = 1.5''$ for SDSS spectrographs. We identify the spectrograph from the date the spectra for the given galaxy was obtained and then apply the relevant correction.

FP residual for a galaxy is defined in terms of the fractional difference between the measured size and the size predicted using FP,

$$\lambda_{N_z} = \ln\frac{R_0}{R_{{\rm FP},N_z}} = \log R_0 - a\log\sigma_0 - b\log I_0 - c - \sum_{i=1}^{N_z} d_i z_{cor}^i, \tag{7.6}$$

where N_z refers to the order of polynomial used for fitting the FP as defined in eq. (7.1). We find that the upto third order polynomials in z in eq. (7.1) are not necesasrily sufficient to fully null the redshift evolution of the λ. To further reduce the effects of redshift evolution, we also fit the FP in redshift bins and will denote λ from such fits as λ^{zb}. We typically use bins with width $\delta z = 0.02$ for such fits to obtain $\lambda^{zb}_{N_z}$ for all galaxies and will subsequently carry out the following analysis in the same manner as λ (i.e. ignoring z binning).

[180] ignored the velocity dispersion measurements in their FP analysis. To study the influence of velocity dispersion, we also define

$$\log R_0 = b^I \log I_0 + c^I + \sum_{i=1}^{N_z} d_i^I z^i \tag{7.7}$$

$$\lambda^I_{N_z} = \ln\frac{R_0}{R^I_{{\rm FP},N_z}} \tag{7.8}$$

where we used superscript I to denote that the FP is only dependent on surface brightness and not the velocity dispersion. We note that this FP definition is not strictly equivalent to the FP definition used by [180] who defined the FP using magnitudes instead of surface brightness.

Galaxy-Galaxy correlations

Estimator

To compute the galaxy-galaxy cross-correlation function between two different samples, we use the Landy-Szalay estimator [156, 242]

$$\xi_{gg}(r_p, \Pi) = \frac{SD - DR_S - SR_D + R_S R_D}{R_S R_D}, \tag{7.9}$$

where r_p is the projected separation for a pair of galaxies, Π is the line of sight separation, D and S refers to the dataset (galaxies) being cross correlated ($D = S$ in case of auto correlations) and R_S and R_D refer to the set of random points corresponding to S and D samples. Product XY (eg. SD) refers to the binned weighted count of pairs across two samples with distances that are within the (r_p, Π) range of the given bin. The weight of a pair is the product of the galaxy weights that are described in section 7.3.

The projected correlation function is then obtained by integrating ξ_{gg} over the bins in Π

$$w_{gg}(r_p) = \sum_{-\Pi_{\max}}^{\Pi_{\max}} \Delta\Pi \, \xi_{gg}(r_p, \Pi). \tag{7.10}$$

Large values of $\Pi_{\max}$ are required to reduce the impact of redshift space distortions (RSD) on measured correlation function, even though measurement noise increases with larger $\Pi_{\max}$ [242]. To reduce the impact of redshift space distortions on the projected correlations, we use $\Pi_{\max} = 100h^{-1}$Mpc, with 20 bins of size $\Delta\Pi = 10h^{-1}$Mpc.

Separately, to analyze the line of sight anisotropy, we also compute the multipoles of the correlation function as

$$\xi_{gg,2\ell}(s) = \frac{2\ell + 1}{2} \int d\mu \xi_{gg}(s, \mu) L_{2\ell}(\mu) d\mu \tag{7.11}$$

where $s = \sqrt{r_p^2 + \Pi^2}$ is the separation between pair of galaxies in the redshift space and $\mu = \Pi/s$.

Modelling

The galaxy cross correlation function between samples S and D in redshift space is given by

$$\begin{aligned}\xi_{gg}(r_p, \Pi) = \int \mathrm{d}z W(z) b_{g,S}(s, z) b_{g,D}(s, z) r_{gg}(s, z) \int \frac{\mathrm{d}^2 k_\perp \mathrm{d}k_z}{(2\pi)^3} \\ \times P_{\delta\delta}(\vec{k}, z)(1 + \beta_S \mu_k^2)(1 + \beta_D \mu_k^2) e^{\mathrm{i}(\vec{r}_p . \vec{k}_\perp + \Pi k_z)}.\end{aligned} \tag{7.12}$$

where $s = \sqrt{r_p^2 + \Pi^2}$, b_g is the galaxy bias and is in general a function of redshift and scale, $P_{\delta\delta}$ is the matter power spectrum. The Kaiser factor $(1 + \beta\mu_k^2)$ accounts for the effects of

redshift space distortions [139] with $\beta = f(z)/b_g$, f is the growth function. We also introduced the cross correlation coefficient, $r_{gg}(s,z)$, between the two samples of galaxies but we will assume that $r_{gg}(s,z) = 1$ on all scales used for fitting the model ($r_p > 5h^{-1}$Mpc). $W(z)$ is the redshift weight accounting for the effective contributions from different redshifts to the measured correlation function and is given by [177]

$$W(z) = \frac{p(z)^2}{\chi^2(z)\mathrm{d}\chi/\mathrm{d}z}\left[\int \frac{p(z)^2}{\chi^2(z)\mathrm{d}\chi/\mathrm{d}z}\mathrm{d}z\right]^{-1}. \tag{7.13}$$

$p(z)$ is the redshift probability distribution for the galaxy sample.

To compute the projected correlation function, we will assume a scale independent bias, b_g and use the effective redshift, z, for our sample computed by integrating over weights $W(z)$. We then integrate over the correlation function multipoles to obtain the projected correlation function as [22]

$$w_{gg}(r_p) = \sum_{\ell=0}^{2} 2\int_0^{\Pi_{\max}} d\Pi \xi_{gg,2\ell}(r) L_{2\ell}(\Pi/r) \tag{7.14}$$

where $L_{2\ell}$ are the Legendre polynomials, prefactor 2 arises because we assume symmetry around $\Pi = 0$ and change the limits of integration. $\xi_{gg,2\ell}(r)$ are the correlation function multipole given as

$$\xi_{gg,2\ell}(r) = (-1)^{\ell}\alpha_{2\ell}(\beta(z))\frac{b_{g,S}b_{g,D}}{2\pi^2}\int dk k^2 P_{\delta\delta}(k) j_{2\ell}(kr) \tag{7.15}$$

where $j_{2\ell}$ are the spherical bessel functions. We use the package `mcfit` [166] to compute the correlation function multipoles. The coefficients $\alpha_{2\ell}(\beta)$ are given by [22]

$$\alpha_0(\beta) = 1 + 1/3(\beta_S + \beta_D) + 1/5\beta_S\beta_D \tag{7.16}$$
$$\alpha_2(\beta) = 2/3(\beta_S + \beta_D) + 4/7\beta_S\beta_D \tag{7.17}$$
$$\alpha_4(\beta) = 8/35\beta_S\beta_D \tag{7.18}$$

Galaxy-FP residual cross correlations

Estimator

To compute the cross correlations between the galaxy density and the FP residuals, we use

$$\xi_{g\lambda}(r_p,\Pi) = \frac{\lambda_S D - \lambda_S R_D}{R_S R_D}, \tag{7.19}$$

λ_S is the FP residuals for sample S , D is the sample of galaxies used as galaxy density tracers and R_S, R_D are the corresponding randoms sample. $\lambda_S D$ effectively refers to the pair counts, weighted with FP residuals λ,

$$\lambda_S D(r_p,\Pi) = \sum_{S,D}\lambda_S w_{SD}. \tag{7.20}$$

The $\sum_{S,D}$ is over all the galaxy pairs with separation within the (r_p, Π) bin limits, w_{SD} is the weight assigned to the pair of galaxies and λ_S is the FP residual from sample S. $S_\lambda R_D$ measures the same quantity with density tracer sample being replaced by the random points. Randoms subtraction can remove potential additive systematics that donot correlate with underlying galaxy density and also leads to optimal covariance [242].

We caution that this estimator can be biased if the $\langle\lambda\rangle$ is not zero, even after including the randoms subtraction. This is because the λ is estimated at the position of the galaxies and is hence weighted by the galaxy density field, which results in contribution from the galaxy clustering in case the $\langle\lambda\rangle$ is not zero, i.e.

$$\begin{aligned}\xi_{g\lambda}(r) &= \left\langle \left[(\lambda_0 + \langle\lambda\rangle)(1+\delta_g^S)\right] \delta_g^D \right\rangle (r) \\ &= \left\langle \lambda_0(1+\delta_g^S)\delta_g^D \right\rangle (r) + \langle\lambda\rangle \left\langle \delta_g^S \delta_g^D \right\rangle (r) \end{aligned} \tag{7.21}$$

where we used λ_0 to explicitly define the mean zero quantity. Thus before computing the correlation function, we subtract out $\langle\lambda\rangle$ even though FP definition and our fitting procedure ensures that it is very small.

The projected correlation function $w_{g\lambda}$ is then obtained by integration over line of sight as in eq. (7.10) and the multipoles are obtained as in eq. (7.11).

Modelling

Following [119], we assume that the deviations from fundamental plane are correlated with the tidal field due to the effects of intrinsic alignments of galaxy shapes, i.e. galaxy shapes are aligned with tidal field in the three dimensions and the projection effects then lead to correlations between tidal field and the projected shape and size of galaxies. Galaxy sizes are affected by the intrinsic alignments along the line of sight. λ can then be described in terms of matter field as

$$\lambda = -A_\lambda \zeta \left[\nabla_z \nabla_z \nabla^{-2} - \frac{1}{3}\right] \delta_m \tag{7.22}$$

$$\lambda = A_\lambda \zeta \left[\frac{1}{3} - \frac{k_z^2}{k^2}\right] \delta_m \tag{7.23}$$

$$\lambda = A_\lambda \frac{\zeta}{3} \left[1 - 3\mu^2\right] \delta_m \tag{7.24}$$

where we used $\zeta = \frac{C_1 \rho_{crit} \Omega_m}{D(z)}$ and $\mu_{\vec{k}} = k_z/\vec{k}$. Our sign convention implies that for $A_\lambda > 0$ galaxies in higher overdensities (larger δ_m) have larger size. Following convention of intrinsic alignments studies [eg. 134], we will use $C_1 \rho_{crit} = 0.0134$.

We note here that in general it is plausible that additional galaxy environment effects also affect the projected galaxy sizes, in which case the deviations from the fundamental plane can be written in terms of the trace of the tidal field, $\lambda \propto \nabla^2\phi \propto \delta_m$. This formalism also results in the similar form for λ as in eq. (7.24), but with different constants and different

line of sight anisotropy term as compared to $(1-3\mu^2)$. Such a model was assumed by [135] when modeling λ. We will use the form in eq. (7.24) to fit the measurements of projected correlation functions (where line of sight anisotropy has negligible effect due to large line of sight integration) and study the deviations from the model by comparing A_λ to the amplitude of intrinsic alignments of galaxies A_{IA}, where the expectation under the model is $A_\lambda = A_{IA}/2$ [119]. The primary difference between our model and that of [135] is that fitted values of A_λ are rescaled by a constant $\zeta/3 = C_1\rho_{crit}\Omega_m/3D(z)$.

To check for the impact of the $(1-3\mu^2)$ term, we will also compute the multipole moments of the galaxy-λ cross correlations and we will replace this factor with $(1+\beta_\lambda\mu^2)$, i.e.

$$\lambda = A_\lambda \frac{\zeta}{3}\left[1+\beta_\lambda\mu^2\right]\delta_m \tag{7.25}$$

where β_λ is a free parameter to be fit, with fiducial value set to $\beta_\lambda = -3$.

The cross correlation function of λ with galaxies in redshift space is given by

$$\begin{aligned}\xi_{g\lambda}(r_p,\Pi) =& A_\lambda\frac{\zeta}{3}\int \mathrm{d}z W(z) b_g(r,z) r_{cc}(r,z)\\ &\int \frac{\mathrm{d}^2k_\perp \mathrm{d}k_z}{(2\pi)^3} P_{\delta\delta}(\vec{k},z)(1+\beta_g\mu_k^2)(1+\beta_\lambda\mu_k^2)e^{\mathrm{i}(\vec{r}_p.\vec{k}_\perp+\Pi k_z)},\end{aligned} \tag{7.26}$$

where $r_{cc}(r,z)$ is the cross correlation coefficient between galaxies and matter. In this work, we will assume $r_{cc}(r,z)=1$ on the scales used to fit the model ($r_p > 5h^{-1}$Mpc). Note that Kaiser factor for RSD $(1+\beta\mu_k^2)$ is different from the galaxy clustering as we assume that only galaxy positions are affected by RSD and λ carries the $(1+\beta_\lambda\mu^2)$ term (we are ignoring the fact that the FP residuals are affected by RSD. RSD effects on λ scale as $\Delta\lambda \propto \frac{v}{H(z)D(z)}$, v is galaxy velocity and $D(z)$ is the line of sight distance to the galaxy).

The projected correlation function and multipoles are then computed using similar transforms as in galaxy clustering, eq. (7.14) and eq. (7.15). When computing the projected correlation function, $w_{g\lambda}$, we fix $\beta_\lambda = -3$, while when fitting the multipoles β_λ is a free parameter.

As shown in eq. (7.21), since the FP residuals are sampled at the position of the galaxies, the measured $\xi_{g\lambda} = \lambda_S(1+\delta_g^S)\delta_g^D$ is weighted by the galaxy density and thus includes higher order terms [see 46, for detailed study of this effect in measurements of intrinsic alignments of galaxies]. Detailed modeling of this effect is outside the scope of current work and we will ignore it in our models.

Galaxy-Intrinsic shear

We follow [241] and [244] for the measurements and modeling of intrinsic alignments. We only briefly describe the methodology here and refer the reader to [241] for more details.

Estimator

The cross correlations between galaxy shapes and the galaxy density field are measured as

$$\xi_{g+}(r_p, \Pi) = \frac{S_+D - S_+R_D}{R_S R_D}, \tag{7.27}$$

where S_+D refers to the summation over the radial shear, $\gamma_{+,S}$, measured in the coordinate frame defined by the pair of galaxies

$$S_+D = \sum_{S,D} \gamma_{+,S} w_{SD}. \tag{7.28}$$

where $\gamma_{+,S}$ is positive for radial alignment and negative for tangential alignments.

The projected correlation function w_{g+} is obtained by integration over line of sight as in eq. (7.10)

Modeling

We assume the nonlinear-alignment model [117] for modeling the alignment signal

$$\begin{aligned}\xi_{g+}(r_p, \Pi) =& A_I \zeta \int \mathrm{d}z W(z) b_g(r, z) r_{cc}(r, z) \\ & \int \frac{\mathrm{d}^2 k_\perp \mathrm{d}k_z}{(2\pi)^3} \times P_{\delta\delta}(\vec{k}, z)(1 + \beta\mu_k^2)(1 - \mu_k^2) e^{\mathrm{i}(\vec{r}_p.\vec{k}_\perp + \Pi k_z)},\end{aligned} \tag{7.29}$$

with the line of sight anisotropy $1 - \mu_k^2$ accounting for the projection effects [see 244, for a detailed analysis]. We will only use the projected correlation functions for intrinsic alignments where these terms have negligible effect.

Redshift-space galaxy power spectrum

The model for the galaxy power spectrum in redshift space is based on [105]. We only briefly summarize the formalism here, referring the reader to [105] for more details.

In this model, we follow the halo model formalism in [193] and separately model the 1-halo (correlations of galaxies in the same halo) and 2-halo (correlations of galaxies in different halos) contributions to the clustering of galaxies. To achieve such modeling, it is convenient to decompose the galaxy density field in redshift space into contributions from centrals and satellites:

$$\delta_g(\mathbf{k}) = (1 - f_s)\delta_c(\mathbf{k}) + f_s \delta_s(\mathbf{k}), \tag{7.30}$$

where f_s is the satellite fraaction, and δ_c and δ_s are the density field of centrals and satellites, respectively. The total galaxy power spectrum in redshift space, in turn, can be modelled as:

$$P_{gg}(\mathbf{k}) = (1 - f_s)^2 P_{cc}(\mathbf{k}) + 2 f_s (1 - f_s) P_{cs}(\mathbf{k}) + f_s^2 P_{ss}(\mathbf{k}), \tag{7.31}$$

where P^{cc}, P^{cs}, and P^{ss} are the centrals auto power, central-satellite cross power, and satellite auto power, respectively. We then separate 1-halo and 2-halo terms by further decomposing the galaxy sample into the following four subsamples: centrals without satellites (denoted as "type A" centrals), centrals with satellites ("type B" centrals), satellites with no other satellites ("type A" satellites), and satellites with other neighboring satellites ("type B" satellites). We also account for the Fingers-of-God (FoG) effect when modeling 1-halo and 2-halo terms in redshift space, by separately modeling the FoG effect from each subsample.

The model for the dark matter halo power spectrum in redshift space is based on the distribution function expansion [236, 191, 192, 272, 273, 45], and Eulerian perturbation theory and halo biasing model is applied to model the halo velocity correlator terms [273]. The results of N-body simulations are also used to calibrate key terms in the model.

The resulting galaxy power spectrum model depends on 13 physically-motivated parameters, which include: the Alcock-Paczynski (AP) effect parameters $\alpha_{||}, \alpha_{\perp}$, the growth rate f and the amplitude of matter fluctuations σ_8 evaluated at the effective redshift of the sample $z_{\rm eff}$, the linear bias of the type A centrals and type A and B satellites $[b_{1\,c_A}, b_{1\,s_A}, b_{1\,s_B}]$, the satellite fraction f_s, the fraction of type B satellites f_{s_B}, the mean number of satellite galaxies in halos with more than one satellite $\langle N_{>1,s}\rangle$, the centrals FoG velocity dispersion σ_c, The type A satellites FoG velocity dispersion σ_{s_A}, and normalization nuisance parameter for the 1-halo amplitude $f^{1h}_{s_B s_B}$. We follow the notations introduced in [105]. In this work, we fix the AP parameters to their fiducial values, 1.

We measure the clustering of galaxies using the multipole moments of the power spectrum $P_l(k)$. In this work, we take the FFT-based algorithm presented in [104], built upon the methods proposed in [41] and [234], and this allows fast evaluation of the estimator in [279]. Using the spherical harmonic addition theorem to expand the Legendre polynomials into spherical harmonics, we write the multipole estimator as:

$$P_l(k) = \frac{2l+1}{A}\int \frac{d\Omega_k}{4\pi} F_0(\mathbf{k}) F_l(-\mathbf{k}), \tag{7.32}$$

where Ω_k is the solid angle in Fourier space, $\mathcal{L}_l$ is the Legendre polynomial, w is the weight, A is the normalization defined as $A \equiv \int d\mathbf{r}[n(\mathbf{r})w(\mathbf{r})]^2$, and

$$\begin{aligned} F_l(\mathbf{k}) &= \int d\mathbf{r} F(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{L}_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}) \\ &= \frac{4\pi}{2l+1}\sum_{m=-l}^{l} Y_{lm}(\hat{\mathbf{k}}) \int d\mathbf{r} F(\mathbf{r}) Y^*_{lm}(\hat{\mathbf{r}}) e^{i\mathbf{k}\cdot\mathbf{r}}. \end{aligned} \tag{7.33}$$

The weighted galaxy density field $F(\mathbf{r})$ is given by

$$F(\mathbf{r}) = \frac{w(\mathbf{r})}{A^{1/2}}[n(\mathbf{r}) - \alpha n_s(\mathbf{r})], \tag{7.34}$$

where n and n_s are the number density field for the galaxy and randoms catalogs respectively, and the factor α normalizes n_s to n.

IA effects on RSD

To account for the effects of intrinsic alignments of galaxies on the selection functions of galaxies and hence the RSD measurements, we follow the formalism in [119, 180], defining the bias in observed galaxy density as:

$$\widehat{\delta}_g(\boldsymbol{x}, \lambda) = \delta_g(\boldsymbol{x}) + \epsilon(\lambda(\boldsymbol{x})) \tag{7.35}$$

To derive the error ϵ, we assume that the probability of a galaxy in the observed sample, O, is given as

$$P(O|T, \lambda) = P(T)P(\lambda)(1 + S(\lambda)) \tag{7.36}$$

Where T is the target sample, λ is the FP residual for the given galaxy and $S(\lambda)$ is the size dependent selection function. We can also write the number of galaxies with the observed value of λ as

$$N_\lambda = \frac{dN}{d\lambda} = NP(\lambda)(1 + S(\lambda)), \tag{7.37}$$

N is the total number of galaxies. Following the ansatz in [119, 180], we assume that the some galaxies are missed when they are aligned with the plane of the sky, i.e. have positive λ. Under this assumption and assuming that the intrinsic distribution λ is symmetric within T, we can write

$$S(\lambda) = \frac{N_\lambda - N_{-\lambda}}{N_\lambda + N_{-\lambda}} \tag{7.38}$$

where $N_{-\lambda}$ is the number of galaxies with a negative value of λ. Note that the mean of $S(\lambda)$ is zero by construction. $S(\lambda)$ is also zero if there are no size dependent selection effects, since we assume the λ distribution to be symmetric ($N_\lambda = N_{-\lambda}$ when no size dependent selection).

Galaxies have some random λ values, λ_R, due to intrinsic variations in galaxy properties as well as random projections. The additional λ_I sourced by intrinsic alignments acts as a small perturbation on top of the λ_R (we assume $\lambda_I << \lambda_R$). We can then write the error ϵ that is relevant for cosmological inferences as

$$\epsilon(\lambda_R(\boldsymbol{x}) + \lambda_I(\boldsymbol{x})) = S(\lambda_R) + \lambda_I \frac{\partial S(\lambda)}{\partial \lambda}|_{\lambda=\lambda_R} \tag{7.39}$$

Since λ_R and hence $S(\lambda_R)$ does not correlate with the density field, the relevant part of ϵ that correlates with the density field can be written as

$$\epsilon(\mathbf{x}) \approx \gamma A_\lambda \zeta \left[\frac{1}{3} - \mu^2\right] \delta_m(\boldsymbol{x}) \tag{7.40}$$

where we used the relation of λ to tidal field as defined in eq. (7.24) and defined the ensemble response γ as

$$\gamma = \left\langle \frac{\partial S(\lambda)}{\partial \lambda} \right\rangle = \frac{1}{\int d\lambda N_\lambda} \int d\lambda N_\lambda \frac{\partial S(\lambda)}{\partial \lambda} \tag{7.41}$$

The error in the growth rate and galaxy bias measurements is given by

$$\Delta f = \widehat{f} - f_0 = -\gamma A_\lambda \zeta \tag{7.42}$$

$$\widehat{b}_g = b_{g,0} + \frac{1}{3}\gamma A_\lambda \zeta = b_{g,0} - \frac{1}{3}\Delta f \tag{7.43}$$

We will determine γ from the distribution of FP residuals. We note that since the datasets we will be using are already affected by the selection effects, $S(\lambda)$, this can introduce a bias the estimations of the γ (the mean of λ distribution is shifted which biases S). We will work under the assumption that the selection effects are small and hence the shift in the mean of the distribution and the bias in γ is also small. A_λ will be determined using the cross correlations between the FP residuals and the galaxy density field and ζ is a cosmology dependent constant. We will also compute the variations in f and b from our RSD analysis and these measurements will allow us to test the eq. (7.42) and eq. (7.43).

