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

Yin Li,[a,b] **Sukhdeep Singh,**[a] **Byeonghee Yu,**[a] **Yu Feng,**[a] **& Uroš Seljak**[a,c]

[a]Berkeley Center for Cosmological Physics, Department of Physics, & Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA

[b]Kavli Institute for the Physics and Mathematics of the Universe (WPI), UTIAS, The University of Tokyo, Chiba 277–8583, Japan

[c]Department of Astronomy, University of California, Berkeley, CA 94720, USA

E-mail: yin.li@berkeley.edu, sukhdeep1@berkeley.edu, bhyu@berkeley.edu, yfeng1@berkeley.edu, useljak@berkeley.edu

Abstract. 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 work, 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. The code that accompanies this paper is available at https://github.com/eelregit/covdisc.

Contents

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) [1–3], baryon acoustic oscillations (BAO) [4–6], weak gravitational lensing [7–10], supernovae [11] and strong lensing [12]. The next generation of galaxy surveys, LSST [13], Euclid [14], DESI [15], WFIRST [16], 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 work 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 [17–19].

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 [20]. To overcome these challenges, several fast methods to run approximate mocks have been developed (see e.g. [21, 22] for comparison of covariances from different mocks). Nonetheless, lack of correct physics in these mocks potentially leads to biased covariances [23]. 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 [24]. 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 [25]. 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 [26–33], the super-sample covariance [34–42], shot noises [26, 43, 44], 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 [45–49]. 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 work, 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 paper we focus on the disconnected covariance, and leave the internal estimation of the connected part as a future project.

This paper is structured as follows. In Sec. 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. 3 with a novel quadrature algorithm. Using the mock simulations described in Sec. 4, we demonstrate the accuracy and efficiency of the analytic method in Sec. 5. App. A gives the detailed calculations that generalize our method for different applications, and App. B provides the mathematical tools useful in the derivations.

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.

2.1 3D Power Spectrum Covariance

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

$$n_{\mathrm{data}}(\boldsymbol{x}) = \sum_i \delta^{\mathrm{D}}(\boldsymbol{x} - \boldsymbol{x}_i), \tag{2.1}$$

whose underlying number density is a continuous field $\bar{n}(\boldsymbol{x}) \equiv \langle n_{\mathrm{data}}(\boldsymbol{x}) \rangle$, 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 $\langle n_{\mathrm{rand}}(\boldsymbol{x}) \rangle = \bar{n}(\boldsymbol{x})/\alpha$, 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[1]

$$\delta_{\text{data}} = -1 + \frac{1}{\bar{n}(\boldsymbol{x})} \sum_{i \in \text{data}} \delta^{\text{D}}(\boldsymbol{x} - \boldsymbol{x}_i),$$

$$\delta_{\text{rand}} = -1 + \frac{\alpha}{\bar{n}(\boldsymbol{x})} \sum_{j \in \text{rand}} \delta^{\text{D}}(\boldsymbol{x} - \boldsymbol{x}_j). \qquad (2.2)$$

Following [25] (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)

$$\delta_W(\boldsymbol{x}) \equiv w(\boldsymbol{x}) \big[n_{\text{data}}(\boldsymbol{x}) - \alpha n_{\text{rand}}(\boldsymbol{x}) \big]$$
$$\equiv W(\boldsymbol{x}) \big[\delta_{\text{data}}(\boldsymbol{x}) - \delta_{\text{rand}}(\boldsymbol{x}) \big] \equiv W(\boldsymbol{x}) \delta(\boldsymbol{x}). \qquad (2.3)$$

On the second line we have rewritten the difference in number densities with the difference in overdensities, and defined a combined overdensity field $\delta \equiv \delta_{\text{data}} - \delta_{\text{rand}}$ 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

$$W(\boldsymbol{x}) \equiv \bar{n}(\boldsymbol{x}) w(\boldsymbol{x}). \qquad (2.4)$$

Throughout this paper 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

$$\hat{P}(\boldsymbol{k}) = \frac{|\delta_W(\boldsymbol{k})|^2}{\mathcal{W}_0} - P_{\text{shot}}, \qquad (2.5)$$

where the shot noise power spectrum is

$$P_{\text{shot}} = \frac{\mathcal{S}_0}{\mathcal{W}_0}, \qquad (2.6)$$

and the constant factors are given by[2]

$$\mathcal{W}_0 = \int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})^2 w(\boldsymbol{x})^2 = \int_{\boldsymbol{x}} W(\boldsymbol{x})^2 = \int_{\boldsymbol{k}} |W(\boldsymbol{k})|^2,$$

$$\mathcal{S}_0 = (1 + \alpha) \int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x}) w(\boldsymbol{x})^2. \qquad (2.7)$$

---

[1] These overdensity fields of discrete catalogs are only formal, and they help to simplify the calculation of shot noise terms, e.g. $\langle \delta_{\text{data}} \delta_{\text{rand}} \rangle$ vanishes in (2.8).

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

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

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

$$\langle\delta_{\text{data}}(\boldsymbol{k})\delta_{\text{data}}(-\boldsymbol{k}')\rangle = (2\pi)^3\,\delta^{\text{D}}(\boldsymbol{k}-\boldsymbol{k}')P(\boldsymbol{k}) + \int_{\boldsymbol{x}}\frac{1}{\bar{n}(\boldsymbol{x})}e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}},$$
$$\langle\delta_{\text{rand}}(\boldsymbol{k})\delta_{\text{rand}}(-\boldsymbol{k}')\rangle = \int_{\boldsymbol{x}}\frac{\alpha}{\bar{n}(\boldsymbol{x})}e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}},$$
$$\langle\delta_{\text{data}}(\boldsymbol{k})\delta_{\text{rand}}(-\boldsymbol{k}')\rangle = 0. \tag{2.8}$$

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 (2.3) and (2.8) into the ensemble average of the estimator in (2.5) we get

$$\langle\hat{P}(\boldsymbol{k})\rangle = \frac{1}{\mathcal{W}_0}\int_{\boldsymbol{q}}P(\boldsymbol{k}-\boldsymbol{q})|W(\boldsymbol{q})|^2 \simeq \frac{P(\boldsymbol{k})}{\mathcal{W}_0}\int_{\boldsymbol{q}}|W(\boldsymbol{q})|^2 = P(\boldsymbol{k}). \tag{2.9}$$

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 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 (2.5) 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

$$\text{Cov}\big[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\big] \equiv \langle\hat{P}(\boldsymbol{k})\hat{P}(\boldsymbol{k}')\rangle - \langle\hat{P}(\boldsymbol{k})\rangle\langle\hat{P}(\boldsymbol{k}')\rangle \tag{2.10}$$

Substituting (2.5) into the above equation, we split the covariance into the disconnected and connected pieces in the multivariate cumulant expansion

$$\text{Cov}\big[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\big] = \text{Cov}^{\text{disc}}\big[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\big] + \text{Cov}^{\text{conn}}\big[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\big],$$
$$\text{Cov}^{\text{disc}}\big[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\big] = \frac{1}{\mathcal{W}_0^2}\big|\langle\delta_W(\boldsymbol{k})\delta_W(-\boldsymbol{k}')\rangle\big|^2 + (\boldsymbol{k}'\leftrightarrow-\boldsymbol{k}'),$$
$$\text{Cov}^{\text{conn}}\big[\hat{P}(\boldsymbol{k}),\hat{P}(\boldsymbol{k}')\big] = \frac{1}{\mathcal{W}_0^2}\langle\delta_W(\boldsymbol{k})\delta_W(-\boldsymbol{k})\delta_W(\boldsymbol{k}')\delta_W(-\boldsymbol{k}')\rangle_{\text{c}}. \tag{2.11}$$

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

In the conventional terminology, Cov^{disc} is referred to as the Gaussian part and Cov^{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 Cov^{disc} and Cov^{conn}, Cov^{disc} is

not purely Gaussian and Cov^{conn} is not completely free of Gaussian contribution. Therefore in this paper we rename this covariance decomposition for clarification.

Let's first look at the disconnected piece. From (2.3) and (2.8)

$$\langle \delta_W(\boldsymbol{k})\delta_W(-\boldsymbol{k}')\rangle = \int_{\boldsymbol{k}''} P(\boldsymbol{k}'')W(\boldsymbol{k}-\boldsymbol{k}'')W(\boldsymbol{k}''-\boldsymbol{k}')+(1+\alpha)\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})w(\boldsymbol{x})^2 e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}, \quad (2.12)$$

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 (2.11)

$$\text{Cov}^{\text{disc}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] \approx \left| \frac{P(\boldsymbol{k})+P(\boldsymbol{k}')}{2}\frac{\mathcal{W}(\boldsymbol{k}-\boldsymbol{k}')}{\mathcal{W}_0} + P_{\text{shot}}\frac{\mathcal{S}(\boldsymbol{k}-\boldsymbol{k}')}{\mathcal{S}_0}\right|^2 + (\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}'), \quad (2.13)$$

where we have introduced the following window factors

$$\mathcal{W}(\boldsymbol{q}) = \int_{\boldsymbol{x}} \mathcal{W}(\boldsymbol{x})e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \equiv \int_{\boldsymbol{x}} W(\boldsymbol{x})^2 e^{-i\boldsymbol{q}\cdot\boldsymbol{x}},$$

$$\mathcal{S}(\boldsymbol{q}) = \int_{\boldsymbol{x}} \mathcal{S}(\boldsymbol{x})e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \equiv (1+\alpha)\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})w(\boldsymbol{x})^2 e^{-i\boldsymbol{q}\cdot\boldsymbol{x}}, \quad (2.14)$$

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 (2.7) are special cases of $\mathcal{W}(\boldsymbol{q})$ and $\mathcal{S}(\boldsymbol{q})$ at $\boldsymbol{q}=0$.

The expansion of (2.13) contains quadratic combinations $\mathcal{W}^2$, $\mathcal{WS}$, 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

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

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

$$\mathcal{Q}_{\times}(\boldsymbol{q}) \equiv \mathcal{W}(\boldsymbol{q})\mathcal{S}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\times}(\boldsymbol{s})e^{-i\boldsymbol{q}\cdot\boldsymbol{s}}, \quad (2.15)$$

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.

$$\mathcal{Q}_{\times}(\boldsymbol{s}) = \int_{\boldsymbol{x}} \mathcal{W}(\boldsymbol{x}+\boldsymbol{s})\mathcal{S}(\boldsymbol{x}). \quad (2.16)$$

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. 2.3.

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

$$\text{Cov}^{\text{disc}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] \approx \frac{1}{\mathcal{W}_0^2}\Big\{ P(\boldsymbol{k})P(\boldsymbol{k}')\mathcal{Q}_{\mathcal{W}}(\boldsymbol{k}-\boldsymbol{k}')$$

$$+ \big[P(\boldsymbol{k})+P(\boldsymbol{k}')\big]\Re\big[\mathcal{Q}_{\times}(\boldsymbol{k}-\boldsymbol{k}')\big] + \mathcal{Q}_{\mathcal{S}}(\boldsymbol{k}-\boldsymbol{k}')\Big\} + (\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}'). \quad (2.17)$$

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. 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 [26, 27] and the super-sample covariance (SSC) [34, 35, 37]. The remaining parts of $\mathrm{Cov}^{\mathrm{conn}}$ are various shot noise terms due to the discrete nature of the tracer field [26], 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 [29, 31], 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 paper 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.

2.1.1 Diagonal Limit

We expect $\mathrm{Cov}^{\mathrm{disc}}$ in (2.17) 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:

$$\mathrm{Cov}^{\mathrm{diag}}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] = (2\pi)^3 \, \delta^{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}')\left\{ \frac{P(\boldsymbol{k})^2}{V_{\mathcal{W}}} + \frac{2P(\boldsymbol{k})P_{\mathrm{shot}}}{V_{\times}} + \frac{P_{\mathrm{shot}}^2}{V_{\mathcal{S}}} \right\} + (\boldsymbol{k}' \leftrightarrow -\boldsymbol{k}'), \quad (2.18)$$

where V's are effective volumes defined as follows

$$V_{\mathcal{W}} \equiv \frac{\mathcal{W}_0^2}{\mathcal{Q}_{\mathcal{W}}(\boldsymbol{s} = 0)} = \frac{\big[\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})^2 w(\boldsymbol{x})^2\big]^2}{\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})^4 w(\boldsymbol{x})^4},$$

$$V_{\mathcal{S}} \equiv \frac{\mathcal{S}_0^2}{\mathcal{Q}_{\mathcal{S}}(\boldsymbol{s} = 0)} = \frac{\big[\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x}) w(\boldsymbol{x})^2\big]^2}{\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})^2 w(\boldsymbol{x})^4},$$

$$V_{\times} \equiv \frac{\mathcal{W}_0 \mathcal{S}_0}{\mathcal{Q}_{\times}(\boldsymbol{s} = 0)} = \frac{\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})^2 w(\boldsymbol{x})^2 \int_{\boldsymbol{x}'} \bar{n}(\boldsymbol{x}') w(\boldsymbol{x}')^2}{\int_{\boldsymbol{x}} \bar{n}(\boldsymbol{x})^3 w(\boldsymbol{x})^4}. \quad (2.19)$$

We call (2.18) 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. 5.1.1.

In the diagonal limit, the $\mathrm{Cov}^{\mathrm{disc}}$ matrix of the band-power, binned in spherical k shells following the later Sec. 2.5, is

$$\mathrm{Cov}^{\mathrm{diag}}\big[\hat{P}(k_i), \hat{P}(k_j)\big] = 2\delta_{ij}^{\mathrm{K}} \int_{k_{i-\frac{1}{2}}}^{k_{i+\frac{1}{2}}} \frac{4\pi k^2 \mathrm{d}k}{V_{k_i}} \left\{ \frac{P(k)^2}{N_{\mathcal{W}}} + \frac{2P(k)P_{\mathrm{shot}}}{N_{\times}} + \frac{P_{\mathrm{shot}}^2}{N_{\mathcal{S}}} \right\}, \quad (2.20)$$

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

$$N_{\mathcal{W}} = \frac{V_k V_{\mathcal{W}}}{(2\pi)^3}, \quad N_{\mathcal{S}} = \frac{V_k V_{\mathcal{S}}}{(2\pi)^3}, \quad N_{\times} = \frac{V_k V_{\times}}{(2\pi)^3}. \tag{2.21}$$

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. 5.1.1). Eq. (2.18) and the constant effective volumes generalize the scale-dependent effective volume commonly used in survey forecast [50] for any weights $w(\boldsymbol{x})$.

2.2 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

$$P(\boldsymbol{k}) = \sum_\ell P_\ell(k)\, \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}), \tag{2.22}$$

in which $\mathcal{L}_\ell$ is the Legendre polynomial of degree ℓ. The multipole moments can be obtained by[3]

$$P_\ell(k) = (2\ell + 1) \int_{\hat{\boldsymbol{k}}} P(\boldsymbol{k})\, \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}). \tag{2.23}$$

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

(2.22) and (2.23) also apply to their estimators $\hat{P}(\boldsymbol{k})$ and $\hat{P}_\ell(k)$, so the disconnected covariance of multipoles follows straightforwardly from (2.17)

$$\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}_\ell(k), \hat{P}_{\ell'}(k')\big] \approx \frac{2(2\ell+1)(2\ell'+1)}{\mathcal{W}_0^2} \int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}})\, \mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}) \Big\{$$
$$P(\boldsymbol{k})P(\boldsymbol{k}')\mathcal{Q}_{\mathcal{W}}(\boldsymbol{k} - \boldsymbol{k}') + \big[P(\boldsymbol{k}) + P(\boldsymbol{k}')\big]\Re\big[\mathcal{Q}_{\times}(\boldsymbol{k} - \boldsymbol{k}')\big] + \mathcal{Q}_{\mathcal{S}}(\boldsymbol{k} - \boldsymbol{k}')\Big\}. \tag{2.24}$$

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 (2.17) simply doubles the first one giving the factor of 2 in the above equation.