In section 7.4, we will also split the galaxy sample into two subsamples according to the sign of FP residuals, following [180]. These subsamples are expected to have almost identical γ values: $\gamma^+ \approx \gamma^-$, where $+$ and $-$ indicates samples with positive and negative FP residuals, respectively. That is because $\partial S(\lambda)/\partial\lambda$ is mirrored across the y-axis. Therefore, we expect that the growth rate measurements between the two FP residual subsamples is consistent: $\Delta f^+ - \Delta f^- = -(\gamma^+ - \gamma^-)A_\lambda\zeta \approx 0$, unlike the model assumed by [180].

Covariance matrices

For the correlation function measurements, we compute the covariance matrices using Jackknife resampling method by splitting the sample into 100 approximately equal area patches (68 patches in North Galactic Cap (NGC) and 32 in South Galactic Cap (SGC) of BOSS data described in section 7.3). Since the jackknife covariances are noisy, leading to biased inverse matrix, we also apply the Hartlap correction [112] when computing the inverse covariance matrix used in the likelihood functions.

For the multipoles in Fourier space, we assume the theoretical Gaussian covariance, following [97]:

$$C_{\ell_1\ell_2}(k_i, k_j) = \frac{2(2\pi)^4}{V_{k_i}^2}\delta_{ij}\int_{k_i-\Delta k/2}^{k_i+\Delta k/2}\sigma^2_{\ell_1\ell_2}(k)k^2\mathrm{d}k, \tag{7.44}$$

where $V_{k_i} = 4\pi[(k_i + \Delta k/2)^3 - (k_i - \Delta k/2)^3)^3]/3$ is the volume of the shell in k-space. When the expected mean number density $\bar{n}(z)$ is varying, the per-mode covariance $\sigma^2_{\ell_1\ell_2}(k)$ is given by [279]:

$$\begin{aligned}\sigma^2_{\ell_1\ell_2}(k) =& \frac{(2\ell_1+1)(2\ell_2+1)}{A^2}\int_{-1}^{1}\mathrm{d}\mu\int_{V_s}\mathrm{d}\boldsymbol{r}\;\bar{n}^4(\boldsymbol{r})w^4(\boldsymbol{r}) \\ & \left[P(k,\mu,z) + \bar{n}^{-1}(\boldsymbol{r})\right]^2\mathcal{L}_{\ell_1}(\mu)\mathcal{L}_{\ell_2}(\mu),\end{aligned} \tag{7.45}$$

where the normalization terms A is defined as $A \equiv \int d\mathbf{r}[\bar{n}(\mathbf{r})w(\mathbf{r})]^2$, following the notations in section 7.2.

For the galaxy subsamples we define in section 7.3, we compute their mean number densities in order to obtain correct covariance matrices using eq. 7.44.

7.3 Data

In this work we use the LOWZ ($0.16 < z < 0.43$) and CMASS ($0.43 < z < 0.7$) spectroscopic samples from SDSS-III BOSS [44, 50, 12, 70, 249] data release 12 [DR12; 16], which are selected using the imaging data from SDSS-I and SDSS-II surveys. The SDSS survey [99, 281, 121, 132, 88, 250, 81, 98, 219, 257] [174, 205, 270, 2, 13, 199]. To fit the FP, we also obtain the photometric measurements for our galaxies, specifically the Radii, axis ratios and magnitudes from de Vaucouleurs fits from the SDSS photometric catalog. The magnitudes are corrected for the extinction and we also apply k-corrections using the formalism of [276].

When computing galaxy clustering and galaxy-FP cross correlations, we apply weights to the galaxies, where the weights are given by [223]

$$w = w_{\rm sys}(w_{no-z} + w_{cp} - 1), \tag{7.46}$$

where $w_{\rm sys}$ weights correct for the variations in the selection function on the sky (important for CMASS) and w_{no-z}, w_{cp} correct for missing redshifts due to failure to obtain redshift (no-z) or fiber collisions for close pairs, *cp*.

For intrinsic alignments of galaxies, the shape sample used to estimate the shear is described in more detail in [216] and [241].

Furthermore, to study the dependence of FP and residuals on galaxy properties, we also split the BOSS samples based on color and luminosity. We follow the procedure from [241], whereby we split the samples in narrow redshift bins based on the percentiles of the color and luminosity. We make 5 color samples, $C_1 - C_5$, with each sub-sample containing 20% of the galaxies such that C_1 starts with reddest galaxies and the subsequent samples contain progressively bluer galaxies. Split in the redshift bins ensures that each sample has the same redshift distribution. We follow the similar procedure for luminosity and make 4 luminosity subsamples, $L_1 - L_4$ with L_1 being brightest and L_4 being faintest. $L_1 - L_3$ contain 20% of the sample each while L_4 contains 40% of galaxies.

For the LOWZ sample, we also identify the galaxies in groups using the counts-in-cylinders technique [214] using the same procedure as was followed in [241]. Galaxies that are not in groups (or are in group of 1) are designated as 'Field' galaxies, the brightest galaxy in a group of more than 1 galaxy is designated as BGG (brightest group galaxy) while all non-BGG galaxies are designated as satellite galaxies.

Since we are splitting the samples by color, luminosity and also FP residuals, the variations in the photometry across the sky can lead to some variations in selection function of the galaxies for these sub-samples. The variations can introduce biases when computing the galaxy auto correlation functions for these galaxies. To avoid this issue, we will use cross

correlations when computing the correlation functions, where we cross correlate the sub-samples with the full sample from which they were selected. As a result the biases in the signal are reduced though the covariance will still be affected [242]. In the power spectrum measurements, we will use the auto correlations. To reduce the impact of selection functions on sub samples, we reweight the randoms provided by BOSS to correspond to these samples. We compute the weight using the ratio of number galaxies within the sub-samples to number of galaxies in the full sample, within each $\sim 80\,\mathrm{deg}^2$ patch we generated for the jackknife calculations. We also tried the weights computed in much smaller patches, but those weights biased the measurements on small scales.

7.4 Results

In this section we present our results, beginning with the fits of FP to LOWZ and CMASS as well as various subsamples and analysis of FP residuals based on some galaxy properties. Then we present the measurements of cross correlations between FP residuals and galaxy density and comparison of these cross correlations with the Intrinsic alignments of galaxies. Following this we present the measurements of redshift space distortions (RSD) and the correlations between RSD constraints and FP residuals.

FP Fits

In this section we present results of fitting FP to full LOWZ and CMASS samples and an analysis of the FP residuals based on the redshift, luminosity and environment of the galaxies. Note that unless mentioned otherwise, in this subsection, FP residuals are obtained from fitting FP to the full samples and not the sub-samples.

In figure 7.1, we show the contour plots of the FP residuals as function of galaxy size, magnitude and velocity dispersion. For the standard FP, λ, we obtain the RMS values of $\lambda_{rms} = 0.22$ for LOWZ and $\lambda_{rms} = 0.26$ for CMASS sample and for the redshift dependent FP, λ_3, we obtain $\lambda_{rms} = 0.15$ for LOWZ and $\lambda_{rms} = 0.16$ for CMASS samples. Note that λ is defined in $\log_e$ (ln) base and hence λ_{rms} is larger than the scatter in $\log_{10} R$ space by a factor of $\ln 10 \sim 2.3$. After accounting for this effect, our results are consistent with the FP scatter of ~ 0.1 obtained by [228, 135], albeit for somewhat different samples.

Also note in figure 7.1 that the FP residuals are correlated with the galaxy properties. We now investigate some of these correlations in more detail.

Redshift dependence

In figure 7.2 we present the redshift dependence of the mean and the RMS of the FP residuals for both LOWZ and CMASS samples. For the standard FP we find a strong correlation of the mean residuals, λ_0, with the redshift. Including redshift dependence within the FP using polynomials reduces the dependence of the mean by a large magnitude and

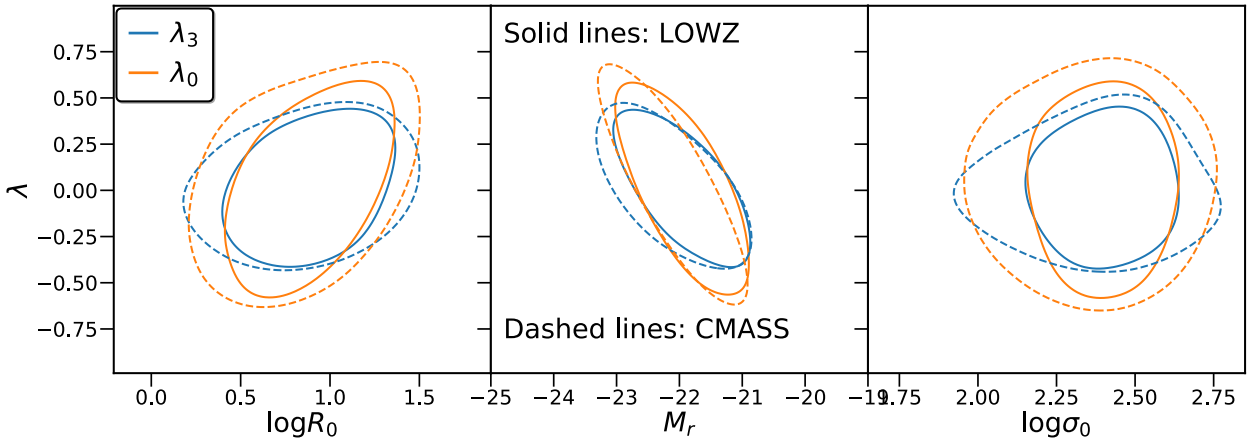


Figure 7.1: The residuals over the FP for standard FP, λ_0 and the redshift dependent FP λ_3 as function of galaxy size, R_0, r band magnitude M_r and the velocity dispersion σ_0. We show the contours covering 95% of the sample for both LOWZ (solid lines) and CMASS (dashed lines) samples.

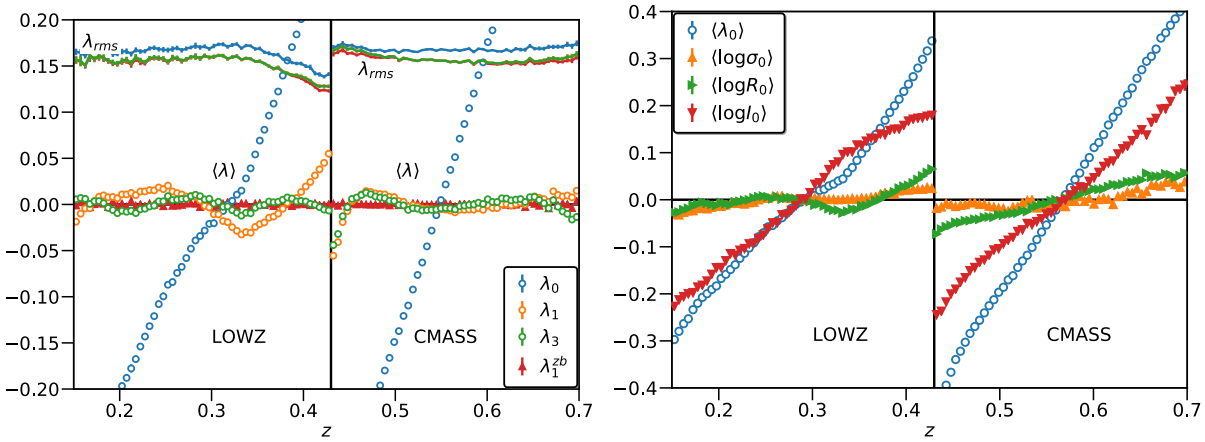


Figure 7.2: a) Redshift dependence of the mean and standard deviation (mean subtracted RMS) of the FP residuals for both LOWZ and CMASS samples. The subscript, λ_i, $i \in [0, 1, 3]$ denotes the order of the polynomial in z used for fitting FP. The superscript 'zb' denotes the sample fitted in narrow redshift bins $\Delta z = 0.02$. The residuals over for the standard fundamental plane, λ_0, have strong dependence on the redshift and including redshift polynomials in the FP reduce this dependence as well as the scatter. Fitting FP within small redshifts further reduces the mean of FP (though there can be evolution within the bins). b)Redshift evolution of galaxy properties that are included in the FP. The redshift dependence of the residuals for standard FP can be explained by the redshift dependence of these properties, especially the surface brightness of galaxies, $\log I$, which has the strong and monotonic dependence on redshift, driven by the $\log(1+z)$ correction for the Tolman dimming.

also reduces the RMS (as function of z) by $\sim 10\%$. The change (reduction) in the mean is largest when using the first order polynomial with some further improvement when going to the third order polynomial. There are still small residuals correlations between the mean and redshift and such correlations can potentially be important for the correlation functions we present in section 7.4 and for the cosmological applications of the FP in general. To further reduce correlation between mean and the redshift, we instead fit the FP in narrow redshift bins, $\Delta z = 0.02$ (this choice is motivated to keep the bin size small but have large enough line of sight size so as to not bias the cross correlation measurements presented in section 7.4). Fits in the bins further reduce the impact of the redshift dependent residuals. We will use λ_3 as our fiducial FP (unbinned with third order polynomial in redshift), but we will test the cross correlation results with the binned FP for comparison.

To study the source of the redshift dependence of FP, we show the redshift dependence of the galaxy properties in figure 7.2. The velocity dispersion and physical radius only have mild dependence on redshift for both LOWZ and CMASS sample. The surface brightness of the galaxies on the other hand evolves strongly with redshift and is the primary driver for the redshift evolution of the standard FP residuals, λ_0. The redshift evolution of surface brightness is driven by redshift dimming [268] of the surface brightness, as a result of which we only observe galaxies with larger surface brightness at higher redshifts. The evolution of surface brightness is very similar to the $\log(1+z)$ correction that was included in the eq. 7.4. Hence, when including the redshift polynomials into the FP, we are essentially undoing the $\log(1+z)$ correction. Including $\log(1+z)$ dependence in the FP is also a possibility instead of the polynomials in z, but we opt for the polynomials (or fit in narrow z bins) as they provide extra degrees of freedom which can account for the small redshift dependence of the velocity dispersion, physical radius and luminosity.

Our results on the redshift dependence of the FP are qualitatively consistent with those from [135], where such trends were observed for the SDSS main galaxy sample. Due to the differences in the samples used in this study and [135], a more quantitative comparison is difficult but we have tested our pipelines on the samples used in [135] and we are able to reproduce their results.

Luminosity and environment dependence

In figure 7.3, we show the dependence of the FP residuals on the environment as function of redshift. As in [135], we observe that the brightest group galaxies, *BGGs* have positive residuals on average (they are larger in size than predicted by FP) while satellites have negative residuals implying they are smaller than average. It is tempting to interpret these results based on the environment dependence of galaxies, whereby centrals or BGGs tend to be larger but less concentrated while satellites under going tidal stripping tend to be smaller and more concentrated. However, in figure 7.3 we observe that the FP residuals are strongly correlated with the galaxy luminosity, where brighter galaxies have larger (or more positive) residuals. These trends can explain most of the variations between different galaxy types in

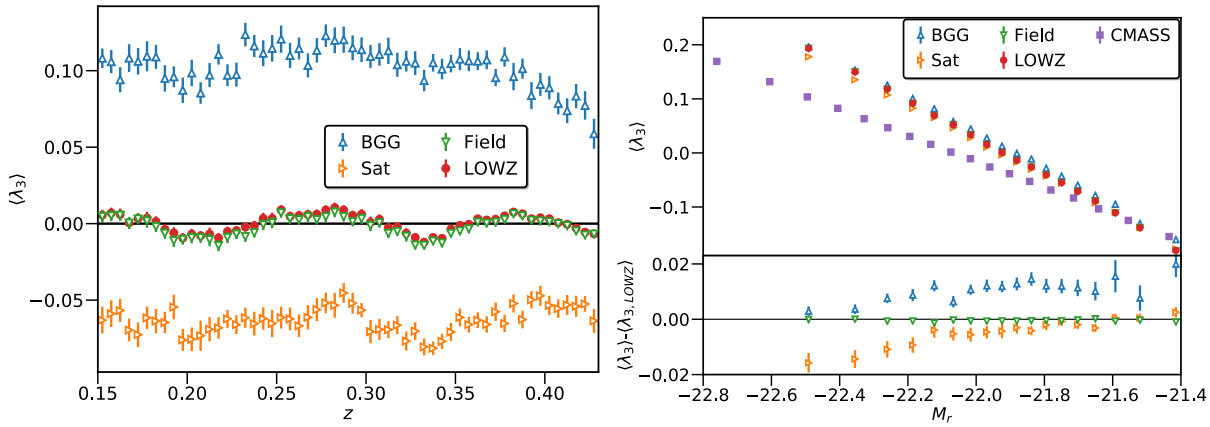


Figure 7.3: a) Dependence of the FP residuals on galaxy types. Similar to [135], we observe that Brightest group galaxies (BGG), satellites and field galaxies have different FP residuals, with BGGs being larger than FP predictions while satellites being smaller. b) Mean FP residuals as function of the r band magnitude. Different types of galaxies, BGGs, Satellites, Field, all give very similar relation which suggests that the dependence in a) can be explained largely by magnitude (or luminosity) dependence of the FP residuals. In the lower panel we show the difference in the mean FP residuals of different samples relative to the full LOWZ sample. BGGs (satellites) are still higher (lower) than the full sample, though the differences are much smaller than in a). We do not observe any significant dependence of RMS of λ with luminosity.

figure 7.3, with BGGs (satellites) being only marginally larger (smaller) after accounting for the luminosity evolution, as shown in the lower panel of figure 7.3.

Given that $\lambda \propto -b \log I \propto bM_r$ (b is negative), the luminosity dependence of FP residuals λ is expected in case the various galaxy properties, namely luminosity, radius and velocity dispersion are not perfectly correlated in a way to cancel such a dependence. While it is possible to include the higher order luminosity dependence in the FP as well, we opt not to do so as FP residuals are correlated with multiple galaxy properties and systematics (see appendices) and including dependence on too many variables complicates the interpretation of FP and its residuals (inclusion of z-dependence is necessary to avoid biases in cross correlation measurements presented in section 7.4). Instead, we split our sample based on luminosity, color and environment as described in section 7.3 and fit the FP separately to those subsamples when studying the dependence of cross correlations on these galaxy properties.

Cross correlations with galaxy density

In this section we present the measurements of galaxy clustering and cross correlations between the galaxy positions and FP residuals. Throughout this section we will use FP with third order polynomial as the primary FP and when considering the subsamples, we will fit

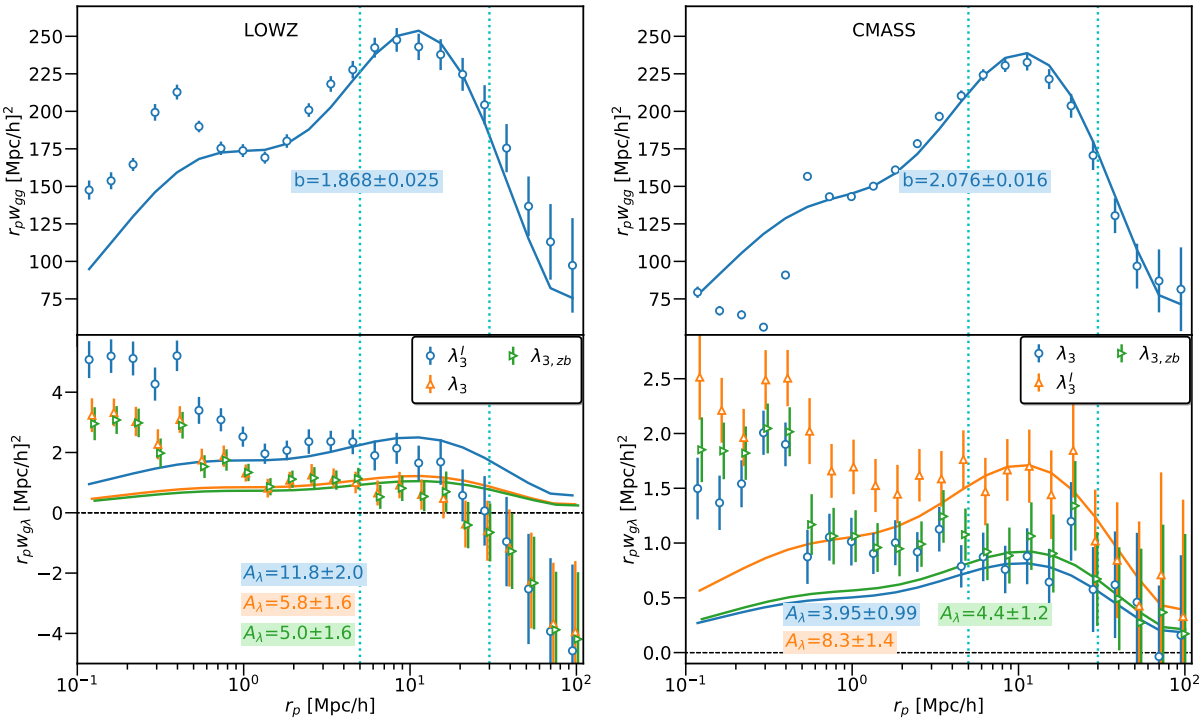


Figure 7.4: Measurements of galaxy clustering (upper panels) and galaxy-λ (lower panels) cross correlation functions for (a) LOWZ and (b) CMASS samples. Lower panel shows measurements with three different FP definitions, λ_3^I (blue, FP without velocity dispersion), λ_3 (orange) and FP fitted in z-bins $\lambda_{3,zb}$. We measure strong correlations between galaxies and FP residuals for both CMASS and LOWZ samples, with λ_3^I signal being factor of ~ 2 larger than the λ_3. Given that the mean of λ_3^I and λ_3 are very similar and λ_3^I rms is larger by $\sim 10\%$, this difference is originating from the intrinsic differences between the two FP planes. Also, the consistency between λ_3 and $\lambda_{3,zb}$ suggests that the negative signal in LOWZ at large scales is unlikely due to any redshift dependent additive systematics in the FP (as observed in figure 1). Numbers quoted in the plots are the best fit galaxy bias and λ amplitude A_λ obtained by fitting the model in range $5 < r_p < 30h^{-1}$Mpc (marked by vertical dashed cyan lines). For the LOWZ sample, the $\chi^2_{dof} \sim 0.7$, even though the fit looks inconsistent with the data. This is due to strong correlations between the bins on large scales, likely driven by systematics.

FP to each of the subsamples separately.

We begin by presenting the measurements of projected correlation functions using full LOWZ and CMASS samples in figure 7.4, with FP fit to the whole sample, λ_3, FP fit in narrow redshift bins, $\lambda_{3,zb}$ and the FP without velocity dispersion, λ_3^I. We fit the models described in section 7.2 to both galaxy clustering and galaxy-λ cross correlations in the range $5 < r_p < 30h^{-1}$Mpc. We do not fit the scales below $r_p < 5h^{-1}$Mpc as our model with assumption of linear bias is not expected to work well on these non-linear scales and we also do not use $r_p > 30h^{-1}$Mpc, as there is some evidence of systematics in the galaxy-λ cross

correlations. We have checked that including scales between $30 < r_p < 70h^{-1}$Mpc does not significantly change the best fit parameters as measurements on those scales are correlated and also have lower signal to noise compared to smaller scales.

For LOWZ sample, we obtain the linear galaxy bias, $b_g = 1.868 \pm 0.025$ and for CMASS sample we obtain $b_g = 2.096 \pm 0.019$. Using galaxy-λ_3 cross correlations , we obtain $A_\lambda = 5.8 \pm 1.6$ for LOWZ sample and $A_\lambda = 5.3 \pm 1.3$ for CMASS sample, with CMASS sample have lower noise due to larger volume of the sample, even though $\lambda_{\rm rms}$ is larger for CMASS sample. The values of A_λ do not change significantly if we fit FP in narrow redshift bins ($\lambda_{3,zb}$ in figure 7.4) and/or if we fit the FP for North and South regions of BOSS separately (measurements not shown). However, fitting the FP without velocity dispersion , λ_3^I, leads to significantly larger amplitude, with A_λ being larger by factor of $\sim 2-5$ depending upon the sample and the FP definition.