We compute the three terms in the curly brackets of (2.24) separately. First we consider the P^2 term, expand the power spectra in multipoles, plug in $\mathcal{Q}_{\mathcal{W}}$ from (2.15), and repeatedly couple Legendre polynomials with (B.1) and apply (B.7) to derive

$$\int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}})\, \mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}) P(\boldsymbol{k})P(\boldsymbol{k}') \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{W}}(\boldsymbol{s}) e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{s}}$$
$$= \sum_{\ell_1\ell_2\ell_3\ell_4} P_{\ell_1}(k)P_{\ell_3}(k')(2\ell_2+1)\begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 (2\ell_4+1)\begin{pmatrix} \ell' & \ell_3 & \ell_4 \\ 0 & 0 & 0 \end{pmatrix}^2$$
$$\times (-i)^{\ell_2-\ell_4} \int 4\pi s^2 \mathrm{d}s \sum_{\ell''} \begin{pmatrix} \ell_2 & \ell_4 & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2 \mathcal{Q}_{\mathcal{W}\ell''}(s) j_{\ell_2}(ks) j_{\ell_4}(k's). \tag{2.25}$$

---

[3] In this paper we use the following shorthand notation for averaging (instead of integrating) over 4π solid angle of any vector $\boldsymbol{v}$

$$\int_{\hat{\boldsymbol{v}}} \to \int \frac{\mathrm{d}\Omega_{\boldsymbol{v}}}{4\pi}.$$

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

$$
\int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}) \, \mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}) \big[P(\boldsymbol{k}) + P(\boldsymbol{k}') \big] \Re \Big[\int_{\boldsymbol{s}} \mathcal{Q}_\times(\boldsymbol{s}) e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{s}} \Big]
$$

$$
= \sum_{\ell_1 \ell_2} P_{\ell_1}(k)(2\ell_2 + 1) \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 (-i)^{\ell_2 - \ell'} \int 4\pi s^2 \mathrm{d}s \sum_{\ell''} \begin{pmatrix} \ell_2 & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2 \mathcal{Q}_{\times \ell''}(s) j_{\ell_2}(ks) j_{\ell'}(k's)
$$

$$
+ (\ell, k \leftrightarrow \ell', k'), \tag{2.26}
$$

and the third (P_{shot}^2) term

$$
\int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}) \, \mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}) \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{S}}(\boldsymbol{s}) e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{s}}
$$

$$
= (-i)^{\ell - \ell'} \int 4\pi s^2 \mathrm{d}s \sum_{\ell''} \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2 \mathcal{Q}_{\mathcal{S}\ell''}(s) j_\ell(ks) j_{\ell'}(k's). \tag{2.27}
$$

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

$$
\mathcal{Q}_\ell(s) = (2\ell + 1) \int_{\hat{\boldsymbol{s}}} \mathcal{Q}(\boldsymbol{s}) \, \mathcal{L}_\ell(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{n}}). \tag{2.28}
$$

As shown in (2.15) and (2.16), 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 (2.25), (2.26) and (2.27) before summing them up in (2.24) 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. 3.

2.3 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. [51] measures the multipoles of galaxy clustering with respect to the midpoint direction of the two galaxies. An alternative estimator also suggested by Ref. [51] 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 [52–56].

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 (2.28) 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 Cov^{disc}.

We use `nbodykit`, which calls `corrfunc` [57], 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

$$\mathcal{Q}_{\mathcal{S}}(s, \mu) \approx \frac{3 \int_{\boldsymbol{x}, \boldsymbol{s}} \mathcal{W}(\boldsymbol{x} + \boldsymbol{s})\mathcal{S}(\boldsymbol{s})}{2\pi(s_i^3 - s_{i-1}^3)(\mu_j - \mu_{j-1})} \approx \frac{3\alpha^2 \sum_{ab} \bar{n}(\boldsymbol{x}_a) w(\boldsymbol{x}_a)^2 w(\boldsymbol{x}_b)^2}{2\pi(s_i^3 - s_{i-1}^3)(\mu_j - \mu_{j-1})}. \tag{2.29}$$

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

$$\mu = \frac{\hat{\boldsymbol{s}} \cdot (\boldsymbol{x} + \boldsymbol{s}/2)}{|\boldsymbol{x} + \boldsymbol{s}/2|}. \tag{2.30}$$

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 (2.28). 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 (2.7) by replacing the integral with a summation over a random catalog, e.g.

$$\mathcal{W}_0 \approx \alpha \sum_a \bar{n}(\boldsymbol{x}_a) w(\boldsymbol{x}_a)^2. \tag{2.31}$$

We have seen in (2.9) 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 [58] with the mask [59] before fitting it to the data (more detailed description in Sec. 5.1 and illustration in Fig. 2). We then use the model before convolution to compute $\mathrm{Cov}^{\mathrm{disc}}$.

2.4 Corollary Covariances

We generalize our formalism below for more applications, and refer the readers to App. A for detailed derivations.

2.4.1 Cross Correlation and Cross Covariance

In Sec. 2.1 and 2.2 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 derive the equations for cross covariance of cross 3D power spectra in App. A.1. The derivation parallels that of Sec. 2.1 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.

2.4.2 Angular Power

We can generalize the formalism developed in Sec 2.2 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. We derive the equations in App. A.2, assuming flat sky and Limber approximation. Compared to the multipole case, the angular window effect is captured by some $\mathcal{Q}$ windows similar to (2.28) 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. 2.3. Our method can be a fast alternative to the calculation on the sphere as is generally done for the CMB measurements [60].

2.4.3 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 [61]. We derive the auto covariance for this case in App. A.3, and also its cross covariance with power spectrum multipoles, which is particularly useful for their joint analysis aimed at testing theories of gravity [8, 62]. 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. We also derive alternative equations for the auto projected covariance similar to the angular case in App. A.2.

2.4.4 Correlation Functions

In configuration space, the Landy-Szalay estimator [63] is typically employed to measure the correlation function, and it can be written in our notation as

$$\hat{\xi}(\boldsymbol{s}) = \frac{1}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{x}} \delta_W(\boldsymbol{x}) \delta_W(\boldsymbol{x} + \boldsymbol{s}), \tag{2.32}$$

where we have defined the normalization factor analogous to (2.16)

$$\mathcal{Q}_W(\boldsymbol{s}) \equiv \int_{\boldsymbol{x}} W(\boldsymbol{x} + \boldsymbol{s}) W(\boldsymbol{x}). \tag{2.33}$$

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

$$\begin{aligned}
\langle \hat{\xi}(\boldsymbol{s}) \rangle &= \frac{1}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{x}} W(\boldsymbol{x}) W(\boldsymbol{x} + \boldsymbol{s}) \langle \delta(\boldsymbol{x}) \delta(\boldsymbol{x} + \boldsymbol{s}) \rangle \\
&\equiv \frac{\xi(\boldsymbol{s})}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{x}} W(\boldsymbol{x}) W(\boldsymbol{x} + \boldsymbol{s}) = \xi(\boldsymbol{s}),
\end{aligned} \tag{2.34}$$

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

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:

$$\begin{aligned}
\xi(\boldsymbol{s}) &= \int_{\boldsymbol{k}} P(\boldsymbol{k}) e^{i\boldsymbol{k} \cdot \boldsymbol{s}}, \\
\hat{\xi}(\boldsymbol{s}) &= \frac{\mathcal{W}_0}{\mathcal{Q}_W(\boldsymbol{s})} \int_{\boldsymbol{k}} \hat{P}(\boldsymbol{k}) e^{i\boldsymbol{k} \cdot \boldsymbol{s}}.
\end{aligned} \tag{2.35}$$

Here $\mathcal{W}_0$ is equal to the value of $\mathcal{Q}_W(\boldsymbol{s} = 0)$ according to (2.14) and (2.33). So the covariance of the two 3D estimators are related by

$$\text{Cov}\big[\hat{\xi}(\boldsymbol{s}), \hat{\xi}(\boldsymbol{s}')\big] = \frac{\mathcal{W}_0^2}{\mathcal{Q}_W(\boldsymbol{s})\mathcal{Q}_W(\boldsymbol{s}')} \int_{\boldsymbol{k},\boldsymbol{k}'} \text{Cov}\big[\hat{P}(\boldsymbol{k}), \hat{P}(\boldsymbol{k}')\big] e^{i\boldsymbol{k}\cdot\boldsymbol{s}} e^{i\boldsymbol{k}'\cdot\boldsymbol{s}'}. \qquad (2.36)$$

We are more interested in the correlation function covariances cast in the multipole, angular, and projected form. We develop their relations to the corresponding power spectrum covariances in App. A.4.

2.4.5 Wedges

In this paper and especially Sec. 2.2, 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 $i = 1, \ldots, N_{\text{wedge}}$. The power spectrum wedges are related to the multipoles by

$$\hat{P}(k, \mu_i) = \int_{\mu_{i-\frac{1}{2}}}^{\mu_{i+\frac{1}{2}}} \frac{\mathrm{d}\mu}{\mu_{i+\frac{1}{2}} - \mu_{i-\frac{1}{2}}} \Big[\sum_{\ell=0}^{\ell_{\max}} \hat{P}_\ell(k)\,\mathcal{L}_\ell(\mu)\Big] = \sum_{\ell=0}^{\ell_{\max}} \hat{P}_\ell(k)\bar{\mathcal{L}}_\ell(\mu_i), \qquad (2.37)$$

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

$$\bar{\mathcal{L}}_\ell(\mu_i) \equiv \int_{\mu_{i-\frac{1}{2}}}^{\mu_{i+\frac{1}{2}}} \frac{\mathcal{L}_\ell(\mu)\,\mathrm{d}\mu}{\mu_{i+\frac{1}{2}} - \mu_{i-\frac{1}{2}}}. \qquad (2.38)$$

Therefore

$$\text{Cov}^{\text{disc}}\big[\hat{P}(k, \mu_i), \hat{P}(k', \mu_j)\big] = \sum_{\ell=0}^{\ell_{\max}} \sum_{\ell'=0}^{\ell_{\max}} \text{Cov}^{\text{disc}}\big[\hat{P}_\ell(k), \hat{P}_{\ell'}(k')\big] \bar{\mathcal{L}}_\ell(\mu_i)\bar{\mathcal{L}}_{\ell'}(\mu_j). \qquad (2.39)$$

The same relation applies to the correlation functions[4] by replacing $P \to \xi$, $k \to s$.

2.5 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. 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 $i = 1, \ldots, N_{\text{bin}}$. In accordance, we integrate the analytic prediction within spherical k-shells

$$\text{Cov}\big[\hat{P}(k_i), \hat{P}(k_j)\big] = \int_{k_{i-\frac{1}{2}}}^{k_{i+\frac{1}{2}}} \frac{4\pi k^2 \mathrm{d}k}{V_{k_i}} \int_{k_{j-\frac{1}{2}}}^{k_{j+\frac{1}{2}}} \frac{4\pi k'^2 \mathrm{d}k'}{V_{k_j}} \text{Cov}\big[\hat{P}(k), \hat{P}(k')\big], \qquad (2.40)$$

where $V_{k_i} = 4\pi(k_{i+\frac{1}{2}}^3 - k_{i-\frac{1}{2}}^3)/3$ 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.[5]

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.

---

[4] 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.

[5] We package and release this simple and general-purpose utility at https://github.com/eelregit/avgem.
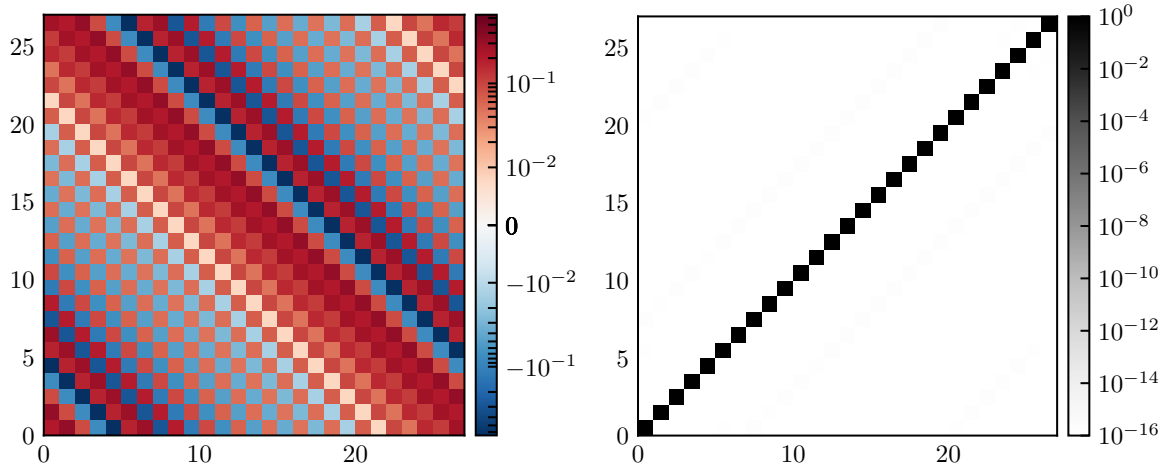
Figure 1. An example 27×27 Hankel matrix H_0, for the J_0 Hankel transform. We visualize it in the left panel, and demonstrate in the right panel that the squared Hankel matrix is an identity matrix. As a circulant matrix, a Hankel matrix carries out the convolution operation in the FFTLog algorithm (Eq. 3.3 and 3.5). It is also the building block of the double Bessel quadrature method (Eq. 3.1 and 3.10).

3 Double Bessel Quadrature

Eq. (2.25), (2.26), and (2.27) all have integrals involving two spherical Bessel functions. More generally, we are interested in evaluating the following integral with Bessel functions

$$G(y, y') = \int_0^\infty x \mathrm{d}x \ F(x) J_\nu(xy) J_{\nu'}(xy'), \tag{3.1}$$

which proves useful also for the applications in Sec.2.4. This applies to (2.25), (2.26), and (2.27) given

$$j_n(t) = \sqrt{\frac{\pi}{2t}} J_{n+\frac{1}{2}}(t). \tag{3.2}$$

Usual quadrature methods struggle to converge for (3.1) because the Bessel kernels are highly oscillatory and damp slowly. Recent solutions proposed by Ref. [64, 65] generalized the FFTLog algorithm [66, 67] to perform the integral transform from $F(x)$ to $G(y, y')$ for each fixed ratio y'/y.