To understand the source of the correlations, in figure 7.5 we show the cross correlations of galaxy properties, the surface brightness $\log I$, physical radius, $\log R$ (orange) and velocity dispersion, $\log v$ in figure 7.5. The surface brightness shows strong correlations with the density field, with a large fraction of the signal driven by its strong evolution with redshift. These correlations provide interesting insights that the larger galaxies and ones with higher velocity dispersion tend to reside in the over dense regions but galaxies in over dense regions tend to have lower surface brightness.

The impact of these parameters can be observed in the lower panel of figure 7.5. The correlations of the FP residuals, λ, are essentially the weighted sum of the correlations of the galaxy properties, where the weights are the parameters of the FP. For λ_0, surface brightness dominates given that the redshift evolution has not been corrected for and leads to large negative correlations. Once redshift evolution is corrected, the correlations of surface brightness decrease and hence the FP cross correlations become positive for λ_1 and λ_3. This further justifies our choice to include the redshift evolution in the FP, as the correlation functions otherwise are dominated by the redshift evolution of the FP which itself is dominated by the redshift evolution of the surface brightness.

We also note that our measurements of galaxy-λ_3 cross correlations are not consistent with the results of [135]. This is because of the very different samples used in the two studies. As observed in section 7.4, the FP residuals are strongly correlated with the galaxy luminosity, with fainter galaxies having negative λ. Fainter sub-samples also show negative correlations between FP residuals and density, in our measurement. Given that SDSS main sample used in [135] was fainter than BOSS samples, we hypothesize that the different measurements in the two studies are primarily due to the different galaxy samples. In order to rule out pipeline differences, we also reanalyzed the data of [135] with our current pipeline and reproduced their results[2] (not shown).

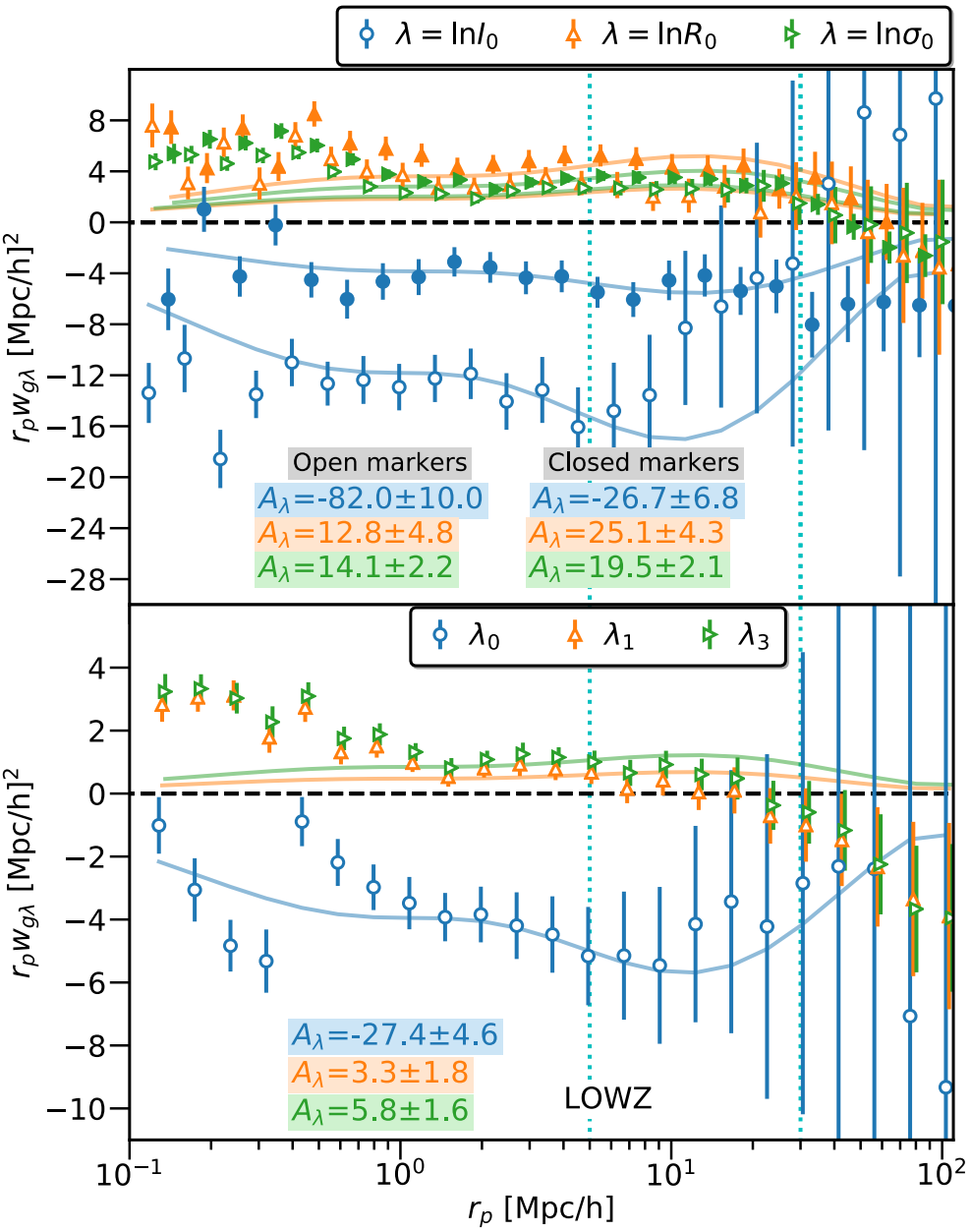


Figure 7.5: Upper panel: Galaxy-λ cross correlation functions, where λ is replaced with different galaxy properties, namely the surface brightness, $\log I$ (blue), physical radius, $\log R$ (orange) and velocity dispersion, $\log v$ (green). For the open points, we set the mean of these galaxy properties to be zero only at the level of full sample, while for the closed points, the mean is set to zero within small redshift bins, $\Delta z = 0.02$. Size and velocity dispersion are positively correlated with the density field, though surface brightness shows negative correlations which are also stronger when redshift evolution is not corrected for. Lower panel: Cross correlation measurements using residuals from different definitions of fundamental plane. Standard FP residuals , λ_0 is negatively correlated driven by the effects of surface brightness, while FP corrected for redshift evolution show positive correlations with density.

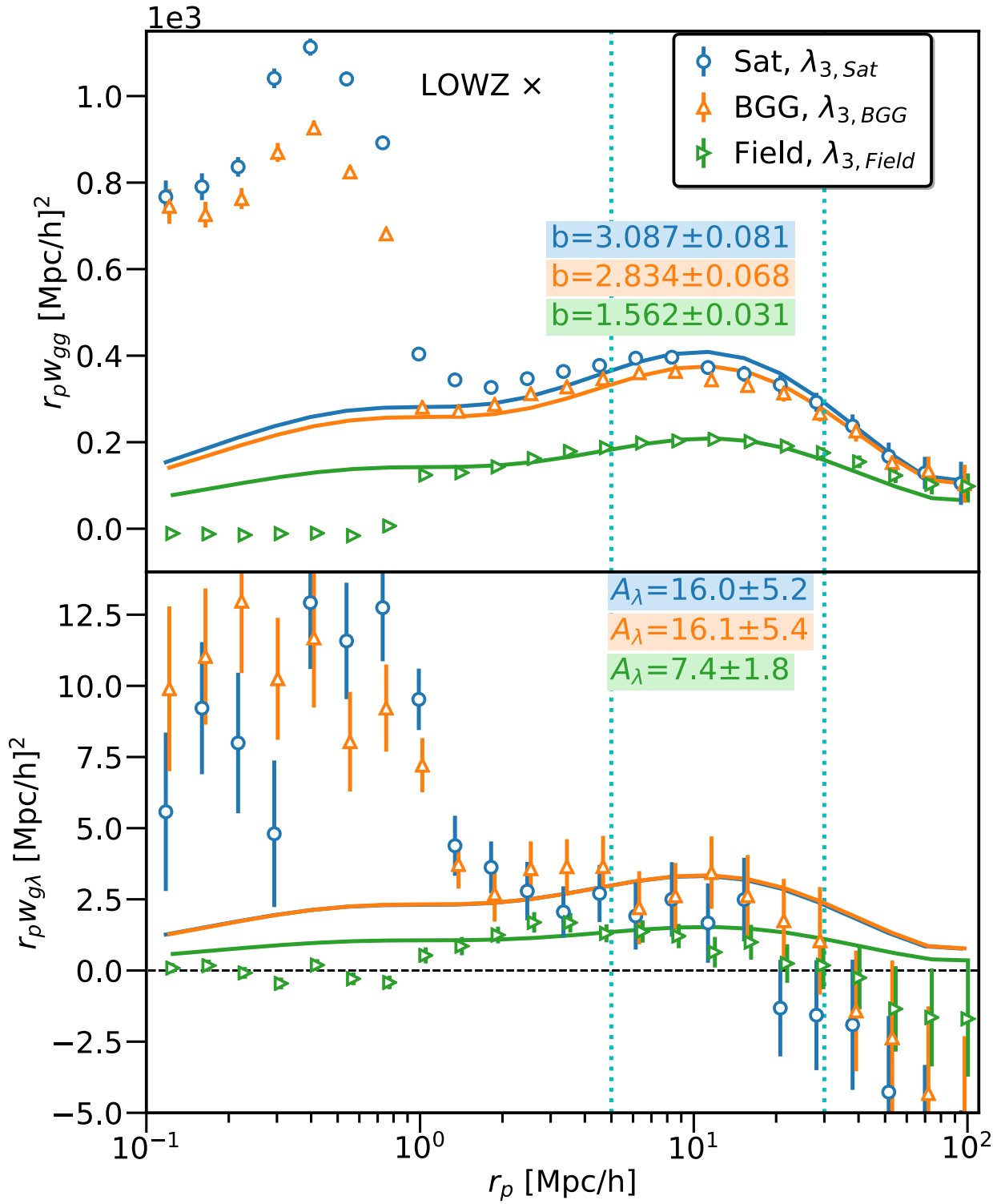


Figure 7.6: Galaxy clustering (cross correlation with LOWZ) and galaxy-λ cross correlation measurements for different environment samples, with full LOWZ sample as density tracers. Group galaxies, BGGs and Satellites, have similar clustering and galaxy-size cross correlations, unlike intrinsic alignments, where satellites do not show large scale shape alignments.

Environment dependence

In figure 7.6, we show the correlations of FP residuals for group galaxies (satellites and BGGs) as well as the field galaxies. Satellites and BGGs have higher galaxy bias as expected since they are on average in more massive halos and hence more dense environments. More interestingly, FP residuals for both satellites and BGGs show much stronger correlations with the density field as compared to the field galaxies.

A_λ is rather large and also very similar for both Satellites and BGGs though the uncertainties for both samples are also large. Primary concern with such large signal is that since these galaxies reside in crowded regions, there can be some residual systematics in the photometry leading to such correlations. Since we subtract the randoms signal, additive systematics are unlikely to be able to lead to such large signals unless they strongly correlate with the galaxy density field. We have some evidence for systematics affecting the measurements, especially for LOWZ sample, but those systematics are predominately on large scales and do not lead to such large and significant A_λ values. We have also tested for the flags in SDSS photometry for blending and other photometry issues and BGGs, satellites and field galaxies all have very similar (low) rate of problematic photometry flags. Thus it is unlikely that satellite and BGGs results in particular are affected by the photometry problems.

Another possible explanations for such similarities is that the galaxy environment plays a strong role in determining the galaxy size, in addition to galaxy properties such as luminosity and size. To further study the impact of the environment on FP residuals, in figure 7.7 we show the A_λ as function of galaxy bias, b_g, where bias is a proxy for the galaxy environment. Though there is considerable scatter, we observe that the galaxies with larger bias, i.e., the galaxies in over dense regions, tend to have larger A_λ. The observed high A_λ for satellite galaxies is consistent with this trend as these galaxies also have higher bias. These results are are not straight forward to interpret within the context of the tidal stripping of satellite galaxies as was used as an explanation in [135]. Our results suggests that FP residuals of the satellite galaxies can primarily be explained by the luminosity dependence of the FP. However, when computing the correlation functions, we fit FP only to the satellite galaxies and within the satellite sample it is possible that stronger tidal stripping in denser environment can imprint some environment dependence on FP residuals leading to stronger correlations. In either case, our results suggest that environment plays a dominant role in determining the FP residuals of a galaxy. More detailed interpretation of these results will require a study using the realistic galaxy simulations to understand the relative importance of various processes involved in determining the galaxy sizes. We leave such a study for the future work.

Correlations with IA

As discussed in section 7.2, intrinsic alignments (IA) of galaxies coupled with the projection effects can lead to the scatter over the FP and the correlations between the FP residuals and the density field. The model in used in fitting the cross correlations between the FP residuals

[2]Both pipelines were developed by S. Singh

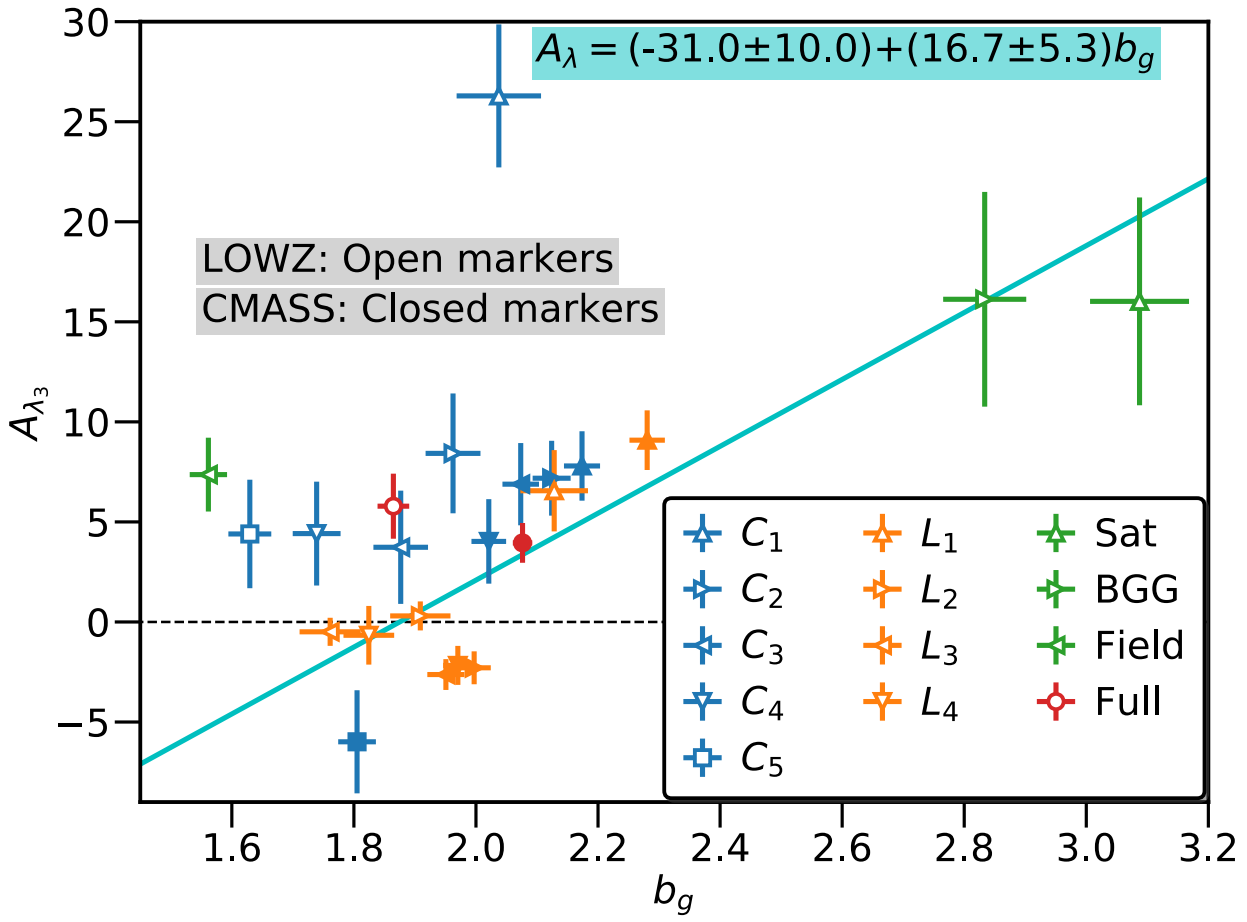


Figure 7.7: Environment dependence (as characterized by linear galaxy bias) of FP residuals. In more biased (overdense) environments, scatter about FP has stronger correlations with the environment.

and the galaxy density field accounts for this effects and indeed if IA is the only cause of correlations of FP residuals, we expect $A_\lambda \sim A_I/2$.

In figure 7.8 we present the comparison of the the intrinsic alignments amplitude (a detailed analysis of IA measurements was presented in [241] and in this work we repeat those measurements using BOSS DR12 data) and the FP residual amplitude A_λ derived from the cross correlations with the density field. For both LOWZ and CMASS samples we observe positive correlations between A_I and A_λ, samples with stronger IA also showing stronger correlations for FP. This is also consistent with the observations that IA and FP correlations have similar environment dependence (see previous section and [241]). However, our measurements are inconsistent with the model predictions of $A_\lambda = A_{IA}/2$. In addition to considerable scatter in the measurements, the best fit linear models we obtained (not shown) deviated significantly from the model.

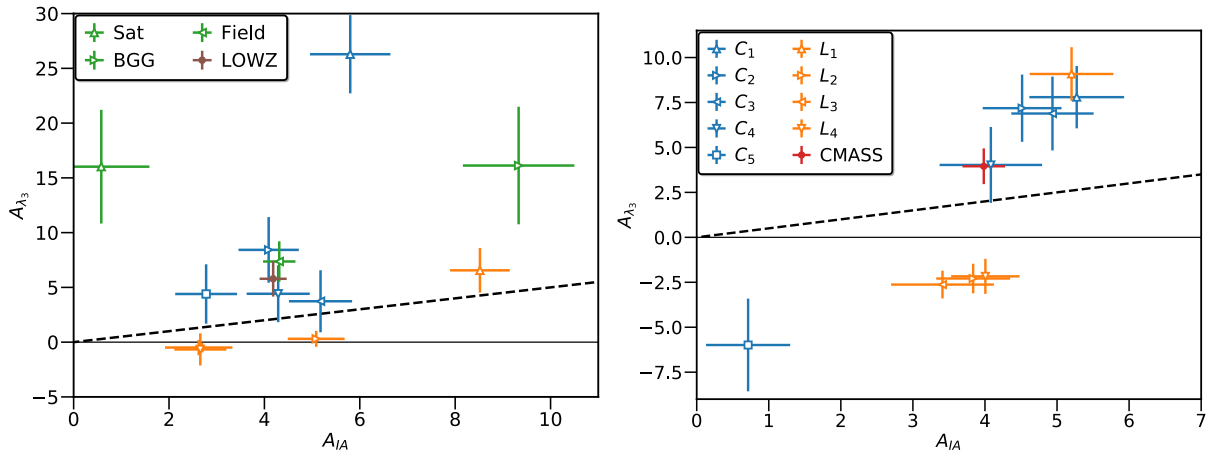


Figure 7.8: Comparison of the intrinsic alignments amplitude (A_{IA}) measured using galaxy shear and the amplitude of galaxy size correlations measured using the fundamental plane residuals (A_{λ_3}), for different subsamples of LOWZ (a) and CMASS (b). Different colors represent different splits and different markers represent different subsamples (labels for color and luminosity subsamples are consistent across two panels). Under the model assumed in section 7.2, size correlations are caused by intrinsic alignments in conjunction with projection effects and we expect $A_{\lambda_3} \propto A_{IA}/2$ (shown by dashed black line). Solid cyan line shows the best fit linear model with parameter as shown in the figures. Data prefers $A_{\lambda_3} \propto 4A_{IA}$, which suggests that in addition to projection effects, galaxy sizes themselves are affected by the tidal field, such that $\lambda_3 \propto \nabla^2\phi$, with similar constants as the IA model.

Our results suggest that the galaxy size correlation (as measured by FP) include contributions beyond the effects of intrinsic alignments and the projection effects. These contributions can come from physical processes such as stronger feedback in over dense regions or observational systematics affecting the estimation of size, magnitude and velocity dispersions of galaxies (eg. errors in PSF modeling, small fiber size used in spectroscopic measurements). The interpretation of these results is further complicated by the fact that IA also depends on the shape measurement methods. As shown in [244], de Vaucouleurs shape results in $\sim 15-20\%$ larger IA amplitude though de Vaucouleurs shapes were also shown to be affected by systematics in the same study because of which we use the re-gaussianzation shapes to measure the IA in this chapter. Hence, it is difficult to fully explain the origin and the magnitude of the size correlation amplitudes and a detailed exploration of the physical origin of these effects will require further study with realistic simulations.

Multipoles

In figure 7.9 we show the measurements of the multipoles of the galaxy clustering and galaxy-FP cross correlation measurements. The monopole and quadrupole of the galaxy clustering are consistent with the expectations from the redshift space distortion measurements.