Here we present a fast and simple algorithm that potentially suits the covariance calculation better than that in Ref. [64, 65]. First let's look at a simpler integral with only one Bessel function

$$B(x) = \int_0^\infty y \mathrm{d}y \ A(y) J_\nu(xy), \tag{3.3}$$

known as the Hankel transform. We can perform integral transforms like this one using the `mcfit` [68] 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 (3.3)

$$xB(x) = \int_0^\infty \frac{dy}{y} \left[yA(y)\right]\left[xyJ_\nu(xy)\right], \tag{3.4}$$

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 (3.4) 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

$$B_i = x_i^{-1} \sum_{j=1}^N H_{\nu,ij} y_j A_j, \tag{3.5}$$

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. [67]. As an example we show in Fig. 1 a 27×27 H_0 matrix with $\Delta = 0.34$. The above rescaling in (3.4) 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. 1 by the nearly perfect agreement between H_0^2 and the identity matrix to machine precision.

Notice that formally (3.5) can be obtained by applying the following replacement rule on (3.4) or (3.3):

$$xyJ_\nu(xy) \longrightarrow \frac{H_{\nu,ij}}{\Delta}, \qquad \int \frac{dy}{y} \longrightarrow \Delta \sum_j . \tag{3.6}$$

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 (3.1) also follows from this rule.

Now we are ready to derive our double Bessel quadrature algorithm. Consider the following integral that further integrates (3.1) with some arbitrary auxiliary functions $A(y)$ and $A'(y')$ over y and y'

$$I = \int_0^\infty y dy \int_0^\infty y' dy' \ A(y) G(y, y') A'(y'). \tag{3.7}$$

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 (3.3) to (3.5), so that
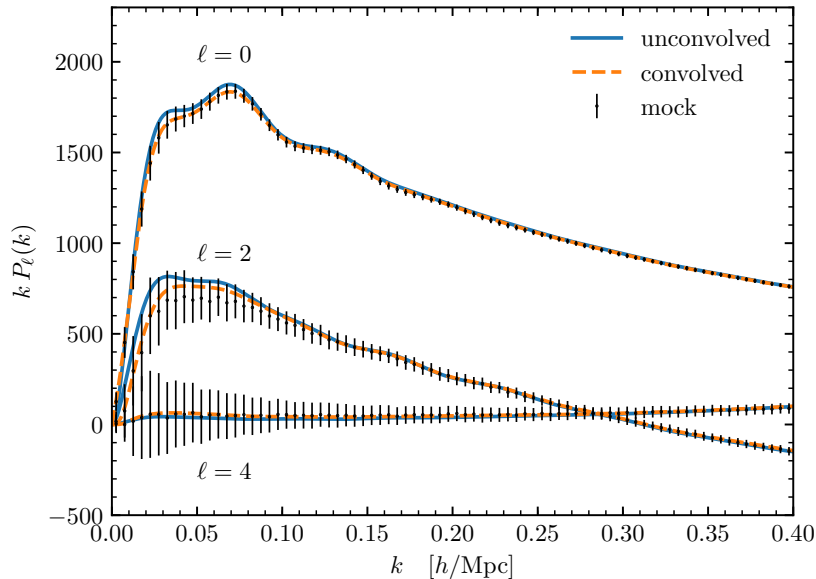
$$I = \int_0^\infty x dx \ F(x) B(x) B'(x) \approx \Delta \sum_i F_i \sum_j H_{\nu,ij} y_j A_j \sum_{j'} H_{\nu',ij'} y_{j'} A'_{j'}. \tag{3.8}$$

Alternatively, a direct discretization of (3.7) reads

$$I \approx \Delta^2 \sum_{jj'} y_j^2 A_j G_{j,j'} y_{j'}^2 A'_{j'}. \tag{3.9}$$

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

$$G_{j,j'} = \frac{1}{\Delta} \sum_i y_j^{-1} H_{\nu,ij} F_i H_{\nu',ij'} y_{j'}^{-1}. \tag{3.10}$$

Figure 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 [58], including the convolution with the appropriate window function [59] 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. 3 to compute the covariance matrices.

This result can also be obtained, formally, by applying the same replacement rule (3.6) to (3.1).

Evaluation of (3.10) 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. [64, 65].

Furthermore, (3.9) 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. 2.5) rather than the finer structure, and (3.10) allows us to perform the numerical integration without scanning a very fine grid. On the other hand, the method in Ref. [64, 65] 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 Simulations

To verify our analytic covariances, we compare them to the empirical results measured from $N_{\rm mock} = 1000$ realizations of the Multidark-Patchy galaxy mocks [69], produced for BOSS
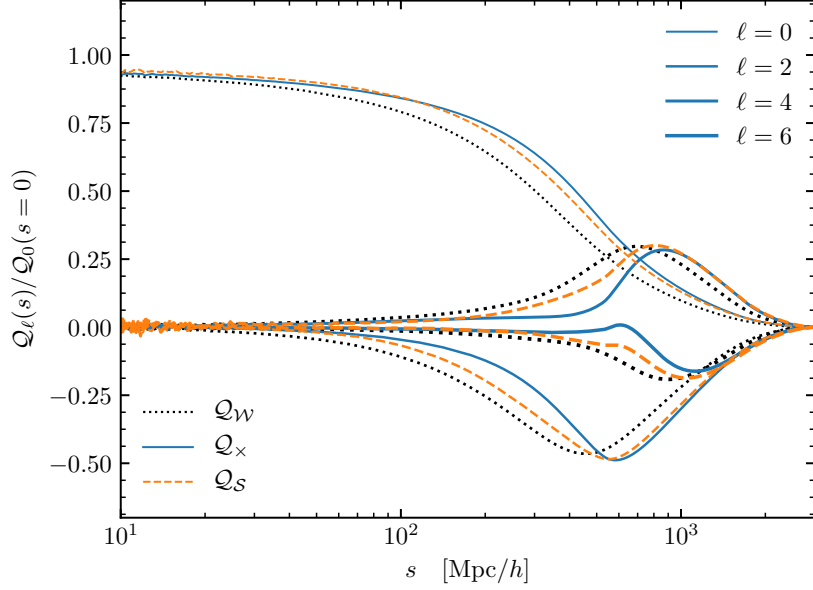
Figure 3. Window function multipoles, of 1000 Patchy mocks for NGC of BOSS DR12 in $0.5 < z < 0.75$. The $\mathcal{Q}$ window functions are directly related to the covariance, defined in (2.15), (2.16) and (2.28). We normalize the window multipoles by their zero-lag monopoles to compare their shapes. As the covariance is $\sim (P + P_{\text{shot}})^2$, three $\mathcal{Q}$ factors in different colors describe the windows for different pieces: $\mathcal{Q}_{\mathcal{W}}$ for P^2, $\mathcal{Q}_{\times}$ for $P \times P_{\text{shot}}$, and $\mathcal{Q}_{\mathcal{S}}$ for P_{shot}^2. 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. 2 and the $\mathcal{Q}_\ell$ here, we can compute the disconnected covariance of $\hat{P}_\ell$ following Sec. 2.2.

data release 12 (DR12) [6]. 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 match the observed galaxy clustering and its redshift evolution. In this work 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 $\Omega_{\text{m}} = 0.307$, $\Omega_{\text{b}} = 0.048$, $h = 0.678$, $\sigma_8 = 0.829$.

We estimate the redshift-space power spectrum multipoles with the endpoint Yamamoto estimator [51] in the third redshift bin. We employ the FFT-accelerated algorithm [52] enabled by multipole decomposition [70], implemented in the large-scale structure toolkit `nbodykit` [71]. We use the standard Landy-Szalay [61, 63] 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. [61] for correlation function with dimensions of length).

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

matrix from the N_{mock} realizations is

$$\widehat{\text{Cov}}(O, O') = \frac{1}{N_{\text{mock}} - 1} \sum_{r=1}^{N_{\text{mock}}} (\hat{O}_r - \bar{O})(\hat{O}_r' - \bar{O}'), \qquad (4.1)$$

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

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

$$w = w_{\text{FKP}} \times w_{\text{veto}}, \qquad (4.2)$$

in which $w_{\text{FKP}} \equiv 1/(1 + \bar{n} \times 10^4 \, \text{Mpc}^3/h^3)$, 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 $\alpha \lesssim 0.2$.

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.

5.1 Power Spectrum Multipole Covariance Matrix

We compute the analytic disconnected covariance of $P_\ell(k)$ using the equations in Sec. 2.2, 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. 2. With all model parameters initialized to their fiducial values, we use the RSD model from [58] to compute a diagonal covariance matrix [72] and fit this model to the mock sample mean over the k range from $0.02 \, h/\text{Mpc}$ to $0.4 \, h/\text{Mpc}$. 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 (2.9), we convolve the model with the appropriate window [59] before fitting it to the mocks. We show both the unconvolved and convolved model curves in Fig. 2.

To measure the window functions represented by their multipole moments, as defined in (2.15), (2.16), and (2.28), we use the random catalog of Patchy mocks, and follow the procedure described in Sec. 2.3. As shown from (2.13) to (2.17), because the covariance has a quadratic form in power spectrum and its shot noise, i.e. $\sim (P + P_{\text{shot}})^2$, the three different window factors describe the covariance shape of different pieces: $\mathcal{Q}_{\mathcal{W}}$ for P^2, $\mathcal{Q}_\times$ for $P \times P_{\text{shot}}$, and $\mathcal{Q}_{\mathcal{S}}$ for P_{shot}^2. They capture the same survey geometry but differ in the weights, and therefore have similar shapes and different normalizations. Our analytic Cov^{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 $\ell_{\text{max}} = 10$. In Fig. 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,
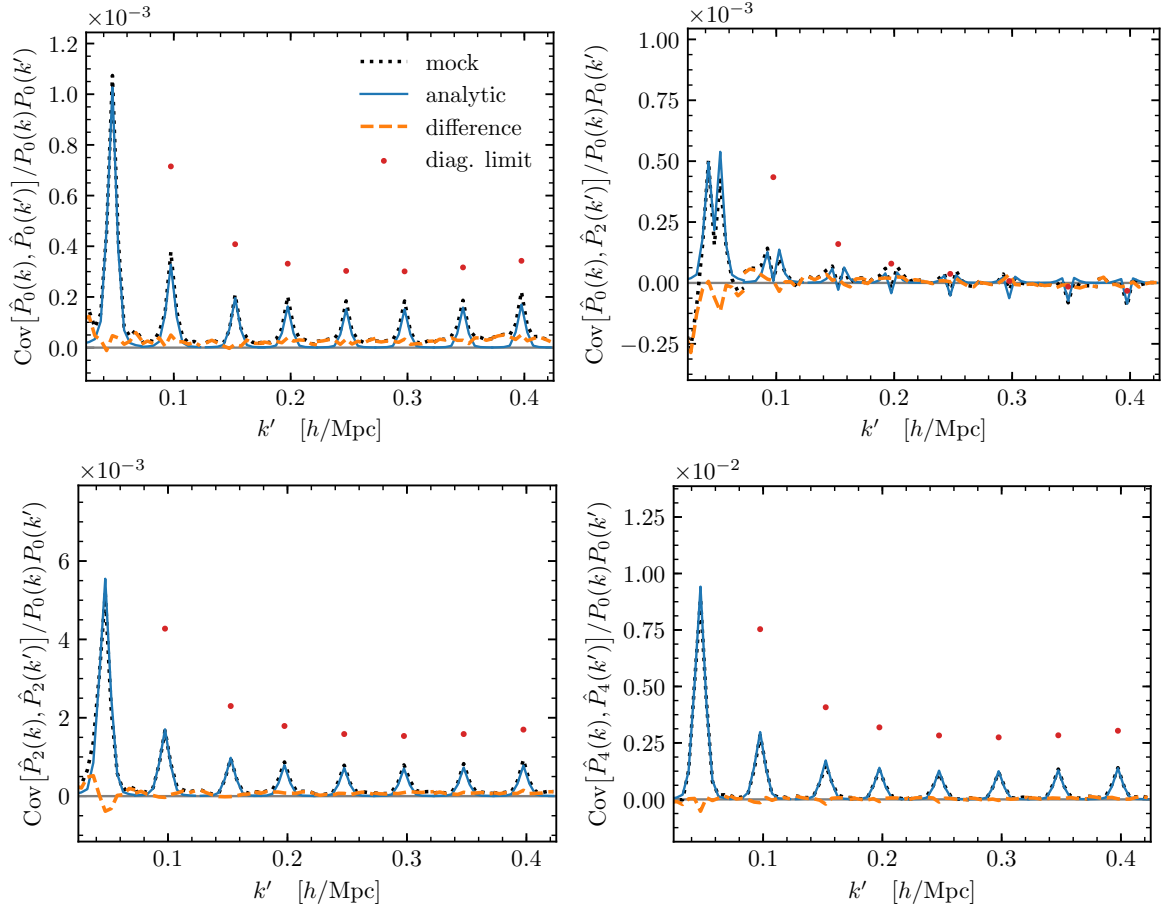
Figure 4. Slices of covariance matrices of power spectrum multipoles, of bin width $0.005\,h/\mathrm{Mpc}$, normalized by the monopole. Different panels shows the auto and cross covariance matrices of different multipoles, as labeled next to the vertical axes. Every spike corresponds to a slice of covariance 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 (2.20) 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.

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. 2.2. The equations involve double Bessel integrals that we compute using the quadrature method introduced in Sec. 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. 2.5. The computation takes 2 minutes on a single CPU.

In Fig. 4, we compare our analytic covariance for power spectrum multipoles, $P_0, P_2, P_4,$

to the mock sample covariance computed using eq. (4.1). 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. 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:

$$\widehat{\mathrm{Cov}}^{\mathrm{conn}} = \widehat{\mathrm{Cov}} - \mathrm{Cov}^{\mathrm{disc}}. \tag{5.1}$$

We also show the differences in mock and analytic covariance in Fig. 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. 5 we present the shape of the full covariance matrix by normalizing it as the linear correlation coefficients given by

$$\mathrm{Corr}(O, O') = \frac{\mathrm{Cov}(O, O')}{\sqrt{\mathrm{Cov}(O, O)\,\mathrm{Cov}(O', O')}}, \tag{5.2}$$

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 (5.1) by the diagonal of the analytic disconnected covariance

$$\mathrm{Corr}^{\mathrm{conn}}(O, O') = \frac{\widehat{\mathrm{Cov}}^{\mathrm{conn}}(O, O')}{\sqrt{\mathrm{Cov}^{\mathrm{disc}}(O, O)\,\mathrm{Cov}^{\mathrm{disc}}(O', O')}}, \tag{5.3}$$

and show it in the lower panel of Fig. 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, 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 (5.1) 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 [29, 31]. We perform principal component analysis on $\mathrm{Corr}^{\mathrm{conn}}$ in (5.3), 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 $\mathrm{Corr^{conn}}$ with this rank 4 substitute, and rescale it by the denominator in (5.3) 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

$$\widehat{\mathrm{Cov}}^{\mathrm{disc}} = \widehat{\mathrm{Cov}} - \widehat{\mathrm{Cov}}^{\mathrm{conn,sm}}. \tag{5.4}$$