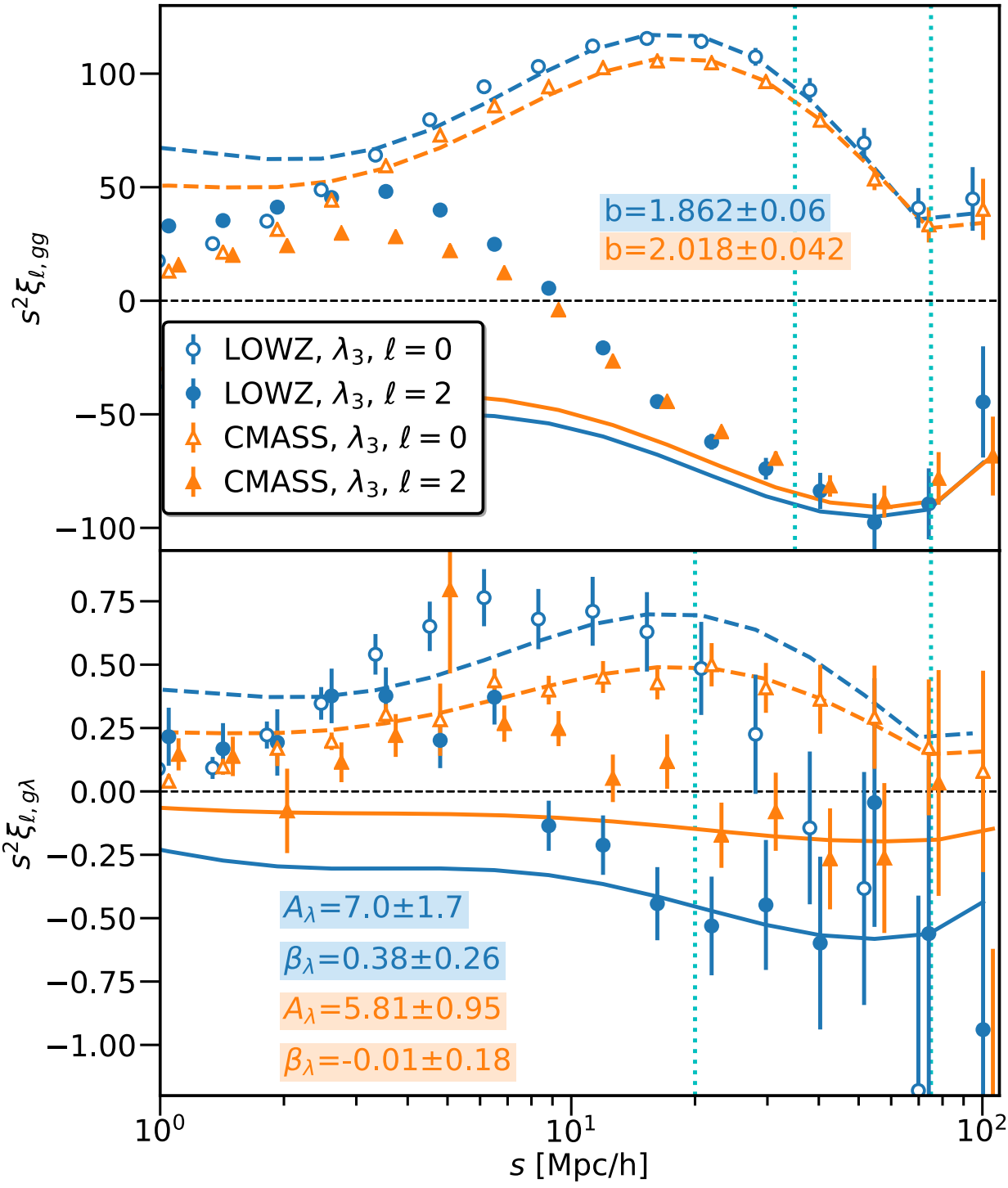


Figure 7.9: Measurements of the multipoles of galaxy clustering (upper panels) and galaxy-λ (lower panels) cross correlation functions for LOWZ (blue) and CMASS (orange) samples. Open points and dashed line shows the monopole and the best fit model for the monopole ($\ell = 0$) while closed points and lines show the same for quadrupole ($\ell = 2$). Vertical cyan lines mark the range over which the model was fit.

Here we fit the simple Kaiser model to fit both the monopole and quadrupole moments and we will present more detailed analysis of galaxy clustering multipoles in the next section. The galaxy bias obtained from these fits is consistent with the values obtained from the projected correlation functions. Since the model does not include the effects of non-linear corrections, it does not fit data well on small scales and hence we only use $30 < s < 70h^{-1}$Mpc to fit the galaxy-galaxy correlation functions and we further fix the growth rate to the value expected from our fiducial cosmology $f = 0.665(0.77)$ for LOWZ (CMASS). Since our focus in this section is to study the anisotropy of the galaxy-FP correlation function, we prefer to use a simpler model with few parameters over a more detailed RSD model presented in the next section.

In the lower panel, we present the measurements and the fits for the galaxy-FP cross correlation function. We detect both the monopole and the quadrupole moments of the correlation functions, pointing to significant line of sight anisotropy in these measurements. Some level of anisotropy is expected as the the galaxy positions are measured in redshift space and the FP residuals are weighted with galaxy density field in the redshift space. To estimate the anisotropy contributions from the FP residuals, we fit for the anisotropy factor, β_λ (see eq. (7.25)) (galaxy anisotropy β_g is obtained from clustering). The best fit values of β_λ we obtain are consistent with zero, contrary to the expectations from the model which predicts $\beta_\lambda = -3$ ($\beta_\lambda = -3$ predicts positive quadrupole moments at large scales). These conclusions are not changed even if we fit with growth rate f as a free parameter, if we vary the minimum scale used in the fits ($r_{p,\rm min} = 20$ or $40h^{-1}$Mpc) and even if we include the hexadecapole measurements.

While our measurements appear to rule out the influence of IA on the FP residuals, we note here that this is only true within the model we assumed in this work. Both IA and FP residuals are also weighted by the galaxy density field, which is measured in the redshift space and introduces higher order terms which can have significant contributions to the measured correlation functions. Furthermore, these higher order terms also contain the line of sight anisotropy terms which can in principle affect the β_λ constraints. Modeling these higher order terms accurately is out of the scope of this work and can be attempted in a future work.

IA effects on RSD

In this section we present the measurements of the galaxy power spectrum multipoles in Fourier space. In figure 7.10, we show the measured monopole $P_0(k)$, quadrupole $P_2(k)$, and hexadecapole $P_4(k)$ of the LOWZ NGC and CMASS NGC galaxies, using the FFT-based galaxy power spectrum estimator described in section 7.2. We then fit the RSD model presented in section 7.2 to the measured multipoles and find that the power spectrum multipoles are accurately modeled, down to scales of $k = 0.4h$Mpc^{-1}, in agreement with [105]. The fits to the SGC galaxies are not shown in the figure, but we also find a good agreement between the model and the SGC samples. We include the hexadecapole $P_4(k)$ because this improves RSD constraints significantly, as reported earlier in [37], [96], and [105]. In our fits, we set the minimum wavenumber $k_{\rm max}$ to 0.05 and 0.02 hMpc^{-1}, respectively for LOWZ and

CMASS, in order to minimize any large-scale effects of the window function. As described in section 7.2, we fix the AP distortion parameters to their fiducial values and constrain 11 free parameters, of which two are primarily of our interests: the growth rate f and the amplitude of matter fluctuations σ_8. Fitting this RSD model to the BOSS DR12 multipole measurements, we obtain a tight constraint on the growth of structure, and more detail can be found in [282].

Fundamental plane cuts

In this section we fit FP to the LOWZ and CMASS samples, for NGC and SGC regions separately, and then split each galaxy sample into two subsamples according to the sign of the FP residuals, with the mean FP residual subtracted from each sample, following [180]. Samples with positive (negative) FP residuals correspond to galaxies larger (smaller) than the FP-predicted size. We then fit the galaxy power spectrum model to each of the two subsamples, constraining 11 free parameters in the RSD model presented in section 7.2.

In the analysis, we consider different types of FP residuals: FP fit with $N_z = 1$ (λ_1), FP fit with $N_z = 1$ in narrow redshift bins ($\lambda_{1,zb}$), FP fit only dependent on the surface brightness with the velocity dispersion measurements ignored (λ_1^I; still with with $N_z = 1$), FP fit equivalent to the one in [180] (λ_1^M), FP fit with $N_z = 3$ (λ_3). Figure 7.11 shows how the model fits to each of the two subsamples of the LOWZ NGC galaxies. The subsample with positive FP residuals (henceforth called the "positive" subsample) has higher galaxy bias than the subsample with negative FP residuals, in agreement with the results in [180]. We also find that disregarding the velocity dispersion in the FP definition causes a larger deviation in the galaxy bias between subsamples. Some deviation in the galaxy bias is expected from the correlations between the FP and the galaxy properties as discussed in section 7.4. Since the "positive" sample preferentially selects brighter galaxies, it is expected to have larger bias.

The monopole $P_0(k)$ and the quadrupole $P_2(k)$ scale as $(b_1\sigma_8)^2$ and $b_1 f\sigma_8^2$, respectively. Hence, we can roughly estimate the ratio of $f\sigma_8$ and $b_1\sigma_8$ values between two subsamples of the FP fit from the ratios of the monopoles and quadrupoles. The square rooted ratio of the monopoles scales as $b_1\sigma_8$; for the LOWZ NGC sample, the subsample with positive λ_1^I has this ratio $\approx 10\%$ larger than the subsample with negative λ_1^I. Adding the velocity dispersion term to the FP definition reduces such deviation to $\approx 6\%$. We can also take the quadrupole ratio between two samples and divide it by the square rooted ratio of their monopoles to remove the bias dependence. This quantity roughly determine the ratio of $f\sigma_8$ values between two samples. Figure 7.12 plots the ratios 1) between the full LOWZ NGC sample and the subsample with positive FP residuals (blue) and 2) between the full LOWZ NGC sample and the subsample with negative FP residuals (red). The measured ratios are well within 1σ from each other, particularly on the scales where non-linear, small effects are not important, suggesting that measurements of the difference in RSD constraints between the FP fit subsamples are not statistically significant.

RSD model fits in figure 7.13 are in agreement with the above observation. For each galaxy sample, we fit the RSD model to the multipoles of positive and negative (FP fit)

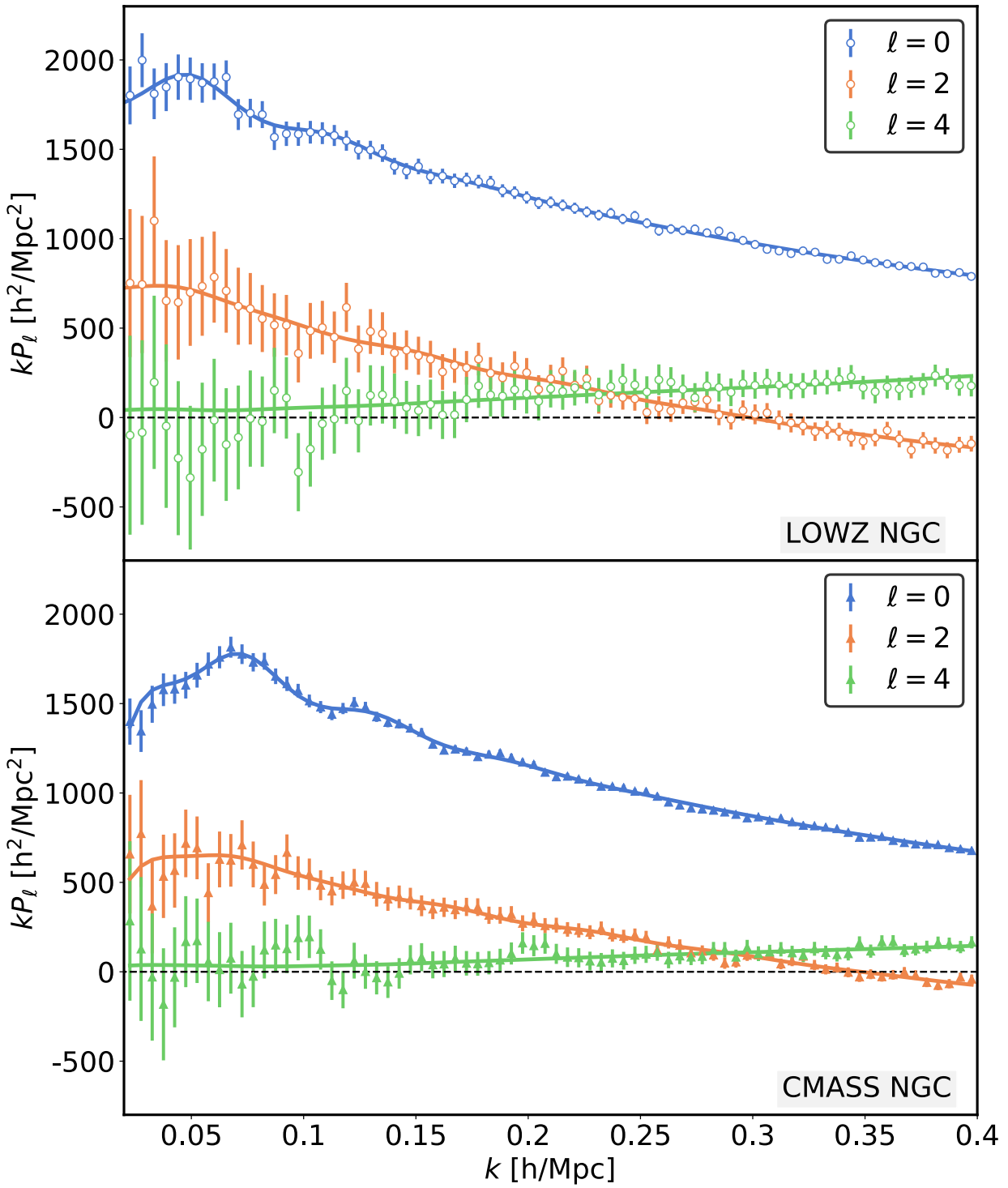


Figure 7.10: The measured galaxy power spectrum multipoles in Fourier space (data points) and the best-fit theory curves (solid lines) for LOWZ NGC (upper panel) and CMASS NGC (lower panel) samples. We fit the model to the monopole (blue), quadrupole (orange), and hexadecapole (green), over the wavenumber range $k = 0.05 - 0.4$ and $0.02 - 0.4h\text{Mpc}^{-1}$ for LOWZ and CMASS galaxies, respectively. Multipoles are accurately modelled, down to $k = 0.4h\text{Mpc}^{-1}$. Although not shown in the figure, we also find an excellent model fit to the SGC samples.

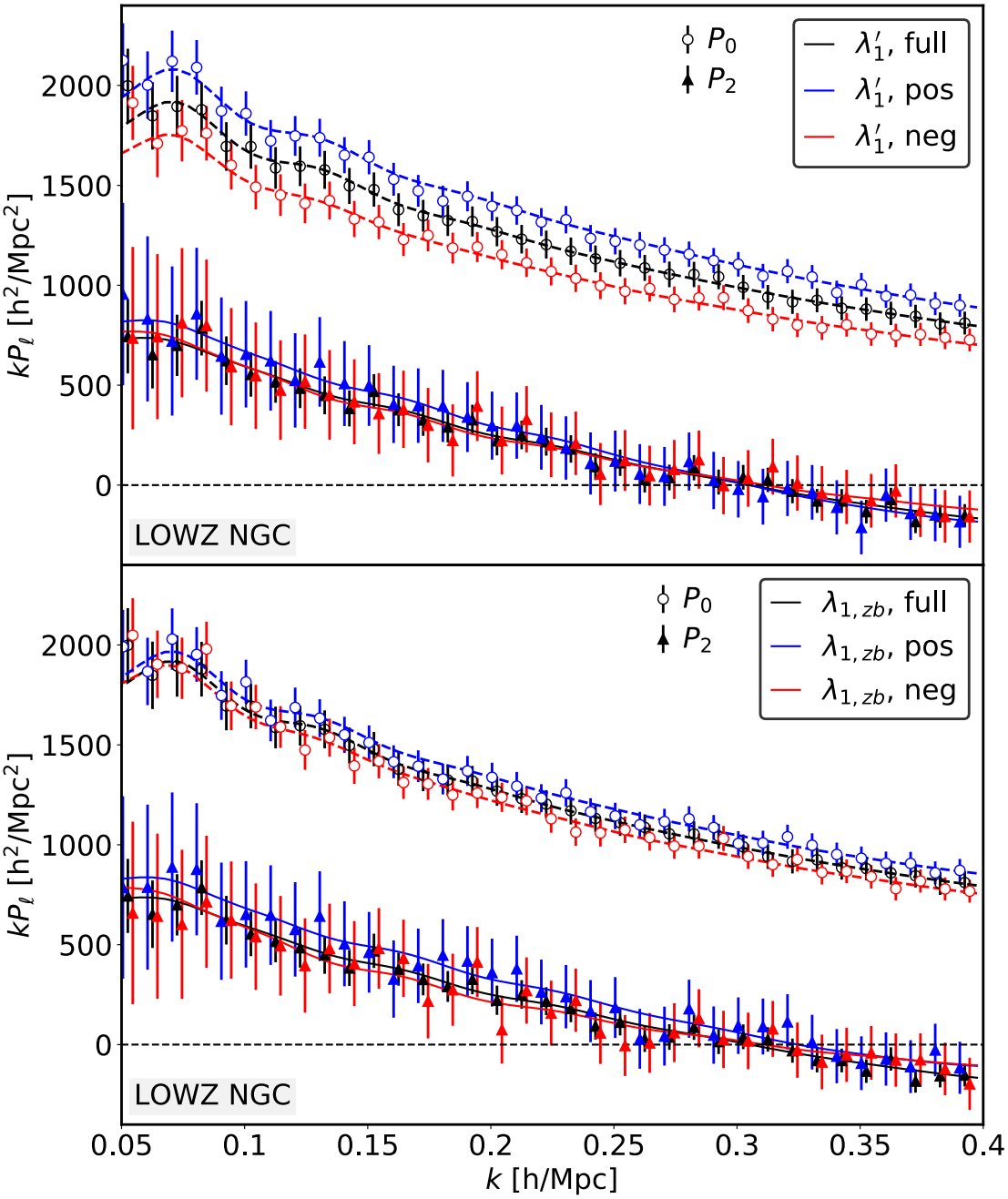


Figure 7.11: Multipole measurements of the LOWZ NGC subsamples with positive (blue) and negative (red) FP residuals, with different types of FP fits. Multipoles of the full LOWZ NGC sample (black) are also provided as a reference, and open circular points and closed triangular points display the monopoles and quadrupoles, respectively. We show the measured multipoles (data points) and the best-fit theory curves (solid lines) for positive and negative subsamples with two different FP definitions: FP fit without velocity dispersion (λ_1^I; upper panel) and FP fit in narrow redshift bins ($\lambda_{1,zb}$; lower panel). The RSD model fits well to all subsamples, and other subsamples with different FP definitions similarly have good model fits, although not shown in the figure. The monopoles of positive and negative subsamples clearly have different amplitudes, suggesting the difference in their galaxy biases.

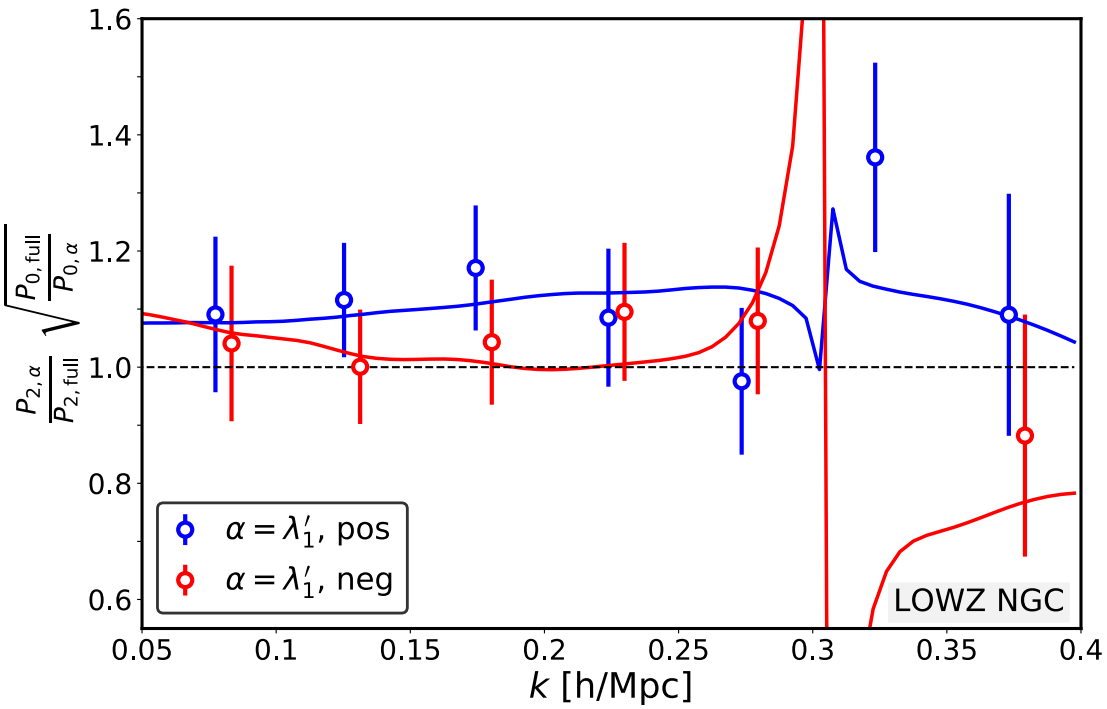


Figure 7.12: A rough estimate of the ratio of $f\sigma_8$ values between the full LOWZ NGC sample and FP fit subsamples, as quantified by the quadrupole ratio $(P_{2,\alpha})/(P_{2,\mathrm{full}})$ divided by $\sqrt{(P_{0,\alpha})/(P_{0,\mathrm{full}})}$, where α denotes the FP fit subsample. Open circular points show the measured ratios, while the dotted lines show the ratios from the best-fit theory curves. The difference in the measured ratios of the two subsamples is not statistically significant, especially on the scales where non-linear, small effects are not important.

subsamples, for both NGC and SGC regions, and then measure the difference in $f\sigma_8$ and $b_1\sigma_8$ constraints between two subsamples. With the convention that the value in the positive subsample is subtracted from the value in the negative subsample, we consider different types of the FP residuals (λ_1^I, λ_1, $\lambda_{1,zb}$, λ_3, and λ_1^M). In figure 7.13, FP fit subsamples selected in the same redshift range and sky region are marked in the same color, and such samples are are not independent, all correlated with one another. First, we find that all positive subsamples have larger galaxy biases than negative subsamples, as expected from figure 7.11, thereby resulting in the sign of $\Delta b_1\sigma_8$ positive in all FP fits. On the contrary, the signs of $\Delta f\sigma_8$ measurements are not consistent across all samples and all within 1σ of the measurements, statistically consistent with the null result. This suggests that there is no evidence of significant bias in RSD measurements due to IA, in tension with the results from [180], which showed a consistent offset in Δf between the FP fit subsamples. In section 7.2, the model expects that $\Delta b_g + \frac{1}{3}\Delta f \approx 0$, where b_g and f are galaxy bias and growth rate parameters, respectively. However, as shown in figure 7.13, the null model ($\Delta f \sim 0$) is

favored over such theory prediction by the data points. Similarly, figure 16 and 18 in [180] also show a deviation between the model and measured data points.

[180] defined the ratios of the measurements to theoretically predicted values in the following way,

$$\frac{\text{Observed}}{\text{Theory}} = \frac{\Sigma_i \left[\frac{B_i^{\text{Theory}}}{\sigma_i}\right]^2 \frac{B_i^{\text{Observed}}}{B_i^{\text{Theory}}}}{\Sigma_i \left[\frac{B_i^{\text{Theory}}}{\sigma_i}\right]^2} \pm \frac{1}{\sqrt{\Sigma_i \left[\frac{B_i^{\text{Theory}}}{\sigma_i}\right]^2}}, \tag{7.47}$$

and presented that Obs/Theory $= 0.61 \pm 0.26$. Assuming the model in [180], we repeat the analysis and obtain Obs/Theory $= -0.05 \pm 0.33$, consistent with zero.

In figure 7.13, we use the MultiDark-Patchy mock catalogues to estimate the size of the error bars of RSD constraints for the FP fit subsamples. In this work, we consider four different galaxy samples: LOWZ NGC, LOWZ SGC, CMASS NGC, and CMASS SGC. For each sample, we take 100 PATCHY mocks and separate each mock into two subsamples, depending on their stellar masses, as the stellar mass is correlated with the luminosity, which in turn is correlated with the FP residuals and also correlated with the galaxy alignment strength [134, 241]). Consequently, we have 100 subsamples with their stellar masses larger than the mean stellar mass of the sample and 100 subsamples with their stellar masses smaller than the mean. For each galaxy sample, we then perform fits to the measured multipoles of PATCHY mocks, and the best-fitting parameters for each of the 200 subsamples are obtained by maximum a posterior (MAP) estimation using the LBFGS algorithm. Subsequently, we measure the standard deviation of the best-fitting values of $f\sigma_8$ and $b_1\sigma_8$ and obtain the propagated error for $\Delta f\sigma_8$ and $\Delta b_1\sigma_8$.

Luminosity/color cuts

In fig 7.14, with luminosity cuts as described in section 7.3, we divide the LOWZ NGC sample into four subsamples, with L_1 being brightest and L_4 being faintest, and show how the RSD model fits the galaxy power spectrum multipoles of all subsamples. Similarly the lower panel plots the quadrupole measurements of all subsamples. The measured multipoles and the best-fit theory model of other samples, such as LOWZ SGC, CMASS NGC, and CMASS SGC, show a similar trend and therefore not shown in the figure.

Fig 7.15 presents the monopole and quadrupole measurements of the LOWZ NGC subsamples based on the color cut, with colors going redder from C_5 to C_1. A redder subsample is shown to have a higher bias. However, the quadrupole measurement of C_1 clearly deviates from those of other subsamples, and fitting the RSD model to all color subsamples, we find that its $f\sigma_8$ constraint is significantly different from $f\sigma_8$ measurements of other subsamples.

In figure 7.17, we show the correlations between the constraints on $f\sigma_8(z_{\text{eff}})$, the product of f and σ_8, each evaluated at the effective redshift of each sample, and the rescaled FP residual amplitude A_{λ_3}, computed at the fiducial σ_8. We divide each $f\sigma_8(z_{\text{eff}})$ measurement

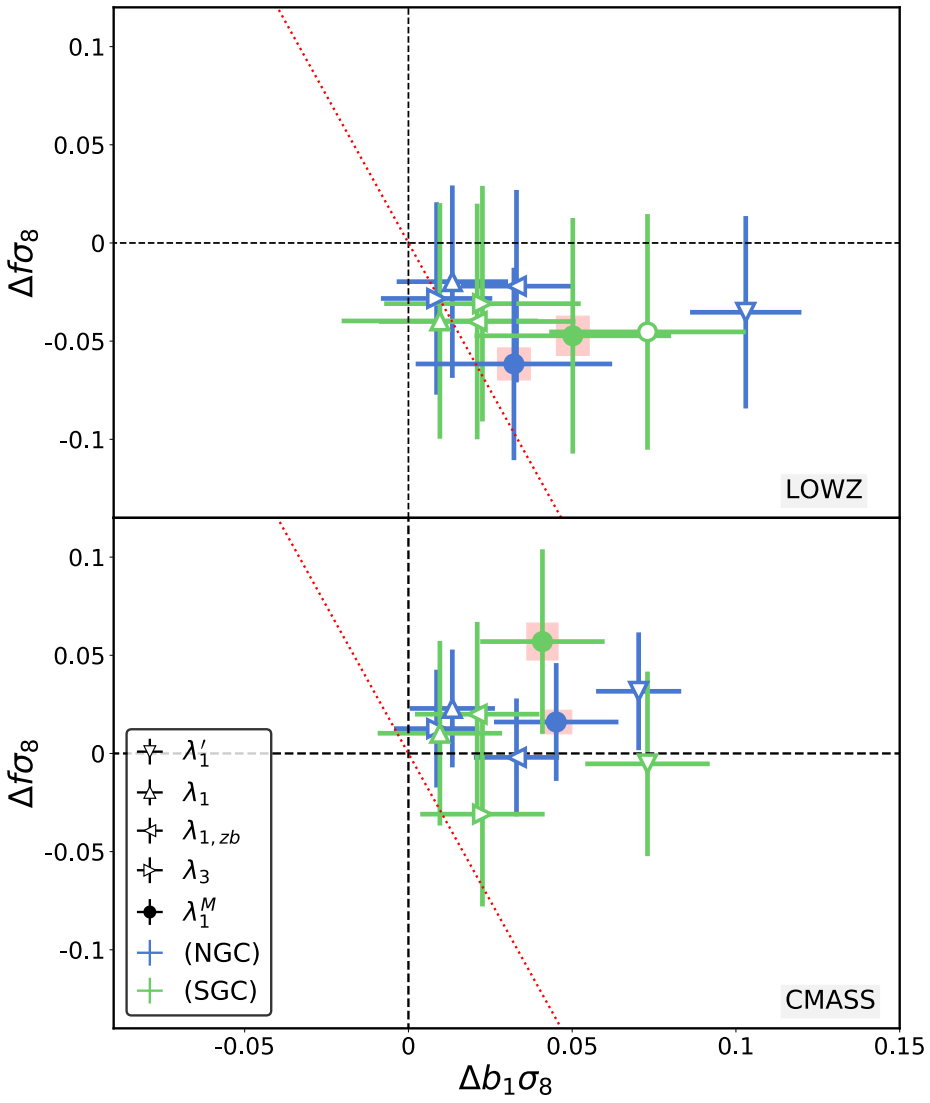


Figure 7.13: The measured difference in $f\sigma_8$ and $b_1\sigma_8$ between positive and negative subsamples, with different definitions of FP fits (indicated with different markers), for the LOWZ (upper panel) and CMASS (lower panel) galaxies in NGC (blue) and SGC (green) regions. Data points highlighted with red rectangles are using the FP measurements from Martens et al. (2018). Note that samples in the same color are all correlated with one another; we take the same galaxy sample and split them into the FP fit subsamples based on different FP definitions. The differences in $f\sigma_8$ values are all statistically consistent with the null results, while the signs of $\Delta b_1\sigma_8$ is consistently positive. For both LOWZ and CMASS, the theory prediction (red dotted lines) from section 7.2, $\Delta f \approx -3\Delta b_1$, is not favored by the data points relative to the null model ($\Delta f \sim 0$).

with the predictions from the Planck 2015 data, to present $f\sigma_8$ measurements independent of the effective redshifts of our samples. To make direct comparisons with the model in section 7.2, we rescale A_{λ_3} and then multiply it with a factor $\gamma/f_{\rm fid}$, where γ is the response parameter as described in section 7.2. The measured γ values (assuming $\lambda_3^{\rm NS}$, where 'NS' denotes fitting NGC and SGC seprately.) are shown in figure 7.16. [3]

As shown in figure 7.17, we also find that there is only a weak evolution of the growth rate measurements with the FP residual amplitude. This agrees with the conclusion in section 7.4 that no significant bias in RSD measurements due to IA is evident. We quantify this correlations by fitting a linear relation between the growth rate f and the FP residual amplitude, and the following are the best-fit models: $(f\sigma_8/f\sigma_{8,\rm fid}) = (-0.05\pm0.02)-(-0.36\pm 0.14)\cdot[-\gamma A_{\lambda_3}\zeta/f_{\rm fid}]$ for NGC and $(f\sigma_8/f\sigma_{8,\rm fid}) = (-0.07\pm0.03)+(-0.10\pm0.04)\cdot[-\gamma A_{\lambda_3}\zeta/f_{\rm fid}]$ for SGC, clearly in tension with the model in section 7.2, which predicted that the growth rate f is larger for a larger FP amplitude; we find such correlations in the opposite direction for both NGC (left panel) and SGC (right panel) samples. Moreover, the slope of this fit is largely driven by the LOWZ C_1 outlier. Without the LOWZ C_1 sample, the slope would be closer to zero. Comparing the measurements to the null detection of IA effects on RSD measurements (shown by brown lines), we obtain χ^2 values of 27.6 and 14.6 for 18 NGC and 18 SGC sub-samples, respectively. For the NGC samples, this corresponds to probability-to-exceed (PTE) of $\approx 7\%$, which indicates that the measurements are consistent with the null result.

7.5 Conclusions

In this work we have presented the estimations of the FP of the BOSS galaxies and the dependence of FP residuals on the galaxy redshift, environment and luminosity. We show that the redshift evolution of the FP observed in earlier works [135, 228] is primarily driven by the redshift evolution of the surface brightness of the galaxies and correcting for this redshift implies that the FP is primarily a relation between the size, luminosity and velocity dispersion of the galaxies. The FP residuals are also strongly correlated with the luminosity of the galaxies and the luminosity evolution of the FP is primarily responsible for the apparent environment dependence of the FP residuals as was first detected by [135]. We also show that the FP residuals are correlated with the observational systematics, most notably the goodness of the galaxy profile fits, the PSF flux and in the case of CMASS sample the stellar density weights.

In section 7.4 we presented the measurements and analysis of the correlations between the FP residuals and the galaxy density field, $w_{g\lambda}$. FP residuals for BOSS Lowz and CMASS samples show similar correlations with the density field. We showed that these correlations are driven by the correlations between the galaxy properties and the galaxy density field. Galaxy luminosity, size and velocity dispersion are positively correlated with the density

[3] Assuming the FP definition in [180], we get γ = -0.21±0.03, 0.11±0.04, 0.18±0.02, 0.22±0.04 for LOWZ NGC, LOWZ SGC, CMASS NGC, and CMASS SGC respectively.

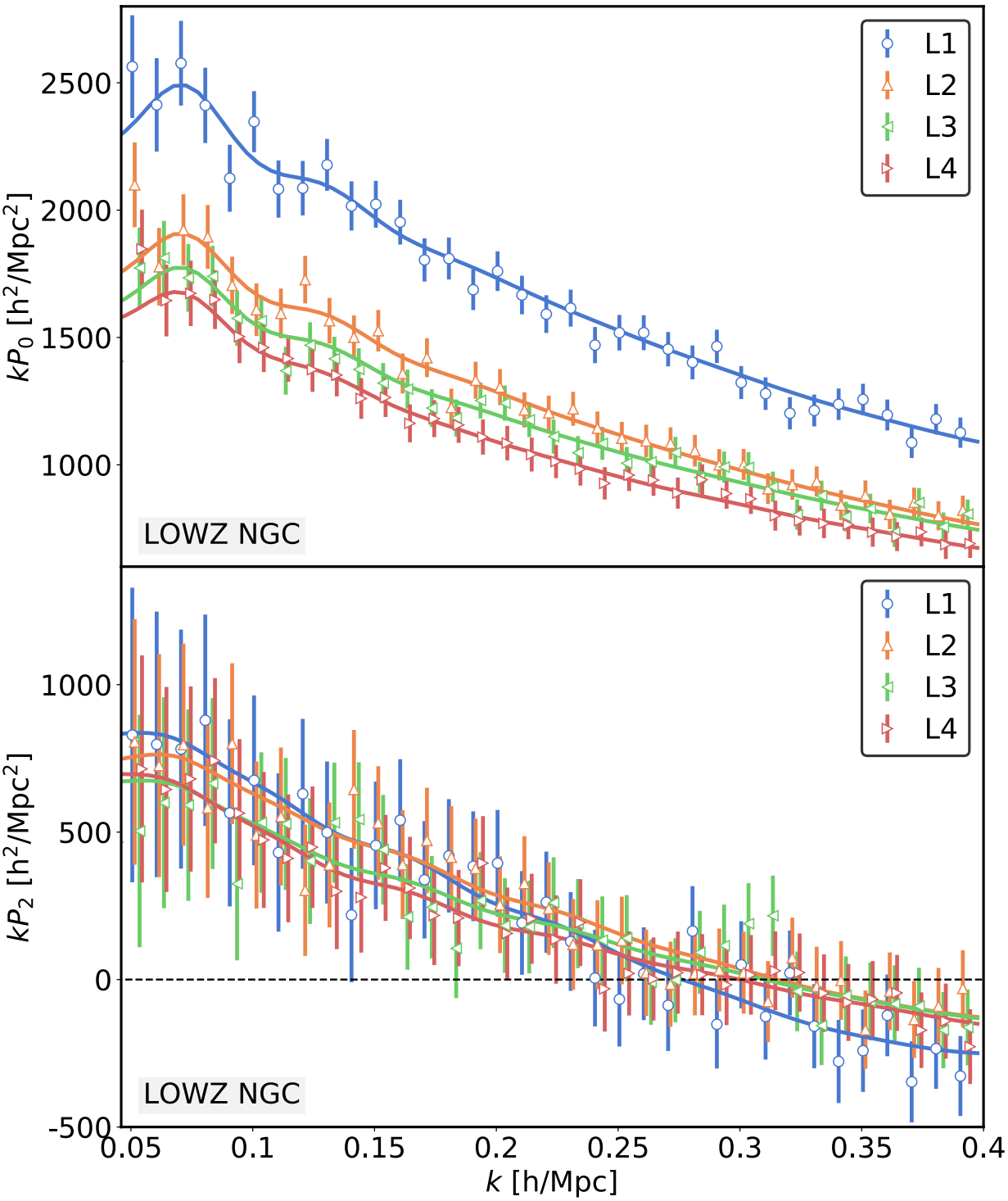


Figure 7.14: The measured galaxy power spectrum multipoles (data points) of four luminosity subsamples, for the LOWZ NGC sample. Solid lines indicate the best-fit theory curves. The monopole measurements (upper panel) show a decreasing trend of bias with luminosity; this suggests that a brighter subsample has a higher galaxy bias. The quadrupole measurements (lower panel) of all luminosity subsamples are within 1 sigma of the quadrupole of the full sample.

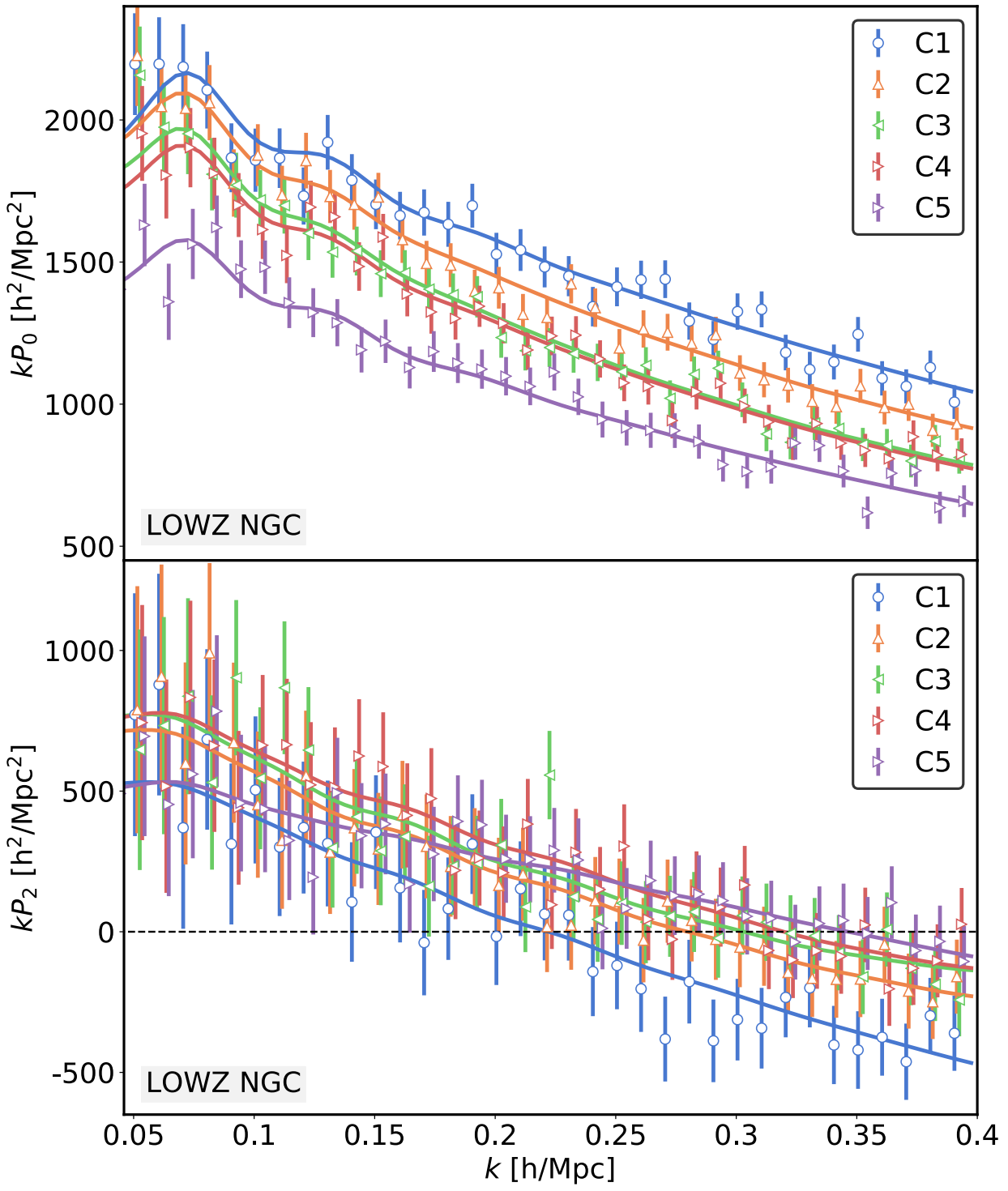


Figure 7.15: Similar to figure 7.14. The multipoles of five color subsamples for LOWZ NGC and the best-fit theory model (solid lines). The monopole measurements (upper panel) show that a redder subsample has a higher bias.

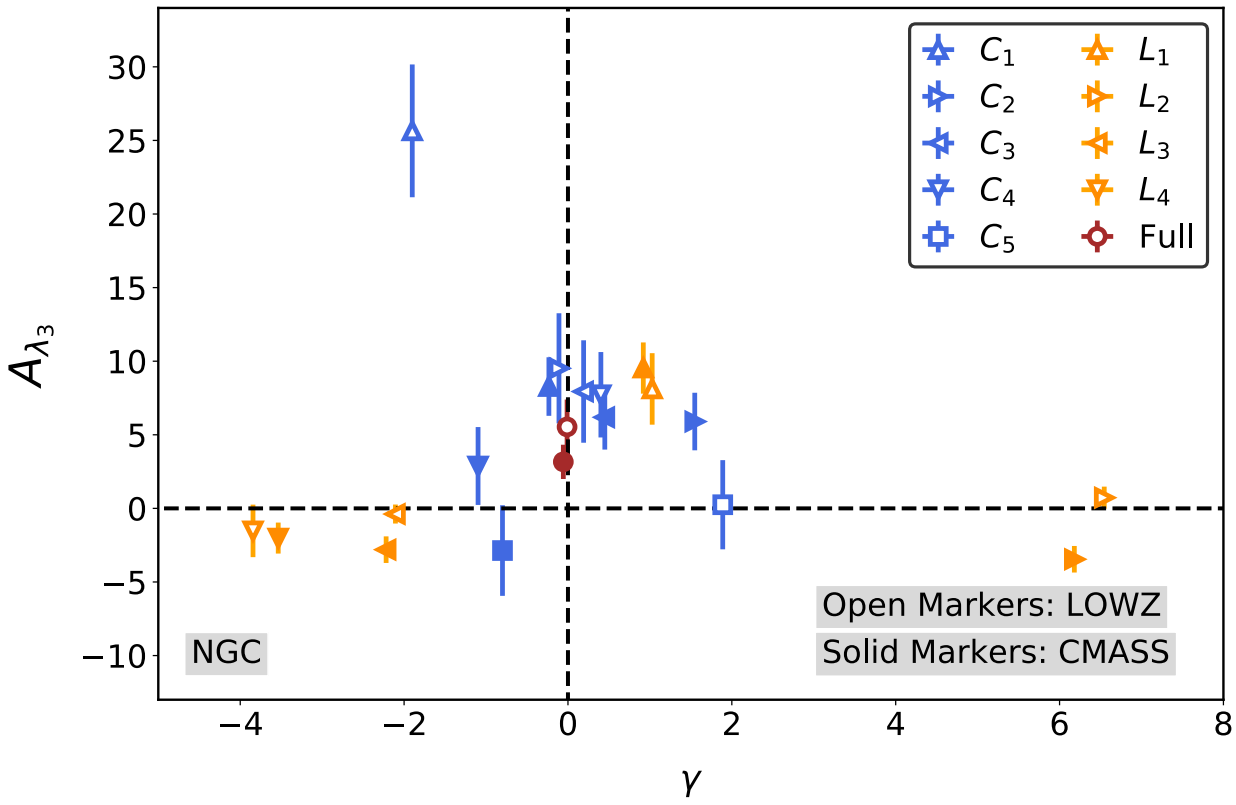


Figure 7.16: Comparison of the FP residuals A_{λ_3} and the selection dependence factor γ for different subsamples of LOWZ NGC and CMASS NGC. Different colors represent different splits, and different markers represent different subsamples. No correlation between A_{λ_3} and γ is evident in the figure.

field while the surface brightness is negatively correlated. The negative correlation of surface brightness is a non-trivial result and is potentially important for HOD modeling, relating galaxies to the halos as well for the modeling of galaxy bias as function of redshift.

We also studied the dependence of the $w_{g\lambda}$ as function of galaxy environment, luminosity as well as color. Brighter galaxies show strong positive correlations between the FP residuals and the density field while the correlation amplitude is lower for fainter samples with the lowest luminosity sample showing negative correlations. Similar trends are also observed for the color splits, with strong positive correlations for the red galaxies with lower correlations for bluer samples. Combining all the samples together, we show that there is strong correlation between the galaxy bias and the amplitude of $w_{g\lambda}$. This implies that the galaxies in more over dense regions show stronger correlations between the FP/size residuals and the environment.

We also compare the amplitude of $w_{g\lambda}$ measurements from model fits with amplitude of the intrinsic alignment (IA) measurements. Since our model assumes that the size correlations are sourced by the effects of IA in three dimensions projected onto the two dimensional plane of the sky, the amplitudes of IA and $w_{g\lambda}$ measurements measurements are expected to be the

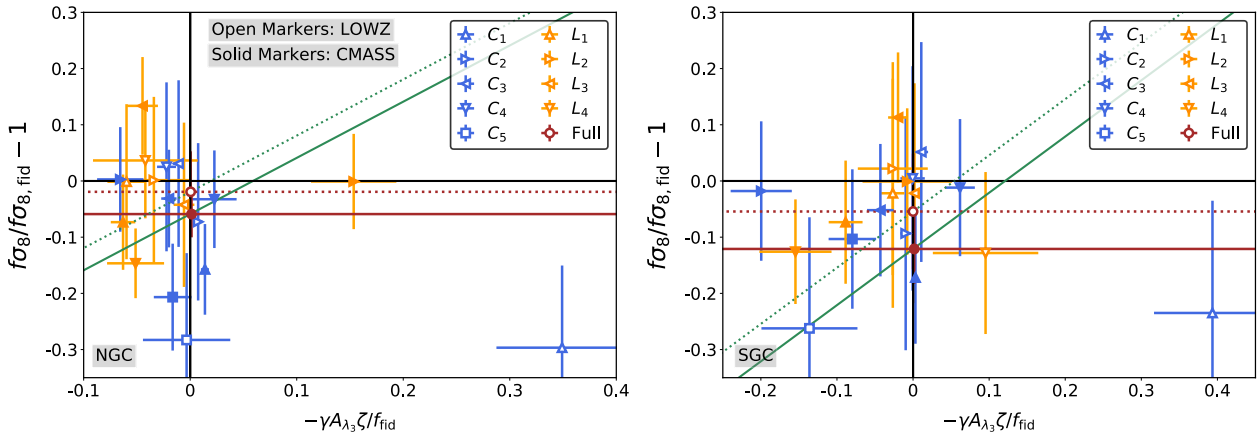


Figure 7.17: Comparison of the RSD measurements of the growth of structure and the rescaled FP residual amplitude, for different subsamples of LOWZ (open markers) and CMASS (solid markers) galaxies in NGC (left panel) and SGC (right panel) regions. Each measurement of $f\sigma_8(z_{\mathrm{eff}})$ is divided by the predicted value assuming the fiducial Planck 2015 cosmology (y-axis), and the amplitude of galaxy size correlations measured using the FP residuals (A_{λ_3}) is rescaled and multiplied by $-\gamma\zeta/f_{\mathrm{fid}}(z_{\mathrm{eff}})$ of each sample (x-axis) so that the expected correlation coefficient between two variables is 1, according to $\Delta f = -\gamma A_\lambda \zeta$ (eq. 7.42). Green lines represent the expected relation between the FP residual and RSD measurements for the full LOWZ (dotted lines) and CMASS (solid lines) sample, and similarly brown lines show the growth of structure measurements for the full LOWZ and CMASS samples. We find that there is only a weak evolution of the growth of structure measurements with the FP residual amplitude, thereby suggesting that no significant bias in RSD measurements due to IA is evident.

same. Though there is considerable scatter, our results are in tension with this prediction. We further tested the model by measuring the multipole moments of the correlation functions, finding again that the measured multipole moments are in tension with the model predictions. We note that our modeling has a limitation as we do not include the density weighting effects. Thus while measurements are in tension with the incomplete model, it is difficult to conclude that the IA does not have any impact on the size correlations as estimated using the FP.