This mock $\widehat{\mathrm{Cov}}^{\mathrm{disc}}$ can be compared directly to the analytic $\mathrm{Cov^{disc}}$, which we present in Fig. 6 as slices across the correlation matrix. The slices are arranged in the same way as in Fig. 4. Also shown is the difference between the analytic and mock results, with everything normalized by the analytic $\mathrm{Cov^{disc}}$ as in (5.3) so that the comparison is fair. This residual contains possible flaw in the analytic $\mathrm{Cov^{disc}}$ 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 $\mathrm{Cov^{disc}}$ on the $\mathcal{Q}$ windows. In this paper 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. 3), one possible simplification is to approximate all $\mathcal{Q}$ windows with the same shape. In practice this can be achieved in (2.13) 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. 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 $\mathrm{Cov^{disc}}$, especially when shot noise is significant.
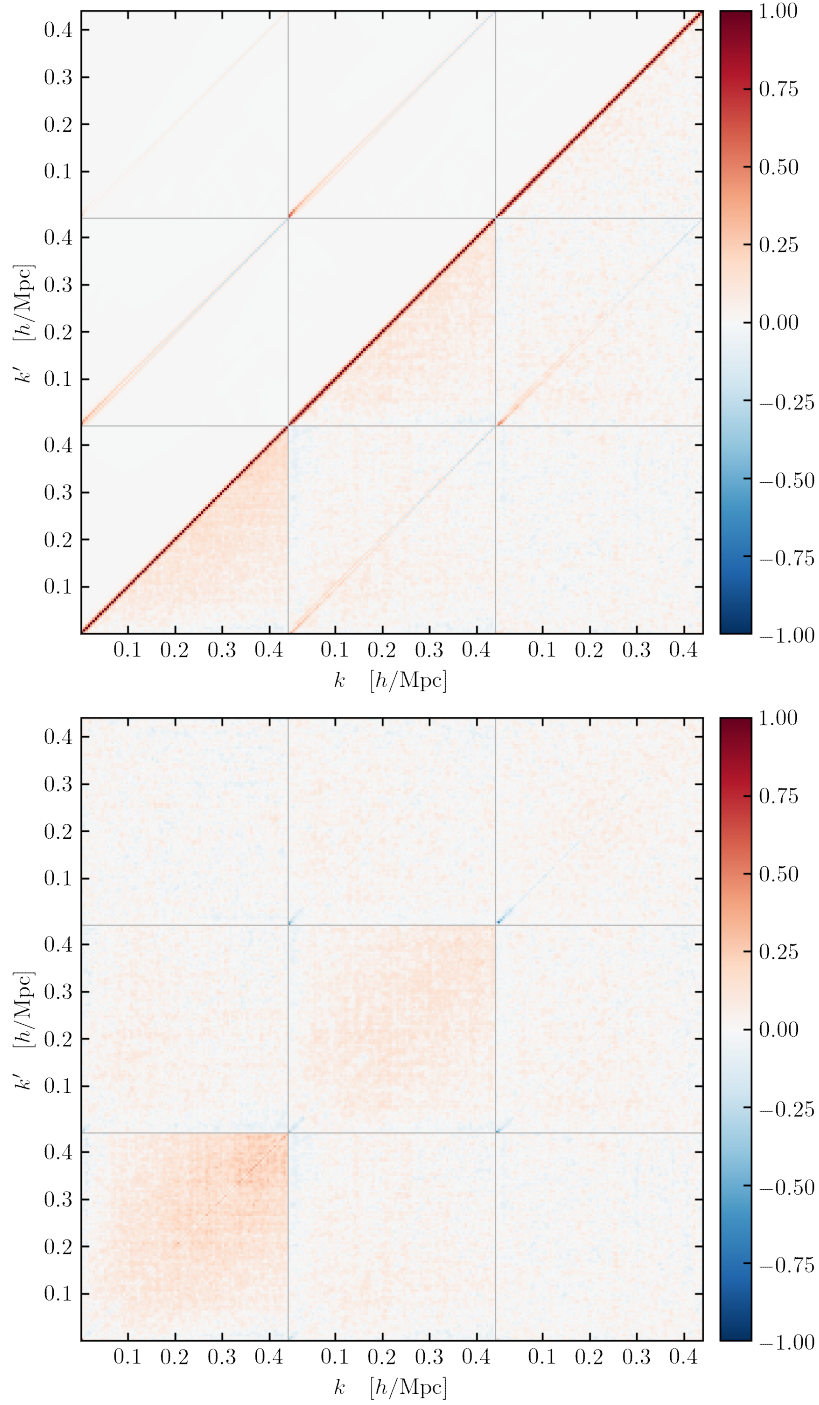
Figure 5. Correlation matrix of the power spectrum multipoles. In the 3×3 blocks, from bottom to top and from left to right, we visualize the auto and cross correlations of $\hat{P}_0$, $\hat{P}_2$, $\hat{P}_4$. The top panel compares the analytic result in its upper triangular block with the mocks covariance 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, 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.
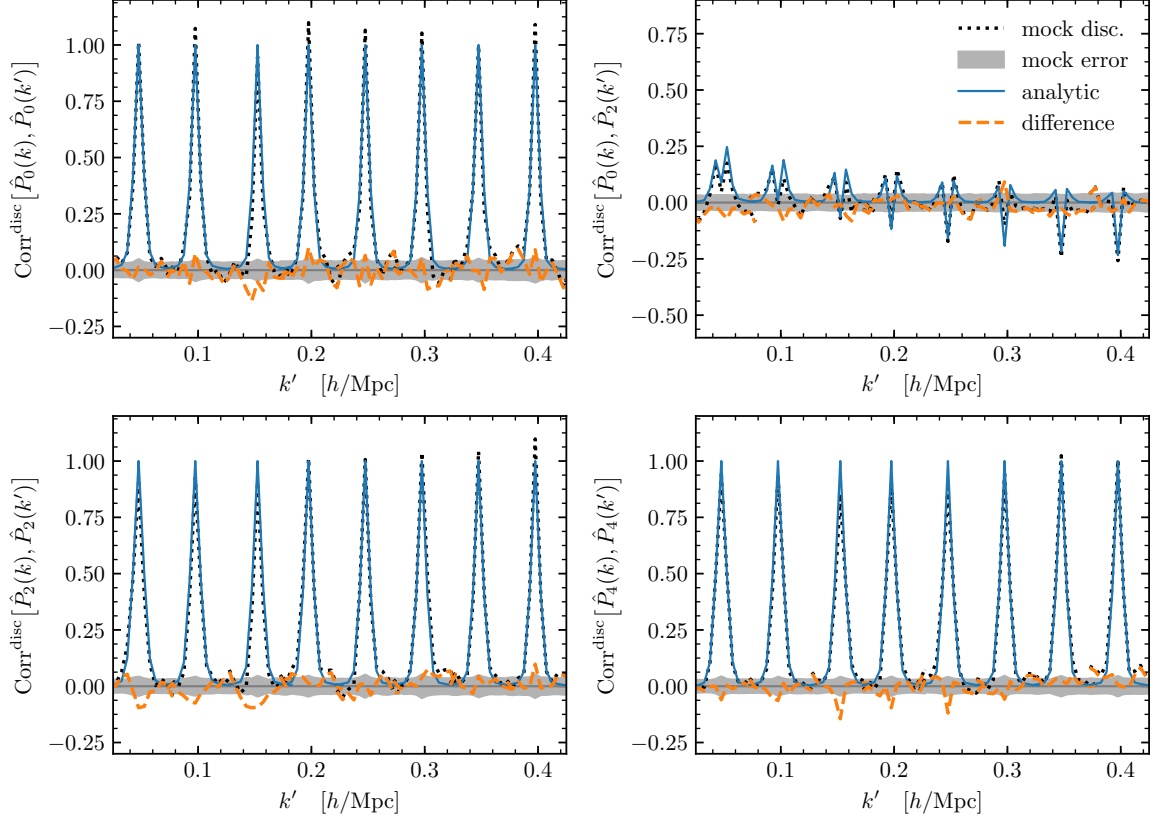
Figure 6. Slices of disconnected correlation matrix of power spectrum multipoles. Every spike corresponds to a slice of covariance matrix near the diagonal ($k' \approx k$) at fixed k (marked by the position of the peak). For fair comparison, all covariance matrices are normalized by the diagonal of 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. 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).
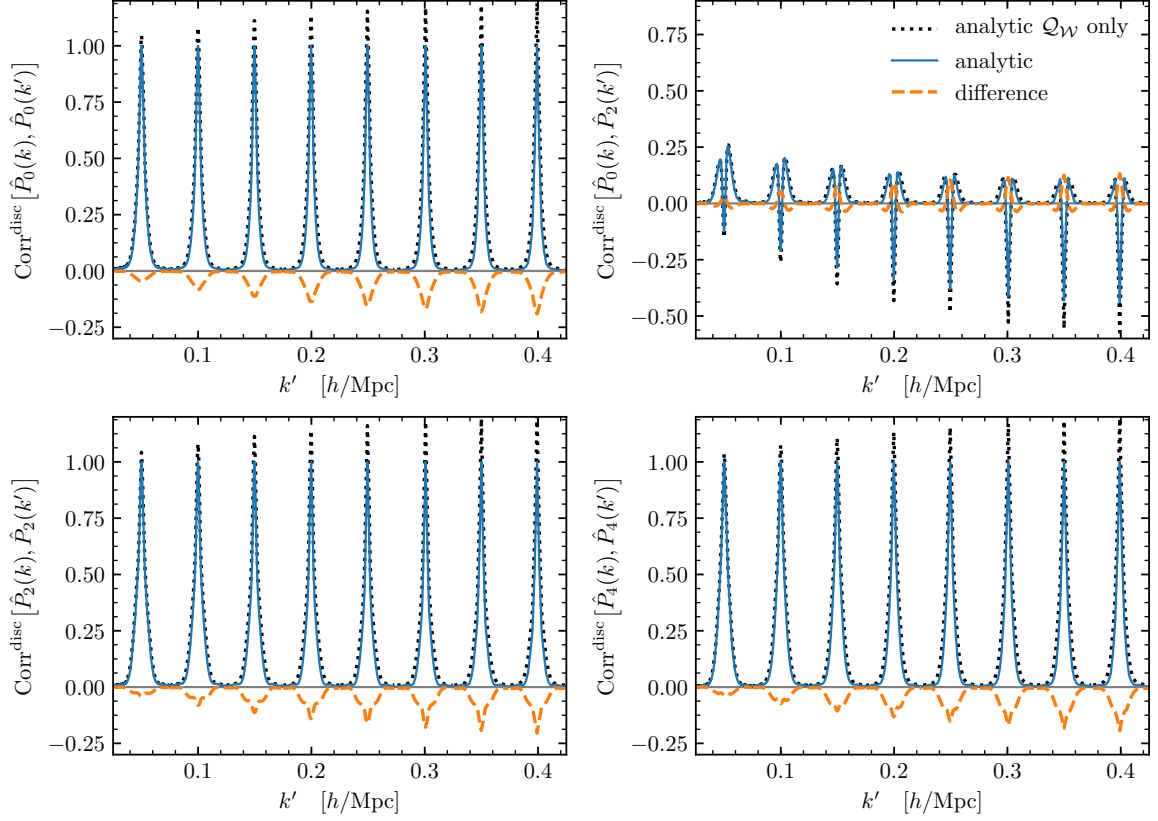
Figure 7. Slices of analytic disconnected correlation matrix of power spectrum multipoles, using different $\mathcal{Q}$ functions. The solid blue lines show the same but unbinned result as in previous figures, with which we normalize all the covariance in this figure for fair comparison. It was computed with the window functions $\mathcal{Q}_{\mathcal{W}}$, $\mathcal{Q}_{\times}$, and $\mathcal{Q}_{\mathcal{S}}$, which are shown in Fig. 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.
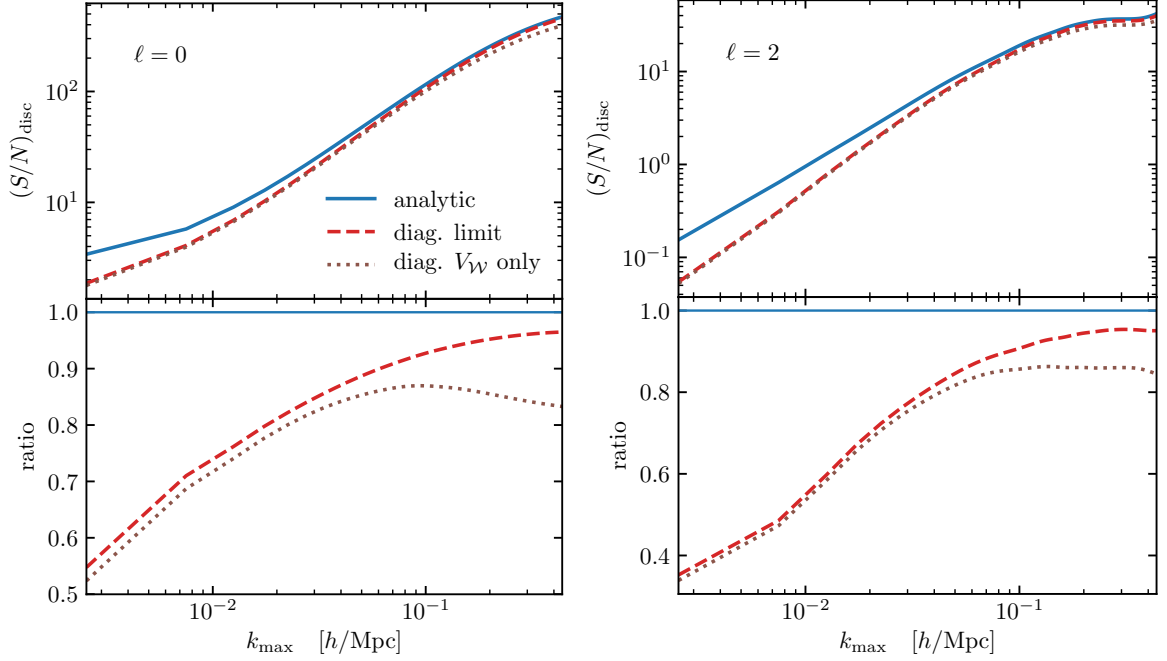
5.1.1 Signal-to-Noise Ratio