Furthermore, in section 7.4 we presented the correlations between FP residuals and RSD constraints, in particular on the growth rate parameter. Splitting the BOSS LOWZ and CMASS galaxies into subsamples based on FP residuals, we fitted the RSD model to the measured multipoles of each subsample and showed that the differences in RSD constraints, across all subsamples, are statistically consistent with the null result. We hence conclude that there is no evidence of significant impact in RSD measurements due to IA, contrary to conclusions drawn by [180]. Moreover, the RSD measurements of the BOSS samples split by luminosity and color further strengthens this argument; despite some scatter, we find only a weak evolution of the RSD constraints with the FP residual amplitude, and in comparison with the model from [180], we find this effect in the opposite direction. This suggest that

there can be other overwhelming effects that impact the FP, and it can be difficult to simply disentangle the relation among FP, IA, and RSD measurements from other effects.

Our work is a step towards improved understanding of FP to use it as a tool for studying galaxies as well as cosmology. The scatter in the FP is similar to the scatter in the ellipticities of the galaxies, suggesting that it can be developed as a probe of gravitational lensing and beyond with similar potency as the galaxy ellipticities (shear). There are notable impediments to such applications as the FP is dependent on the galaxy properties, environment, selection effects and photometry errors. A deeper understanding of the impact of galaxy physics on the FP will require a similar study as ours using realistic cosmological simulations that we plan to pursue in near future. A detailed understanding of observational effects will require image simulations as were performed for the context of galaxy shape measurements [e.g. 178]. A more detailed modeling for the FP correlations with density field also needs to be developed to capture the information from small scales that we ignored in this work as well as to capture the impact of density weighting terms. The promise of galaxy sizes for cosmological applications and the upcoming large dataset from DESI, LSST, Euclid and WFIRST makes it an opportune moment to study FP and galaxy sizes in general in more detail.

Bibliography

[1] A. Lewis and A. Challinor. *CAMB: Code for Anisotropies in the Microwave Background*. 2019. URL: https://camb.info/.

[2] K. N. Abazajian et al. "The Seventh Data Release of the Sloan Digital Sky Survey". In: *ApJS* 182 (June 2009), pp. 543–558. DOI: 10.1088/0067-0049/182/2/543. arXiv: 0812.0649.

[3] Kevork N Abazajian and Scott Dodelson. "Neutrino mass and dark energy from weak lensing". In: *Physical Review Letters* 91.4 (2003), p. 041301.

[4] Kevork N Abazajian et al. "CMB-S4 science book". In: *arXiv preprint arXiv:1610.02743* (2016).

[5] Kevork N Abazajian et al. "Neutrino physics from the cosmic microwave background and large scale structure". In: *Astroparticle Physics* 63 (2015), pp. 66–80.

[6] Paul A Abell et al. *Lsst science book, version 2.0*. arXiv:0912.0201, 2009.

[7] PAR Ade et al. "Planck 2015 results. XV. Gravitational lensing". In: *Astronomy & Astrophysics/Astronomie et Astrophysique* 594.A15 (2016).

[8] Peter Ade et al. "The Simons Observatory: Science goals and forecasts". In: *JCAP* 02 (2019), p. 056. DOI: 10.1088/1475-7516/2019/02/056. arXiv: 1808.07445 [astro-ph.CO].

[9] Peter AR Ade et al. "Planck 2015 results-XV. Gravitational lensing". In: *Astronomy & Astrophysics* 594 (2016), A15.

[10] Amir Aghamousa et al. "The DESI experiment part I: science, targeting, and survey design". In: *arXiv preprint arXiv:1611.00036* (2016).

[11] N. Aghanim et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astron. Astrophys.* 641 (2020). [Erratum: Astron.Astrophys. 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

[12] C. P. Ahn et al. "The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey". In: *ApJS* 203, 21 (Dec. 2012), p. 21. DOI: 10.1088/0067-0049/203/2/21. arXiv: 1207.7137 [astro-ph.IM].

[13] H. Aihara et al. "The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III". In: *ApJS* 193, 29 (Apr. 2011), p. 29. DOI: 10.1088/0067-0049/193/2/29. arXiv: 1101.1559 [astro-ph.IM].

[14] K. Akitsu, M. Takada, and Y. Li. "Large-scale tidal effect on redshift-space power spectrum in a finite-volume survey". In: *Phys.Rev.D* 95.8, 083522 (Apr. 2017), p. 083522. DOI: 10.1103/PhysRevD.95.083522. arXiv: 1611.04723.

[15] S. Alam et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". In: *MNRAS* 470 (Sept. 2017), pp. 2617–2652. DOI: 10.1093/mnras/stx721. arXiv: 1607.03155.

[16] S. Alam et al. "The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III". In: *ApJS* 219, 12 (July 2015), p. 12. DOI: 10.1088/0067-0049/219/1/12. arXiv: 1501.00963 [astro-ph.IM].

[17] Andreas Albrecht et al. "Report of the dark energy task force". In: *arXiv preprint astro-ph/0609591* (2006).

[18] Rupert Allison et al. "Towards a cosmological neutrino mass detection". In: *Physical Review D* 92.12 (2015), p. 123535.

[19] L. Amendola et al. "Cosmology and Fundamental Physics with the Euclid Satellite". In: *Living Reviews in Relativity* 16, 6 (Sept. 2013), p. 6. DOI: 10.12942/lrr-2013-6. arXiv: 1206.1225.

[20] L. Anderson et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples". In: *MNRAS* 441 (June 2014), pp. 24–62. DOI: 10.1093/mnras/stu523. arXiv: 1312.4877.

[21] V. Assassi, M. Simonović, and M. Zaldarriaga. "Efficient evaluation of angular power spectra and bispectra". In: *J. Cosmology Astropart. Phys.* 11, 054 (Nov. 2017), p. 054. DOI: 10.1088/1475-7516/2017/11/054. arXiv: 1705.05022.

[22] T. Baldauf et al. "Algorithm for the direct reconstruction of the dark matter correlation function from weak lensing and galaxy clustering". In: *Phys.Rev.D* 81.6, 063531 (Mar. 2010), p. 063531. DOI: 10.1103/PhysRevD.81.063531. arXiv: 0911.4973 [astro-ph.CO].

[23] Tobias Baldauf et al. "LSS constraints with controlled theoretical uncertainties". In: *arXiv preprint arXiv:1602.00674* (2016).

[24] Arka Banerjee et al. "Tests of neutrino and dark radiation models from galaxy and CMB surveys". In: *Journal of Cosmology and Astroparticle Physics* 2018.01 (2018), p. 022.

[25] E. M. Barboza Jr. and J. S. Alcaniz. "A parametric model for dark energy". In: *Phys. Lett. B* 666 (2008), pp. 415–419. DOI: 10.1016/j.physletb.2008.08.012. arXiv: 0805.1713 [astro-ph].

[26] A. Barreira, E. Krause, and F. Schmidt. "Complete super-sample lensing covariance in the response approach". In: *J. Cosmology Astropart. Phys.* 6, 015 (June 2018), p. 015. DOI: 10.1088/1475-7516/2018/06/015. arXiv: 1711.07467.

[27] A. Barreira and F. Schmidt. "Response approach to the matter power spectrum covariance". In: *J. Cosmology Astropart. Phys.* 11, 051 (Nov. 2017), p. 051. DOI: 10.1088/1475-7516/2017/11/051. arXiv: 1705.01092.

[28] Bruce A Bassett et al. "Fisher matrix preloaded—Fisher4Cast". In: *International Journal of Modern Physics D* 20.13 (2011), pp. 2559–2598.

[29] F. Baumgarten and C.-H. Chuang. "Robustness of the covariance matrix for galaxy clustering measurements". In: *MNRAS* 480 (Oct. 2018), pp. 2535–2543. DOI: 10.1093/mnras/sty1971. arXiv: 1802.04462.

[30] Rachel Bean and Olivier Doré. "Probing for Dark Energy Perturbations using the CMB and Large Scale Structure?" In: *AIP Conference Proceedings*. Vol. 743. 1. AIP. 2004, pp. 88–95.

[31] Mariangela Bernardi et al. "Early-Type Galaxies in the Sloan Digital Sky Survey. III. The Fundamental Plane". In: *The Astronomical Journal* 125.4 (Apr. 2003), pp. 1866–1881. DOI: 10.1086/367794. arXiv: astro-ph/0301626 [astro-ph].

[32] G. Bertin and M. Lombardi. "Looking at the Fundamental Plane through a Gravitational Lens". In: *ApJ* 648 (Sept. 2006), pp. L17–L20. DOI: 10.1086/507298. eprint: astro-ph/0606672.

[33] D. Bertolini et al. "Non-Gaussian covariance of the matter power spectrum in the effective field theory of large scale structure". In: *Phys.Rev.D* 93.12, 123505 (June 2016), p. 123505. DOI: 10.1103/PhysRevD.93.123505. arXiv: 1512.07630.

[34] M. Betancourt. "A Conceptual Introduction to Hamiltonian Monte Carlo". In: *ArXiv e-prints* (Jan. 2017). arXiv: 1701.02434 [stat.ME].

[35] Michael Betancourt. "A conceptual introduction to Hamiltonian Monte Carlo". In: *arXiv preprint arXiv:1701.02434* (2017).

[36] F. Beutler et al. "The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant". In: *MNRAS* 416 (Oct. 2011), pp. 3017–3032. DOI: 10.1111/j.1365-2966.2011.19250.x. arXiv: 1106.3366.

[37] Florian Beutler et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Anisotropic galaxy clustering in Fourier-space". In: *Mon. Not. Roy. Astron. Soc.* 466.2 (2017), pp. 2242–2260. DOI: 10.1093/mnras/stw3298. arXiv: 1607.03150 [astro-ph.CO].

[38] Florian Beutler, Emanuele Castorina, and Pierre Zhang. "Interpreting measurements of the anisotropic galaxy power spectrum". In: *JCAP* 03 (2019), p. 040. DOI: 10.1088/1475-7516/2019/03/040. arXiv: 1810.05051 [astro-ph.CO].

[39] Florian Beutler and Patrick McDonald. "Unified galaxy power spectrum measurements from 6dFGS, BOSS, and eBOSS". In: *JCAP* 11 (2021), p. 031. DOI: 10.1088/1475-7516/2021/11/031. arXiv: 2106.06324 [astro-ph.CO].

[40] Florian Beutler et al. "The 6dF Galaxy Survey: z ≈ 0 measurements of the growth rate and σ_8". In: *Monthly Notices of the Royal Astronomical Society* 423.4 (2012), pp. 3430–3444.

[41] Davide Bianchi et al. "Measuring line-of-sight dependent Fourier-space clustering using FFTs". In: *Mon. Not. Roy. Astron. Soc.* 453.1 (2015), pp. L11–L15. DOI: 10.1093/mnrasl/slv090. arXiv: 1505.05341 [astro-ph.CO].

[42] C. Blake, P. Carter, and J. Koda. "Power spectrum multipoles on the curved sky: an application to the 6-degree Field Galaxy Survey". In: *MNRAS* 479 (Oct. 2018), pp. 5168–5183. DOI: 10.1093/mnras/sty1814. arXiv: 1801.04969.

[43] C. Blake et al. "The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9". In: *MNRAS* 415 (Aug. 2011), pp. 2876–2891. DOI: 10.1111/j.1365-2966.2011.18903.x. arXiv: 1104.2948.

[44] M. R. Blanton et al. "An Efficient Targeting Strategy for Multiobject Spectrograph Surveys: the Sloan Digital Sky Survey "Tiling" Algorithm". In: *AJ* 125 (Apr. 2003), pp. 2276–2286. DOI: 10.1086/344761. eprint: arXiv:astro-ph/0105535.

[45] J. Blazek, M. McQuinn, and U. Seljak. "Testing the tidal alignment model of galaxy intrinsic alignment". In: *J. Cosmology Astropart. Phys.* 5, 010 (May 2011), p. 10. DOI: 10.1088/1475-7516/2011/05/010. arXiv: 1101.4017 [astro-ph.CO].

[46] Jonathan Blazek, Zvonimir Vlah, and Uroš Seljak. "Tidal alignment of galaxies". In: *J. Cosmology Astropart. Phys.* 2015.8, 015 (Aug. 2015), p. 015. DOI: 10.1088/1475-7516/2015/08/015. arXiv: 1504.02510 [astro-ph.CO].

[47] LE Bleem et al. "A measurement of the correlation of galaxy surveys with CMB lensing convergence maps from the South Pole Telescope". In: *The Astrophysical Journal Letters* 753.1 (2012), p. L9.

[48] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational Inference: A Review for Statisticians". In: *CoRR* abs/1601.00670 (2016). arXiv: 1601.00670. URL: http://arxiv.org/abs/1601.00670.

[49] L. Blot et al. "Comparing approximate methods for mock catalogues and covariance matrices II: Power spectrum multipoles". In: *ArXiv e-prints* (June 2018). arXiv: 1806.09497.

[50] A. S. Bolton et al. "Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey". In: *AJ* 144, 144 (Nov. 2012), p. 144. DOI: 10.1088/0004-6256/144/5/144. arXiv: 1207.7326 [astro-ph.CO].

[51] J. R. Bond and A. S. Szalay. "The collisionless damping of density fluctuations in an expanding universe". In: *APJ* 274 (Nov. 1983), pp. 443–468. DOI: 10.1086/161460.

[52] Camille Bonvin et al. "Dipolar modulation in the size of galaxies: the effect of Doppler magnification". In: *Monthly Notices of the Royal Astronomical Society* 472.4 (Dec. 2017), pp. 3936–3951. DOI: 10.1093/mnras/stx2049. arXiv: 1610.05946 [astro-ph.CO].

[53] V. Bonvin et al. "H0LiCOW - V. New COSMOGRAIL time delays of HE 0435-1223: H_0 to 3.8 per cent precision from strong lensing in a flat ΛCDM model". In: *MNRAS* 465 (Mar. 2017), pp. 4914–4930. DOI: 10.1093/mnras/stw3006. arXiv: 1607.01790.

[54] Aoife Boyle and Eiichiro Komatsu. "Deconstructing the neutrino mass constraint from galaxy redshift surveys". In: *Journal of Cosmology and Astroparticle Physics* 2018.03 (2018), p. 035.

[55] Thejs Brinckmann et al. "The promising future of a robust cosmological neutrino mass measurement". In: *Journal of Cosmology and Astroparticle Physics* 2019.01 (2019), p. 059.

[56] Bob Carpenter et al. "Stan: A probabilistic programming language". In: *Journal of statistical software* 76.1 (2017).

[57] Michael J. Chapman et al. "The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the growth rate of structure from the small-scale clustering of the luminous red galaxy sample". In: (June 2021). arXiv: 2106.14961 [astro-ph.CO].

[58] Shi-Fan Chen, Zvonimir Vlah, and Martin White. "A new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO". In: *JCAP* 02.02 (2022), p. 008. DOI: 10.1088/1475-7516/2022/02/008. arXiv: 2110.05530 [astro-ph.CO].

[59] Shi-Fan Chen et al. "Cosmological Analysis of Three-Dimensional BOSS Galaxy Clustering and Planck CMB Lensing Cross Correlations via Lagrangian Perturbation Theory". In: *arXiv preprint arXiv:2204.10392* (2022).

[60] Shu-Fan Chen, Hayden Lee, and Cora Dvorkin. "Precise and accurate cosmology with CMB×LSS power spectra and bispectra". In: *JCAP* 05 (2021), p. 030. DOI: 10.1088/1475-7516/2021/05/030. arXiv: 2103.01229 [astro-ph.CO].

[61] Michel Chevallier and David Polarski. "Accelerating universes with scaling dark matter". In: *Int. J. Mod. Phys. D* 10 (2001), pp. 213–224. DOI: 10.1142/S0218271801000822. arXiv: gr-qc/0009008.

[62] J. D. Cohn. "Power spectrum and correlation function errors: Poisson vs. Gaussian shot noise". In: *New A* 11 (Jan. 2006), pp. 226–239. DOI: 10.1016/j.newast.2005.08.002. eprint: astro-ph/0503285.

[63] Eoin Ó. Colgáin, M. M. Sheikh-Jabbari, and Lu Yin. "Can dark energy be dynamical?" In: *Phys. Rev. D* 104.2 (2021), p. 023510. DOI: 10.1103/PhysRevD.104.023510. arXiv: 2104.01930 [astro-ph.CO].

[64] DESI Collaboration. *DESI: The Dark Energy Spectroscopic Instrument*. URL: http://desi.lbl.gov/.

[65] A. Cooray and W. Hu. "Power Spectrum Covariance of Weak Gravitational Lensing". In: *APJ* 554 (June 2001), pp. 56–66. DOI: 10.1086/321376. eprint: astro-ph/0012087.

[66] Asantha Cooray and Wayne Hu. "Imprint of reionization on the cosmic microwave background bispectrum". In: *The Astrophysical Journal* 534.2 (2000), p. 533.

[67] Martın Crocce et al. "Galaxy clustering, photometric redshifts and diagnosis of systematics in the DES Science Verification data". In: *Monthly Notices of the Royal Astronomical Society* 455.4 (2015), pp. 4301–4324.

[68] Guido D'Amico et al. "The Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure". In: *JCAP* 05 (2020), p. 005. DOI: 10.1088/1475-7516/2020/05/005. arXiv: 1909.05271 [astro-ph.CO].

[69] Omar Darwish et al. "The Atacama Cosmology Telescope: a CMB lensing mass map over 2100 square degrees of sky and its cross-correlation with BOSS-CMASS galaxies". In: *MNRAS* 500.2 (Jan. 2021), pp. 2250–2263. DOI: 10.1093/mnras/staa3438. arXiv: 2004.01139 [astro-ph.CO].

[70] K. S. Dawson et al. "The Baryon Oscillation Spectroscopic Survey of SDSS-III". In: *AJ* 145, 10 (Jan. 2013), p. 10. DOI: 10.1088/0004-6256/145/1/10. arXiv: 1208.0022 [astro-ph.CO].

[71] R. R. de Carvalho and S. Djorgovski. "Systematic Differences between the Field and Cluster Elliptical Galaxies". In: *The Astrophysical Journal* 389 (Apr. 1992), p. L49. DOI: 10.1086/186346.

[72] DESI Collaboration et al. "The DESI Experiment Part I: Science,Targeting, and Survey Design". In: *arXiv e-prints*, arXiv:1611.00036 (Oct. 2016), arXiv:1611.00036. arXiv: 1611.00036 [astro-ph.IM].

[73] Adji Bousso Dieng et al. "Variational Inference via \chi Upper Bound Minimization". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*. 2017, pp. 2729–2738. URL: http://papers.nips.cc/paper/6866-variational-inference-via-chi-upper-bound-minimization.

[74] S. Djorgovski and Marc Davis. "Fundamental Properties of Elliptical Galaxies". In: *The Astrophysical Journal* 313 (Feb. 1987), p. 59. DOI: 10.1086/164948.

[75] S. Dodelson and M. D. Schneider. "The effect of covariance estimator error on cosmological parameter constraints". In: *Phys.Rev.D* 88.6, 063537 (Sept. 2013), p. 063537. DOI: 10.1103/PhysRevD.88.063537. arXiv: 1304.2593 [astro-ph.CO].

[76] O. Doré et al. "WFIRST Science Investigation Team "Cosmology with the High Latitude Survey" Annual Report 2017". In: *ArXiv e-prints* (Apr. 2018). arXiv: 1804.03628.

[77] Alan Dressler et al. "Spectroscopy and Photometry of Elliptical Galaxies. I. New Distance Estimator". In: *The Astrophysical Journal* 313 (Feb. 1987), p. 42. DOI: 10.1086/164947.

[78] G. Efstathiou. "Constraining the equation of state of the universe from distant type Ia supernovae and cosmic microwave background anisotropies". In: *Mon. Not. Roy. Astron. Soc.* 310 (1999), pp. 842–850. DOI: 10.1046/j.1365-8711.1999.02997.x. arXiv: astro-ph/9904356.

[79] Tim Eifler et al. "Cosmology with the Roman Space Telescope - synergies with the Rubin Observatory Legacy Survey of Space and Time". In: *MNRAS* (Mar. 2021). DOI: 10.1093/mnras/stab533. arXiv: 2004.04702 [astro-ph.CO].

[80] D. J. Eisenstein and W. Hu. "Baryonic Features in the Matter Transfer Function". In: *APJ* 496 (Mar. 1998), p. 605.

[81] D. J. Eisenstein et al. "Spectroscopic Target Selection for the Sloan Digital Sky Survey: The Luminous Red Galaxy Sample". In: *AJ* 122 (Nov. 2001), pp. 2267–2280.

[82] Daniel J. Eisenstein et al. "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". In: *Astrophys. J.* 633 (2005), pp. 560–574. DOI: 10.1086/466512. arXiv: astro-ph/0501171.

[83] Xiao Fang et al. "Cosmology from Clustering, Cosmic Shear, CMB Lensing, and Cross Correlations: Combining Rubin Observatory and Simons Observatory". In: *ArXiv e-prints* (Aug. 2021). arXiv: 2108.00658 [astro-ph.CO].

[84] H. A. Feldman, N. Kaiser, and J. A. Peacock. "Power-spectrum analysis of three-dimensional redshift surveys". In: *APJ* 426 (May 1994), pp. 23–37. DOI: 10.1086/174036. eprint: arXiv:astro-ph/9304022.

[85] F. Feroz, M. P. Hobson, and M. Bridges. "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics". In: *Mon. Not. Roy. Astron. Soc.* 398 (2009), pp. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437 [astro-ph].

[86] D. Foreman-Mackey et al. "emcee: The MCMC Hammer". In: *PASP* 125 (Mar. 2013), p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[87] Daniel Foreman-Mackey et al. "emcee: The MCMC Hammer". In: *PASP* 125.925 (Mar. 2013), p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[88] M. Fukugita et al. "The Sloan Digital Sky Survey Photometric System". In: *AJ* 111 (Apr. 1996), pp. 1748–+.

[89] Carlos Garcıa-Garcıa et al. "The growth of density perturbations in the last ∼10 billion years from tomographic large-scale structure data". In: *arXiv e-prints*, arXiv:2105.12108 (May 2021), arXiv:2105.12108. arXiv: 2105.12108 [astro-ph.CO].

[90] Stuart Geman and Donald Geman. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6 (1984), pp. 721–741. DOI: 10.1109/TPAMI.1984.4767596.

[91] Martina Gerbino. "Neutrino properties from cosmology". In: *Prospects in Neutrino Physics*. Mar. 2018, pp. 52–52. arXiv: 1803.11545 [astro-ph.CO].

[92] T Giannantonio et al. "CMB lensing tomography with the DES Science Verification galaxies". In: *Monthly Notices of the Royal Astronomical Society* 456.3 (2016), pp. 3213–3244.

[93] J González-Nuevo et al. "H-ATLAS/GAMA: magnification bias tomography. Astrophysical constraints above 1 arcmin". In: *Journal of Cosmology and Astroparticle Physics* 2017.10 (2017), p. 024.

[94] Alexia Gorecki et al. "A new method to improve photometric redshift reconstruction. Applications to the Large Synoptic Survey Telescope". In: *A&A* 561, A128 (Jan. 2014), A128. DOI: 10.1051/0004-6361/201321102. arXiv: 1301.3010 [astro-ph.CO].

[95] H. S. Grasshorn Gebhardt and D. Jeong. "Fast and accurate computation of projected two-point functions". In: *Phys.Rev.D* 97.2, 023504 (Jan. 2018), p. 023504. DOI: 10.1103/PhysRevD.97.023504. arXiv: 1709.02401.

[96] J. N. Grieb et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the Fourier space wedges of the final sample". In: *ArXiv e-prints* (July 2016). arXiv: 1607.03143.

[97] Jan Niklas Grieb et al. "Gaussian covariance matrices for anisotropic galaxy clustering measurements". In: *Mon. Not. Roy. Astron. Soc.* 457.2 (2016), pp. 1577–1592. DOI: 10.1093/mnras/stw065. arXiv: 1509.04293 [astro-ph.CO].

[98] J. E. Gunn et al. "The 2.5 m Telescope of the Sloan Digital Sky Survey". In: *AJ* 131 (Apr. 2006), pp. 2332–2359. DOI: 10.1086/500975. eprint: astro-ph/0602326.

[99] J. E. Gunn et al. "The Sloan Digital Sky Survey Photometric Camera". In: *AJ* 116 (Dec. 1998), pp. 3040–3081.

[100] A. Hall and A. Taylor. "A Bayesian method for combining theoretical and simulated covariance matrices for large-scale structure surveys". In: *ArXiv e-prints* (July 2018). arXiv: 1807.06875.

[101] A. J. S. Hamilton. "Uncorrelated modes of the non-linear power spectrum". In: *MNRAS* 312 (Feb. 2000), pp. 257–284. DOI: 10.1046/j.1365-8711.2000.03071.x. eprint: astro-ph/9905191.

[102] A. J. S. Hamilton, C. D. Rimes, and R. Scoccimarro. "On measuring the covariance matrix of the non-linear power spectrum from simulations". In: *MNRAS* 371 (Sept. 2006), pp. 1188–1204. DOI: 10.1111/j.1365-2966.2006.10709.x. eprint: astro-ph/0511416.

[103] P. Hammad. "Mesure d'ordre α de l'information au sens de Fisher". In: *Revue de Statistique Appliquee* 26 (1978), pp. 73–84.

[104] Nick Hand et al. "An optimal FFT-based anisotropic power spectrum estimator". In: *JCAP* 1707.07 (2017), p. 002. DOI: 10.1088/1475-7516/2017/07/002. arXiv: 1704.02357 [astro-ph.CO].

[105] Nick Hand et al. "Extending the modeling of the anisotropic galaxy power spectrum to $k = 0.4\ h\mathrm{Mpc}^{-1}$". In: *JCAP* 1710.10 (2017), p. 009. DOI: 10.1088/1475-7516/2017/10/009. arXiv: 1706.02362 [astro-ph.CO].

[106] Nick Hand et al. "nbodykit: an open-source, massively parallel toolkit for large-scale structure". In: *The Astronomical Journal* 156.4 (2018), p. 160.

[107] Qianjun Hang et al. "Galaxy clustering in the DESI Legacy Survey and its imprint on the CMB". In: *MNRAS* 501.1 (Feb. 2021), pp. 1481–1498. DOI: 10.1093/mnras/staa3738. arXiv: 2010.00466 [astro-ph.CO].

[108] Steen Hannestad, Huitzu Tu, and Yvonne YY Wong. "Measuring neutrino masses and dark energy with weak lensing tomography". In: *Journal of Cosmology and Astroparticle Physics* 2006.06 (2006), p. 025.

[109] Duncan Hanson. *quicklens*. 2018. URL: https://github.com/dhanson/quicklens.

[110] Yuichi Harikane et al. "GOLDRUSH. II. Clustering of galaxies at z 4–6 revealed with the half-million dropouts over the 100 deg2 area corresponding to 1 Gpc3". In: *Publications of the Astronomical Society of Japan* 70.SP1 (2018), S11.

[111] J. Harnois-Déraps and U.-L. Pen. "Non-Gaussian error bars in galaxy surveys - I". In: *MNRAS* 423 (July 2012), pp. 2288–2307. DOI: 10.1111/j.1365-2966.2012.21039.x. arXiv: 1109.5746.

[112] J. Hartlap, P. Simon, and P. Schneider. "Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix". In: *A&A* 464 (Mar. 2007), pp. 399–404. DOI: 10.1051/0004-6361:20066170. eprint: astro-ph/0608064.

[113] Jan Hartlap et al. "A bias in cosmic shear from galaxy selection: results from ray-tracing simulations". In: *Astronomy & Astrophysics* 528 (2011), A51.

[114] A. F. Heavens et al. "Generalized Fisher matrices". In: *Mon. Not. Roy. Astron. Soc.* 445.2 (2014), pp. 1687–1693. DOI: 10.1093/mnras/stu1866. arXiv: 1404.2854 [astro-ph.CO].

[115] C. Hikage et al. "Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data". In: *ArXiv e-prints* (Sept. 2018). arXiv: 1809.09148.

[116] H Hildebrandt et al. "KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing". In: *Monthly Notices of the Royal Astronomical Society* 465.2 (2016), pp. 1454–1498.

[117] C. Hirata and U. Seljak. "Shear calibration biases in weak-lensing surveys". In: *MNRAS* 343 (Aug. 2003), pp. 459–480.

[118] Christopher M Hirata and Uroš Seljak. "Reconstruction of lensing from the cosmic microwave background polarization". In: *Physical Review D* 68.8 (2003), p. 083002.

[119] Christopher M. Hirata. "Tidal alignments as a contaminant of redshift space distortions". In: *MNRAS* 399.2 (Oct. 2009), pp. 1074–1087. DOI: 10.1111/j.1365-2966.2009.15353.x. arXiv: 0903.4929 [astro-ph.CO].

[120] Matthew D Hoffman, Andrew Gelman, et al. "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1593–1623.

[121] D. W. Hogg et al. "A Photometricity and Extinction Monitor at the Apache Point Observatory". In: *AJ* 122 (Oct. 2001), pp. 2129–2138.

[122] C. Howlett and W. J. Percival. "Galaxy two-point covariance matrix estimation for next generation surveys". In: *MNRAS* 472 (Dec. 2017), pp. 4935–4952. DOI: 10.1093/mnras/stx2342. arXiv: 1709.03057.

[123] http://camb.readthedocs.io/en/latest/. In: (2021).

[124] https://github.com/dhanson/quicklens. In: (2017).

[125] W. Hu and A. V. Kravtsov. "Sample Variance Considerations for Cluster Surveys". In: *APJ* 584 (Feb. 2003), pp. 702–715. DOI: 10.1086/345846. eprint: astro-ph/0203169.

[126] Wayne Hu and Daniel J Eisenstein. "Small-scale perturbations in a general mixed dark matter cosmology". In: *The Astrophysical Journal* 498.2 (1998), p. 497.

[127] Wayne Hu and Takemi Okamoto. "Mass reconstruction with cosmic microwave background polarization". In: *The Astrophysical Journal* 574.2 (2002), p. 566.

[128] Eric M. Huff and Genevieve J. Graves. "Magnificent Magnification: Exploiting the Other Half of the Lensing Signal". In: *arXiv e-prints*, arXiv:1111.1070 (Nov. 2011), arXiv:1111.1070. arXiv: 1111.1070 [astro-ph.CO].

[129] Aapo Hyvärinen. "Estimation of Non-Normalized Statistical Models by Score Matching". In: *Journal of Machine Learning Research* 6 (2005), pp. 695–709. URL: http://www.jmlr.org/papers/v6/hyvarinen05a.html.

[130] Mikhail M. Ivanov. "Cosmological constraints from the power spectrum of eBOSS emission line galaxies". In: *Phys. Rev. D* 104.10 (2021), p. 103514. DOI: 10.1103/PhysRevD.104.103514. arXiv: 2106.12580 [astro-ph.CO].

[131] Mikhail M. Ivanov, Marko Simonović, and Matias Zaldarriaga. "Cosmological Parameters from the BOSS Galaxy Power Spectrum". In: *JCAP* 05 (2020), p. 042. DOI: 10.1088/1475-7516/2020/05/042. arXiv: 1909.05277 [astro-ph.CO].

[132] Ž. Ivezić et al. "SDSS data management and photometric quality assessment". In: *Astronomische Nachrichten* 325 (2004), pp. 583–589.

[133] Harvinder Kaur Jassal, J. S. Bagla, and T. Padmanabhan. "Observational constraints on low redshift evolution of dark energy: How consistent are different observations?" In: *Phys. Rev. D* 72 (2005), p. 103503. DOI: 10.1103/PhysRevD.72.103503. arXiv: astro-ph/0506748.

[134] B. Joachimi et al. "Constraints on intrinsic alignment contamination of weak lensing surveys using the MegaZ-LRG sample". In: *A&A* 527, A26 (Mar. 2011), A26. DOI: 10.1051/0004-6361/201015621. arXiv: 1008.3491 [astro-ph.CO].

[135] Benjamin Joachimi, Sukhdeep Singh, and Rachel Mandelbaum. "Detection of spatial correlations of Fundamental Plane residuals, and cosmological implications". In: *MNRAS* 454.1 (Nov. 2015), pp. 478–488. DOI: 10.1093/mnras/stv1962. arXiv: 1504.02662 [astro-ph.CO].

[136] S. Joudaki et al. "KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering". In: *MNRAS* 474 (Mar. 2018), pp. 4894–4924. DOI: 10.1093/mnras/stx2820. arXiv: 1707.06627.

[137] Shahab Joudaki and Manoj Kaplinghat. "Dark energy and neutrino masses from future measurements of the expansion history and growth of structure". In: *Physical Review D* 86.2 (2012), p. 023526.

[138] N. Kaiser. "Clustering in real space and in redshift space". In: *MNRAS* 227 (July 1987), pp. 1–21.

[139] N. Kaiser, G. Squires, and T. Broadhurst. "A Method for Weak Lensing Observations". In: *ApJ* 449 (Aug. 1995), pp. 460–+.

[140] Nick Kaiser. "Weak gravitational lensing of distant galaxies". In: *The Astrophysical Journal* 388 (1992), pp. 272–286.

[141] M. Kaplinghat, L. Knox, and Y. Song. "Determining Neutrino Mass from the Cosmic Microwave Background Alone". In: *Physical Review Letters* 91.24 (Dec. 2003), pp. 241301–+.

[142] Manoj Kaplinghat, Lloyd Knox, and Yong-Seon Song. "Determining neutrino mass from the cosmic microwave background alone". In: *Physical Review Letters* 91.24 (2003), p. 241301.

[143] Lavrentios Kazantzidis and Leandros Perivolaropoulos. "Evolution of the $f\sigma_8$ tension with the Planck15/ΛCDM determination and implications for modified gravity theories". In: *Phys. Rev. D* 97.10 (2018), p. 103503. DOI: 10.1103/PhysRevD.97.103503. arXiv: 1803.01337 [astro-ph.CO].

[144] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[145] Ellie Kitanidis and Martin White. "Cross-correlation of Planck CMB lensing with DESI-like LRGs". In: *MNRAS* 501.4 (Mar. 2021), pp. 6181–6198. DOI: 10.1093/mnras/staa3927. arXiv: 2010.04698 [astro-ph.CO].

[146] F.-S. Kitaura et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release". In: *MNRAS* 456 (Mar. 2016), pp. 4156–4173. DOI: 10.1093/mnras/stv2826. arXiv: 1509.06400.

[147] Lloyd Knox, Yong-Seon Song, and J Anthony Tyson. "Distance-redshift and growth-redshift relations as two windows on acceleration and gravitation: Dark energy or new gravity?" In: *Physical Review D* 74.2 (2006), p. 023512.

[148] Yosuke Kobayashi et al. "Full-shape cosmology analysis of the SDSS-III BOSS galaxy power spectrum using an emulator-based halo model: A 5% determination of $\sigma 8$". In: *Phys. Rev. D* 105.8 (2022), p. 083517. DOI: 10.1103/PhysRevD.105.083517. arXiv: 2110.06969 [astro-ph.CO].

[149] Arthur Kosowsky, Milos Milosavljevic, and Raul Jimenez. "Efficient cosmological parameter estimation from microwave background anisotropies". In: *Phys. Rev. D* 66 (2002), p. 063007. DOI: 10.1103/PhysRevD.66.063007. arXiv: astro-ph/0206014.

[150] Elisabeth Krause and Tim Eifler. "cosmolike - cosmological likelihood analyses for photometric galaxy surveys". In: *MNRAS* 470.2 (Sept. 2017), pp. 2100–2112. DOI: 10.1093/mnras/stx1261. arXiv: 1601.05779 [astro-ph.CO].

[151] Alex Krolewski, Simone Ferraro, and Martin White. "Cosmological constraints from unWISE and Planck CMB lensing tomography". In: *arXiv e-prints*, arXiv:2105.03421 (May 2021), arXiv:2105.03421. arXiv: 2105.03421 [astro-ph.CO].

[152] Alex Krolewski et al. "unWISE tomography of Planck CMB lensing". In: *J. Cosmology Astropart. Phys.* 2020.5, 047 (May 2020), p. 047. DOI: 10.1088/1475-7516/2020/05/047. arXiv: 1909.07412 [astro-ph.CO].

[153] Alp Kucukelbir et al. "Automatic Differentiation Variational Inference". In: *Journal of Machine Learning Research* 18 (2017), 14:1–14:45. URL: http://jmlr.org/papers/v18/16-107.html.

[154] Solomon Kullback and Richard A Leibler. "On information and sufficiency". In: *The annals of mathematical statistics* 22.1 (1951), pp. 79–86.

[155] F. Lacasa. "Covariance of the galaxy angular power spectrum with the halo model". In: *A&A* 615, A1 (July 2018), A1. DOI: 10.1051/0004-6361/201732343. arXiv: 1711.07372.

[156] S. D. Landy and A. S. Szalay. "Bias and variance of angular correlation functions". In: *APJ* 412 (July 1993), pp. 64–71. DOI: 10.1086/172900.

[157] Johannes U. Lange et al. "Five per cent measurements of the growth rate from simulation-based modelling of redshift-space clustering in BOSS LOWZ". In: *Mon. Not. Roy. Astron. Soc.* 509.2 (2021), pp. 1779–1804. DOI: 10.1093/mnras/stab3111. arXiv: 2101.12261 [astro-ph.CO].

[158] R. Laureijs et al. "Euclid Definition Study Report". In: *arXiv e-prints*, arXiv:1110.3193 (Oct. 2011), arXiv:1110.3193. arXiv: 1110.3193 [astro-ph.CO].

[159] J. Lesgourgues. "The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview". In: *ArXiv e-prints* (Apr. 2011). arXiv: 1104.2932 [astro-ph.IM].

[160] Julien Lesgourgues and Sergio Pastor. "Massive neutrinos and cosmology". In: *Physics Reports* 429.6 (2006), pp. 307–379.

[161] Antony Lewis, Anthony Challinor, and Anthony Lasenby. "Efficient computation of cosmic microwave background anisotropies in closed Friedmann-Robertson-Walker models". In: *The Astrophysical Journal* 538.2 (2000), p. 473.

[162] Y. Li, W. Hu, and M. Takada. "Super-sample covariance in simulations". In: *Phys.Rev.D* 89.8, 083519 (Apr. 2014), p. 083519. DOI: 10.1103/PhysRevD.89.083519. arXiv: 1401.0385.

[163] Y. Li, W. Hu, and M. Takada. "Super-sample signal". In: *Phys.Rev.D* 90.10, 103530 (Nov. 2014), p. 103530. DOI: 10.1103/PhysRevD.90.103530. arXiv: 1408.1081.

[164] Y. Li, M. Schmittfull, and U. Seljak. "Galaxy power-spectrum responses and redshift-space super-sample effect". In: *J. Cosmology Astropart. Phys.* 2, 022 (Feb. 2018), p. 022. DOI: 10.1088/1475-7516/2018/02/022. arXiv: 1711.00018.

[165] Yin Li. *Multiplicatively Convolutional Fast Integral Transforms*. https://github.com/eelregit/mcfit.

[166] Yin Li et al. "Disconnected covariance of 2-point functions in large-scale structure". In: *J. Cosmology Astropart. Phys.* 2019.1, 016 (Jan. 2019), p. 016. DOI: 10.1088/1475-7516/2019/01/016. arXiv: 1811.05714 [astro-ph.CO].

[167] D Nelson Limber. "The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field." In: *The Astrophysical Journal* 117 (1953), p. 134.

[168] D. Nelson Limber. "The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II." In: *APJ* 119 (May 1954), p. 655. DOI: 10.1086/145870.

[169] Wu Lin, Mohammad Emtiyaz Khan, and Mark W. Schmidt. "Fast and Simple Natural-Gradient Variational Inference with Mixture of Exponential-family Approximations". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. 2019, pp. 3992–4002. URL: http://proceedings.mlr.press/v97/lin19b.html.

[170] Eric V. Linder. "Exploring the expansion history of the universe". In: *Phys. Rev. Lett.* 90 (2003), p. 091301. DOI: 10.1103/PhysRevLett.90.091301. arXiv: astro-ph/0208512.

[171] M. Lippich et al. "Comparing approximate methods for mock catalogues and covariance matrices I: correlation function". In: *ArXiv e-prints* (June 2018). arXiv: 1806.09477.

[172] Marilena LoVerde. "Neutrino mass without cosmic variance". In: *Physical Review D* 93.10 (2016), p. 103526.

[173] LSST Dark Energy Science Collaboration. "Large Synoptic Survey Telescope: Dark Energy Science Collaboration". In: *ArXiv e-prints* (Nov. 2012). arXiv: 1211.0310 [astro-ph.CO].

[174] R. Lupton et al. "The SDSS Imaging Pipelines". In: *Astronomical Data Analysis Software and Systems X*. Ed. by F. R. Harnden Jr., F. A. Primini, and H. E. Payne. Vol. 238. Astronomical Society of the Pacific Conference Series. 2001, p. 269. eprint: astro-ph/0101420.

[175] Chung-Pei Ma. "Linear power spectra in cold+ hot dark matter models: Analytical approximations and applications". In: *The Astrophysical Journal* 471.1 (1996), p. 13.

[176] Mathew S Madhavacheril, Nicholas Battaglia, and Hironao Miyatake. "Fundamental physics from future weak-lensing calibrated Sunyaev-Zel'dovich galaxy cluster counts". In: *Physical Review D* 96.10 (2017), p. 103525.

[177] R. Mandelbaum et al. "The WiggleZ Dark Energy Survey: direct constraints on blue galaxy intrinsic alignments at intermediate redshifts". In: *MNRAS* 410 (Jan. 2011), pp. 844–859. DOI: 10.1111/j.1365-2966.2010.17485.x. arXiv: 0911.5347 [astro-ph.CO].

[178] Rachel Mandelbaum et al. "Weak lensing shear calibration with simulations of the HSC survey". In: *MNRAS* 481.3 (Dec. 2018), pp. 3170–3195. DOI: 10.1093/mnras/sty2420. arXiv: 1710.00885 [astro-ph.CO].

[179] Gabriela A. Marques and Armando Bernui. "Tomographic analyses of the CMB lensing and galaxy clustering to probe the linear structure growth". In: *Journal of Cosmology and Astroparticle Physics* 2020.05 (May 2020), pp. 052–052. DOI: 10.1088/1475-7516/2020/05/052. URL: https://doi.org/10.1088/1475-7516/2020/05/052.

[180] Daniel Martens et al. "A radial measurement of the galaxy tidal alignment magnitude with BOSS data". In: *MNRAS* 478.1 (July 2018), pp. 711–732. DOI: 10.1093/mnras/sty1100. arXiv: 1802.07708 [astro-ph.CO].

[181] José Manuel Zorrilla Matilla et al. "Geometry and growth contributions to cosmic shear observables". In: *Physical Review D* 96.2 (2017), p. 023513.

[182] A. Meiksin and M. White. "The growth of correlations in the matter power spectrum". In: *MNRAS* 308 (Oct. 1999), pp. 1179–1184. DOI: 10.1046/j.1365-8711.1999.02825.x. eprint: astro-ph/9812129.

[183] Siddharth Mishra-Sharma, David Alonso, and Joanna Dunkley. "Neutrino masses and beyond-Λ CDM cosmology with LSST and future CMB experiments". In: *Physical Review D* 97.12 (2018), p. 123544.

[184] Chirag Modi, Martin White, and Zvonimir Vlah. "Modeling CMB lensing cross correlations with CLEFT". In: *Journal of Cosmology and Astroparticle Physics* 2017.08 (2017), p. 009.

[185] I. Mohammed, U. Seljak, and Z. Vlah. "Perturbative approach to covariance matrix of the matter power spectrum". In: *MNRAS* 466 (Apr. 2017), pp. 780–797. DOI: 10.1093/mnras/stw3196. arXiv: 1607.00043.

[186] Toshiya Namikawa, Shun Saito, and Atsushi Taruya. "Probing dark energy and neutrino mass from upcoming lensing experiments of CMB and galaxies". In: *Journal of Cosmology and Astroparticle Physics* 2010.12 (2010), p. 027.

[187] A. Nigoche-Netro, A. Ruelas-Mayorga, and A. Franco-Balderas. "The Fundamental Plane for early-type galaxies: dependence on the magnitude range". In: *MNRAS* 392.3 (Jan. 2009), pp. 1060–1069. DOI: 10.1111/j.1365-2966.2008.14145.x. arXiv: 0805.1142 [astro-ph].

[188] P. Norberg et al. "Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics". In: *MNRAS* 396 (June 2009), pp. 19–38. DOI: 10.1111/j.1365-2966.2009.14389.x. arXiv: 0810.1885.

[189] R. O'Connell and D. J. Eisenstein. "Large Covariance Matrices: Accurate Models Without Mocks". In: *ArXiv e-prints* (Aug. 2018). arXiv: 1808.05978.

[190] R. O'Connell et al. "Large covariance matrices: smooth models from the two-point correlation function". In: *MNRAS* 462 (Nov. 2016), pp. 2681–2694. DOI: 10.1093/mnras/stw1821. arXiv: 1510.01740.

[191] Teppei Okumura, Uroš Seljak, and Vincent Desjacques. "Distribution function approach to redshift space distortions. Part III: halos and galaxies". In: *J. Cosmology Astropart. Phys.* 2012.11, 014 (Nov. 2012), p. 014. DOI: 10.1088/1475-7516/2012/11/014. arXiv: 1206.4070 [astro-ph.CO].

[192] Teppei Okumura et al. "Distribution function approach to redshift space distortions. Part II: N-body simulations". In: *J. Cosmology Astropart. Phys.* 2012.2, 010 (Feb. 2012), p. 010. DOI: 10.1088/1475-7516/2012/02/010. arXiv: 1109.1609 [astro-ph.CO].

[193] Teppei Okumura et al. "Galaxy power spectrum in redshift space: combining perturbation theory with the halo model". In: *Phys. Rev.* D92.10 (2015), p. 103516. DOI: 10.1103/PhysRevD.92.103516. arXiv: 1506.05814 [astro-ph.CO].

[194] Y. Omori et al. "Dark Energy Survey Year 1 Results: tomographic cross-correlations between DES galaxies and CMB lensing from SPT+Planck". In: *arXiv e-prints*, arXiv:1810.02342 (Oct. 2018), arXiv:1810.02342. arXiv: 1810.02342 [astro-ph.CO].

[195] Yoshiaki Ono et al. "Great Optically Luminous Dropout Research Using Subaru HSC (GOLDRUSH). I. UV Luminosity Functions at z=4-7 Derived with the Half-Million Dropouts on the 100 deg Sky". In: *arXiv preprint arXiv:1704.06004* (2017).

[196] M. Opper et al. "Perturbation Theory for Variational Inference". In: *Advances in Neural Information Processing Systems, Workshop on Advances in Approximate Bayesian Inference*. 2015.

[197] Manfred Opper and Cédric Archambeau. "The Variational Gaussian Approximation Revisited". In: *Neural Computation* 21.3 (2009), pp. 786–792. DOI: 10.1162/neco.2008.08-07-592. URL: https://doi.org/10.1162/neco.2008.08-07-592.

[198] Luis E. Padilla et al. "Cosmological Parameter Inference with Bayesian Statistics". In: *Universe* 7.7 (2021), p. 213. DOI: 10.3390/universe7070213. arXiv: 1903.11127 [astro-ph.CO].

[199] N. Padmanabhan et al. "An Improved Photometric Calibration of the Sloan Digital Sky Survey Imaging Data". In: *ApJ* 674 (Feb. 2008), pp. 1217–1233. DOI: 10.1086/524677. eprint: astro-ph/0703454.

[200] Zhen Pan and Lloyd Knox. "Constraints on neutrino mass from cosmic microwave background and large-scale structure". In: *Monthly Notices of the Royal Astronomical Society* 454.3 (2015), pp. 3200–3206.

[201] S. Pandey et al. "Dark Energy Survey Year 3 Results: Constraints on cosmological parameters and galaxy bias models from galaxy clustering and galaxy-galaxy lensing using the redMaGiC sample". In: *ArXiv e-prints* (May 2021). arXiv: 2105.13545 [astro-ph.CO].

[202] D. W. Pearson and L. Samushia. "Estimating the power spectrum covariance matrix with fewer mock samples". In: *MNRAS* 457 (Mar. 2016), pp. 993–999. DOI: 10.1093/mnras/stw062. arXiv: 1509.00064.

[203] Ue-Li Pen. "Beating lensing cosmic variance with galaxy tomography". In: *Monthly Notices of the Royal Astronomical Society* 350.4 (2004), pp. 1445–1448.

[204] Oliver H. E. Philcox and Mikhail M. Ivanov. "BOSS DR12 full-shape cosmology: Λ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole". In: *PRD* 105.4, 043517 (Feb. 2022), p. 043517. DOI: 10.1103/PhysRevD.105.043517. arXiv: 2112.04515 [astro-ph.CO].

[205] J. R. Pier et al. "Astrometric Calibration of the Sloan Digital Sky Survey". In: *AJ* 125 (Mar. 2003), pp. 1559–1579.

[206] Planck Collaboration et al. "Planck 2013 results. XVII. Gravitational lensing by large-scale structure". In: *A&A* 571, A17 (Nov. 2014), A17. DOI: 10.1051/0004-6361/201321543. arXiv: 1303.5077.

[207] Planck Collaboration et al. "Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters". In: *A&A* 594, A11 (Sept. 2016), A11. DOI: 10.1051/0004-6361/201526926. arXiv: 1507.02704.

[208] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: *ArXiv e-prints* (Feb. 2015). arXiv: 1502.01589.

[209] Planck Collaboration et al. "Planck 2018 results. VIII. Gravitational lensing". In: *ArXiv e-prints* (July 2018). arXiv: 1807.06210.

[210] A. C. Pope and I. Szapudi. "Shrinkage estimation of the power spectrum covariance matrix". In: *MNRAS* 389 (Sept. 2008), pp. 766–774. DOI: 10.1111/j.1365-2966.2008.13561.x. arXiv: 0711.2509.

[211] William H. Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. 3rd ed. USA: Cambridge University Press, 2007. ISBN: 0521880688.

[212] Rajesh Ranganath, Sean Gerrish, and David M Blei. "Black box variational inference". In: *arXiv preprint arXiv:1401.0118* (2013).

[213] B. Reid et al. "SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues". In: *MNRAS* 455 (Jan. 2016), pp. 1553–1573. DOI: 10.1093/mnras/stv2382. arXiv: 1509.06529.

[214] B. A. Reid and D. N. Spergel. "Constraining the Luminous Red Galaxy Halo Occupation Distribution Using Counts-In-Cylinders". In: *ApJ* 698 (June 2009), pp. 143–154. DOI: 10.1088/0004-637X/698/1/143. arXiv: 0809.4505.

[215] Beth A. Reid et al. "A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies". In: *Mon. Not. Roy. Astron. Soc.* 444.1 (2014), pp. 476–502. DOI: 10.1093/mnras/stu1391. arXiv: 1404.3742 [astro-ph.CO].

[216] R. Reyes et al. "Optical-to-virial velocity ratios of local disc galaxies from combined kinematics and galaxy-galaxy lensing". In: *MNRAS* 425 (Oct. 2012), pp. 2610–2640. DOI: 10.1111/j.1365-2966.2012.21472.x.

[217] Danilo Jimenez Rezende and Shakir Mohamed. "Variational Inference with Normalizing Flows". In: *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*. 2015, pp. 1530–1538. URL: http://jmlr.org/proceedings/papers/v37/rezende15.html.

[218] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*. 2014, pp. 1278–1286. URL: http://jmlr.org/proceedings/papers/v32/rezende14.html.

[219] G. T. Richards et al. "Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Quasar Sample". In: *AJ* 123 (June 2002), pp. 2945–2975. DOI: 10.1086/340187. eprint: astro-ph/0202251.

[220] A. G. Riess et al. "A 2.4% Determination of the Local Value of the Hubble Constant". In: *ApJ* 826, 56 (July 2016), p. 56. DOI: 10.3847/0004-637X/826/1/56. arXiv: 1604.01424.

[221] Joseph Rocca. *Bayesian Inference Problem, MCMC and Variational Inference*. 2019. URL: https://towardsdatascience.com/bayesian-inference-problem-mcmc-and-variational-inference-25a8aa9bce29.

[222] Geoffrey Roeder, Yuhuai Wu, and David K. Duvenaud. "Sticking the Landing: Simple, Lower-Variance Gradient Estimators for Variational Inference". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*. 2017, pp. 6928–6937.

[223] A. J. Ross et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: analysis of potential systematics". In: *MNRAS* 424 (July 2012), pp. 564–590. DOI: 10.1111/j.1365-2966.2012.21235.x. arXiv: 1203.6499.

[224] John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. "Probabilistic programming in Python using PyMC3". In: *PeerJ Computer Science* 2 (2016), e55.

[225] Larissa Santos et al. "Neutrinos and dark energy constraints from future galaxy surveys and CMB lensing information". In: *Physical Review D* 88.4 (2013), p. 043505.

[226] Siddharth Satpathy et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: On the measurement of growth rate using galaxy correlation functions". In: *Mon. Not. Roy. Astron. Soc.* 469.2 (2017), pp. 1369–1382. DOI: 10.1093/mnras/stx883. arXiv: 1607.03148 [astro-ph.CO].

[227] Christoph Saulder et al. "Calibrating the fundamental plane with SDSS DR8 data". In: *A&A* 557, A21 (Sept. 2013), A21. DOI: 10.1051/0004-6361/201321466. arXiv: 1306.0285 [astro-ph.CO].

[228] Christoph Saulder et al. "Distance measurements to early-type galaxies by improving the fundamental plane". In: *arXiv e-prints*, arXiv:1905.12970 (May 2019), arXiv:1905.12970. arXiv: 1905.12970 [astro-ph.CO].

[229] Emmanuel Schaan, Simone Ferraro, and Uroš Seljak. "Photo-z outlier self-calibration in weak lensing surveys". In: *JCAP* 12 (2020), p. 001. DOI: 10.1088/1475-7516/2020/12/001. arXiv: 2007.12795 [astro-ph.CO].

[230] Emmanuel Schaan et al. "Looking through the same lens: shear calibration for LSST, Euclid, and WFIRST with stage IV CMB Lensing". In: *Physical Review D* 95.12 (2017).

[231] Marcel Schmittfull and Uroš Seljak. "Parameter constraints from cross-correlation of CMB lensing with galaxy clustering". In: *Physical Review D* 97.12 (2018), p. 123540.

[232] R. L. Schuhmann, B. Joachimi, and H. V. Peiris. "Gaussianization for fast and accurate inference from cosmological data". In: *MNRAS* 459 (June 2016), pp. 1916–1928. DOI: 10.1093/mnras/stw738. arXiv: 1510.00019.

[233] R. Scoccimarro, M. Zaldarriaga, and L. Hui. "Power Spectrum Correlations Induced by Nonlinear Clustering". In: *APJ* 527 (Dec. 1999), pp. 1–15. DOI: 10.1086/308059. eprint: astro-ph/9901099.

[234] Roman Scoccimarro. "Fast Estimators for Redshift-Space Clustering". In: *Phys. Rev.* D92.8 (2015), p. 083532. DOI: 10.1103/PhysRevD.92.083532. arXiv: 1506.02729 [astro-ph.CO].

[235] M. Scodeggio et al. "The tilt of the Fundamental Plane of early-type galaxies: wavelength dependence". In: *MNRAS* 301.4 (Dec. 1998), pp. 1001–1018. DOI: 10.1046/j.1365-8711.1998.02106.x. arXiv: astro-ph/9809191 [astro-ph].

[236] Uros Seljak and Patrick McDonald. "Distribution function approach to redshift space distortions". In: *JCAP* 1111 (2011), p. 039. DOI: 10.1088/1475-7516/2011/11/039. arXiv: 1109.1888 [astro-ph.CO].

[237] Uros Seljak and Byeonghee Yu. "Posterior inference unchained with EL_2O". In: *arXiv e-prints*, arXiv:1901.04454 (Jan. 2019), arXiv:1901.04454. arXiv: 1901.04454 [stat.ML].

[238] E. Sellentin and A. F. Heavens. "Parameter inference with estimated covariance matrices". In: *MNRAS* 456 (Feb. 2016), pp. L132–L136. DOI: 10.1093/mnrasl/slv190. arXiv: 1511.05969.

[239] Blake D. Sherwin et al. "The Atacama Cosmology Telescope: Cross-Correlation of CMB Lensing and Quasars". In: *Phys. Rev.* D86 (2012), p. 083006. DOI: 10.1103/PhysRevD.86.083006. arXiv: 1207.4543 [astro-ph.CO].

[240] Fergus Simpson and Sarah Bridle. "Illuminating dark energy with cosmic shear". In: *Physical Review D* 71.8 (2005), p. 083501.

[241] S. Singh, R. Mandelbaum, and S. More. "Intrinsic alignments of SDSS-III BOSS LOWZ sample galaxies". In: *MNRAS* 450 (June 2015), pp. 2195–2216. DOI: 10.1093/mnras/stv778. arXiv: 1411.1755.

[242] S. Singh et al. "Galaxy-galaxy lensing estimators and their covariance properties". In: *MNRAS* 471 (Nov. 2017), pp. 3827–3844. DOI: 10.1093/mnras/stx1828. arXiv: 1611.00752.

[243] S. Singh et al. "Probing gravity with a joint analysis of galaxy and CMB lensing and SDSS spectroscopy". In: *MNRAS* (Oct. 2018). DOI: 10.1093/mnras/sty2681. arXiv: 1803.08915.

[244] Sukhdeep Singh and Rachel Mandelbaum. "Intrinsic alignments of BOSS LOWZ galaxies - II. Impact of shape measurement methods". In: *MNRAS* 457.3 (Apr. 2016), pp. 2301–2317. DOI: 10.1093/mnras/stw144. arXiv: 1510.06752 [astro-ph.CO].

[245] M. Sinha and L. Garrison. *Corrfunc: Blazing fast correlation functions on the CPU*. Astrophysics Source Code Library. Mar. 2017. ascl: 1703.003.

[246] John Skilling. "Nested Sampling". In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Ed. by Rainer Fischer, Roland Preuss, and Udo Von Toussaint. Vol. 735. American Institute of Physics Conference Series. Nov. 2004, pp. 395–405. DOI: 10.1063/1.1835238.

[247] Z. Slepian and D. J. Eisenstein. "Accelerating the two-point and three-point galaxy correlation functions using Fourier transforms". In: *MNRAS* 455 (Jan. 2016), pp. L31–L35. DOI: 10.1093/mnrasl/slv133. arXiv: 1506.04746.

[248] Z. Slepian and D. J. Eisenstein. "Computing the three-point correlation function of galaxies in O(N^2) time". In: *MNRAS* 454 (Dec. 2015), pp. 4142–4158. DOI: 10.1093/mnras/stv2119. arXiv: 1506.02040.

[249] S. A. Smee et al. "The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey". In: *AJ* 146, 32 (Aug. 2013), p. 32. DOI: 10.1088/0004-6256/146/2/32. arXiv: 1208.2233 [astro-ph.IM].

[250] J. A. Smith et al. "The ugriz Standard-Star System". In: *AJ* 123 (Apr. 2002), pp. 2121–2144.

[251] Kendrick M Smith, Wayne Hu, and Manoj Kaplinghat. "Cosmological information from lensed CMB power spectra". In: *Physical Review D* 74.12 (2006), p. 123002.

[252] Kendrick M Smith et al. "Delensing CMB polarization with external datasets". In: *Journal of Cosmology and Astroparticle Physics* 2012.06 (2012), p. 014.

[253] R. E. Smith. "Covariance of cross-correlations: towards efficient measures for large-scale structure". In: *MNRAS* 400 (Dec. 2009), pp. 851–865. DOI: 10.1111/j.1365-2966.2009.15490.x. arXiv: 0810.1960.

[254] Robert E Smith et al. "Stable clustering, the halo model and non-linear cosmological power spectra". In: *Monthly Notices of the Royal Astronomical Society* 341.4 (2003), pp. 1311–1332.

[255] Y.-S. Song et al. "The Far-Infrared Background Correlation with Cosmic Microwave Background Lensing". In: *Astrophys. J.* 590 (2003), p. 664.

[256] M. A. Strauss and J. A. Willick. "The density and peculiar velocity fields of nearby galaxies". In: *Phys.Rep.* 261 (Jan. 1995), pp. 271–431. DOI: 10.1016/0370-1573(95)00013-7. arXiv: astro-ph/9502079 [astro-ph].

[257] M. A. Strauss et al. "Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample". In: *AJ* 124 (Sept. 2002), pp. 1810–1824.

[258] N. S. Sugiyama, M. Shiraishi, and T. Okumura. "Limits on statistical anisotropy from BOSS DR12 galaxies using bipolar spherical harmonics". In: *MNRAS* 473 (Jan. 2018), pp. 2737–2752. DOI: 10.1093/mnras/stx2333. arXiv: 1704.02868.

[259] M. Takada and B. Jain. "The impact of non-Gaussian errors on weak lensing surveys". In: *MNRAS* 395 (June 2009), pp. 2065–2086. DOI: 10.1111/j.1365-2966.2009.14504.x. arXiv: 0810.4170.

[260] Masahiro Takada and Wayne Hu. "Power spectrum super-sample covariance". In: *Phys.Rev.D* 87.12, 123504 (June 2013), p. 123504. DOI: 10.1103/PhysRevD.87.123504. arXiv: 1302.6994 [astro-ph.CO].

[261] R. Takahashi et al. "Revising the Halofit Model for the Nonlinear Matter Power Spectrum". In: *ApJ* 761, 152 (Dec. 2012), p. 152. DOI: 10.1088/0004-637X/761/2/152. arXiv: 1208.2701.

[262] Ryuichi Takahashi et al. "Revising the halofit model for the nonlinear matter power spectrum". In: *The Astrophysical Journal* 761.2 (2012), p. 152.

[263] Yoshitaka Takeuchi and Kenji Kadota. "Probing neutrinos from Planck and forthcoming galaxy redshift surveys". In: *Journal of Cosmology and Astroparticle Physics* 2014.01 (2014), p. 046.

[264] J. D. Talman. "Numerical Fourier and Bessel Transforms in Logarithmic Variables". In: *Journal of Computational Physics* 29 (Oct. 1978), pp. 35–48. DOI: 10.1016/0021-9991(78)90107-9.

[265] A. Taylor, B. Joachimi, and T. Kitching. "Putting the precision in precision cosmology: How accurate should your data covariance matrix be?" In: *MNRAS* 432 (July 2013), pp. 1928–1946. DOI: 10.1093/mnras/stt270. arXiv: 1212.4359.

[266] M. Tegmark. "Measuring Cosmological Parameters with Galaxy Surveys". In: *Physical Review Letters* 79 (Nov. 1997), pp. 3806–3809. DOI: 10.1103/PhysRevLett.79.3806. eprint: astro-ph/9706198.

[267] Max Tegmark, Andy N Taylor, and Alan F Heavens. "Karhunen-Loeve eigenvalue problems in cosmology: how should we tackle large data sets?" In: *The Astrophysical Journal* 480.1 (1997), p. 22.

[268] Richard C. Tolman. "On the Estimation of Distances in a Curved Universe with a Non-Static Line Element". In: *Proceedings of the National Academy of Science* 16.7 (July 1930), pp. 511–520. DOI: 10.1073/pnas.16.7.511.

[269] M. A. Troxel et al. "Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear". In: *Phys.Rev.D* 98.4, 043528 (Aug. 2018), p. 043528. DOI: 10.1103/PhysRevD.98.043528. arXiv: 1708.01538.

[270] D. L. Tucker et al. "The Sloan Digital Sky Survey monitor telescope pipeline". In: *Astronomische Nachrichten* 327 (2006), pp. 821–+. DOI: 10.1002/asna.200610655. eprint: arXiv:astro-ph/0608575.

[271] S Unnikrishnan et al. "Effect of dark energy sound speed and equation of state on CDM power spectrum". In: *Journal of Physics: Conference Series*. Vol. 484. 1. IOP Publishing. 2014, p. 012048.

[272] Zvonimir Vlah et al. "Distribution function approach to redshift space distortions. Part IV: perturbation theory applied to dark matter". In: *J. Cosmology Astropart. Phys.* 2012.11, 009 (Nov. 2012), p. 009. DOI: 10.1088/1475-7516/2012/11/009. arXiv: 1207.0839 [astro-ph.CO].

[273] Zvonimir Vlah et al. "Distribution function approach to redshift space distortions. Part V: perturbation theory applied to dark matter halos". In: *JCAP* 1310 (2013), p. 053. DOI: 10.1088/1475-7516/2013/10/053. arXiv: 1308.6294 [astro-ph.CO].

[274] Digvijay Wadekar, Mikhail M. Ivanov, and Roman Scoccimarro. "Cosmological constraints from BOSS with analytic covariance matrices". In: *Phys. Rev. D* 102 (2020), p. 123521. DOI: 10.1103/PhysRevD.102.123521. arXiv: 2009.00622 [astro-ph.CO].

[275] Martin J. Wainwright and Michael I. Jordan. "Graphical Models, Exponential Families, and Variational Inference". In: *Foundations and Trends in Machine Learning* 1.1-2 (2008), pp. 1–305. DOI: 10.1561/2200000001. URL: https://doi.org/10.1561/2200000001.

[276] D. A. Wake et al. "The 2df SDSS LRG and QSO survey: evolution of the luminosity function of luminous red galaxies to z = 0.6". In: *MNRAS* 372 (Oct. 2006), pp. 537–550. DOI: 10.1111/j.1365-2966.2006.10831.x. eprint: astro-ph/0607629.

[277] M. J. Wilson et al. "Rapid modelling of the redshift-space power spectrum multipoles for a masked density field". In: *MNRAS* 464 (Jan. 2017), pp. 3121–3130. DOI: 10.1093/mnras/stw2576. arXiv: 1511.07799.

[278] WLK Wu et al. "A guide to designing future ground-based cosmic microwave background experiments". In: *The Astrophysical Journal* 788.2 (2014), p. 138.

[279] Kazuhiro Yamamoto et al. "A Measurement of the quadrupole power spectrum in the clustering of the 2dF QSO Survey". In: *Publ. Astron. Soc. Jap.* 58 (2006), pp. 93–102. DOI: 10.1093/pasj/58.1.93. arXiv: astro-ph/0505115 [astro-ph].

[280] Yuling Yao et al. "Yes, but Did It Work?: Evaluating Variational Inference". In: *PMLR* 80 (2018), pp. 5581–5590. URL: http://proceedings.mlr.press/v80/yao18a.html.

[281] D. G. York et al. "The Sloan Digital Sky Survey: Technical Summary". In: *AJ* 120 (Sept. 2000), pp. 1579–1587.

[282] Byeonghee Yu et al. "RSD measurements from BOSS galaxy power spectrum using the halo perturbation theory model". In: (Nov. 2022). arXiv: 2211.16794 [astro-ph.CO].

[283] Byeonghee Yu et al. "Towards Neutrino Mass from Cosmology without Optical Depth Information". In: *https://arxiv.org/abs/1809.02120* (2018).

[284] Sihan Yuan et al. "Stringent σ_8 constraints from small-scale galaxy clustering using a hybrid MCMC+emulator framework". In: (Mar. 2022). arXiv: 2203.11963 [astro-ph.CO].

[285] Zhongxu Zhai et al. "The Aemulus Project V: Cosmological constraint from small-scale clustering of BOSS galaxies". In: (Mar. 2022). arXiv: 2203.08999 [astro-ph.CO].

[286] Hu Zhan. "Cosmic tomographies: baryon acoustic oscillations and weak lensing". In: *Journal of Cosmology and Astroparticle Physics* 2006.08 (2006), p. 008.

[287] Hu Zhan and Lloyd Knox. "How tomographic cosmic shear maps lead to constraints on dark energy properties". In: *arXiv preprint astro-ph/0611159* (2006).

[288] Hu Zhan, Lloyd Knox, and J Anthony Tyson. "Distance, growth factor, and dark energy constraints from photometric baryon acoustic oscillation and weak lensing measurements". In: *The Astrophysical Journal* 690.1 (2008), p. 923.

[289] Hu Zhan and J Anthony Tyson. "Cosmology with the Large Synoptic Survey Telescope: an Overview". In: *Reports on Progress in Physics* 81.6 (2018), p. 066901.

[290] Jun Zhang, Lam Hui, and Albert Stebbins. "Isolating geometry in weak-lensing measurements". In: *The Astrophysical Journal* 635.2 (2005), p. 806.

[291] Pierre Zhang et al. "BOSS Correlation Function analysis from the Effective Field Theory of Large-Scale Structure". In: *JCAP* 02.02 (2022), p. 036. DOI: 10.1088/1475-7516/2022/02/036. arXiv: 2110.07539 [astro-ph.CO].

[292] Pierre Zhang et al. "BOSS correlation function analysis from the effective field theory of large-scale structure". In: *Journal of Cosmology and Astroparticle Physics* 2022.02 (2022), p. 036.