Our analytic disconnected covariance $\mathrm{Cov}^{\mathrm{disc}}$ captures the correlation between neighboring k bins and multipoles due to the window effect. In Sec. 2.1.1 we have derived its diagonal limit $\mathrm{Cov}^{\mathrm{diag}}$, that only captures the survey size but neglect its inhomogeneity and anisotropy. Fig. 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

$$\left(\frac{S}{N}\right)^2_{\mathrm{disc}} = \sum_{k_i, k_j < k_{\mathrm{max}}} P_\ell(k_i)\, \mathrm{Cov}^{\mathrm{disc}}\left[\hat{P}(k_i), \hat{P}(k_j)\right]^{-1} P_\ell(k_j), \tag{5.5}$$

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. 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 $0.1\, h/\mathrm{Mpc}$, and $\approx 5\%$ even between $0.2\, h/\mathrm{Mpc}$ and $0.4\, h/\mathrm{Mpc}$. 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. 3. We also show that it further biases S/N on small scales if one ignores the differences among the three effective volumes in (2.18) 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 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 Cov^{disc} and its diagonal limit (Sec. 2.1.1). The latter captures the window size with three effective volumes in (2.18), but ignores the shape of the window, resulting in very different covariance matrices (see Fig. 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. 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.

5.2 Projected-Multipole Correlation Function Covariance Matrix

Using the equations derived in Sec. 2.4.4 and App. A.4, 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 [61] 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 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. 5.1, to compute the auto covariance of the projected power spectrum $P_\perp$, and the cross covariance between $P_\perp$ and P_ℓ's, following App. A.3. Together with the auto covariance of P_ℓ's obtained in Sec. 5.1, 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
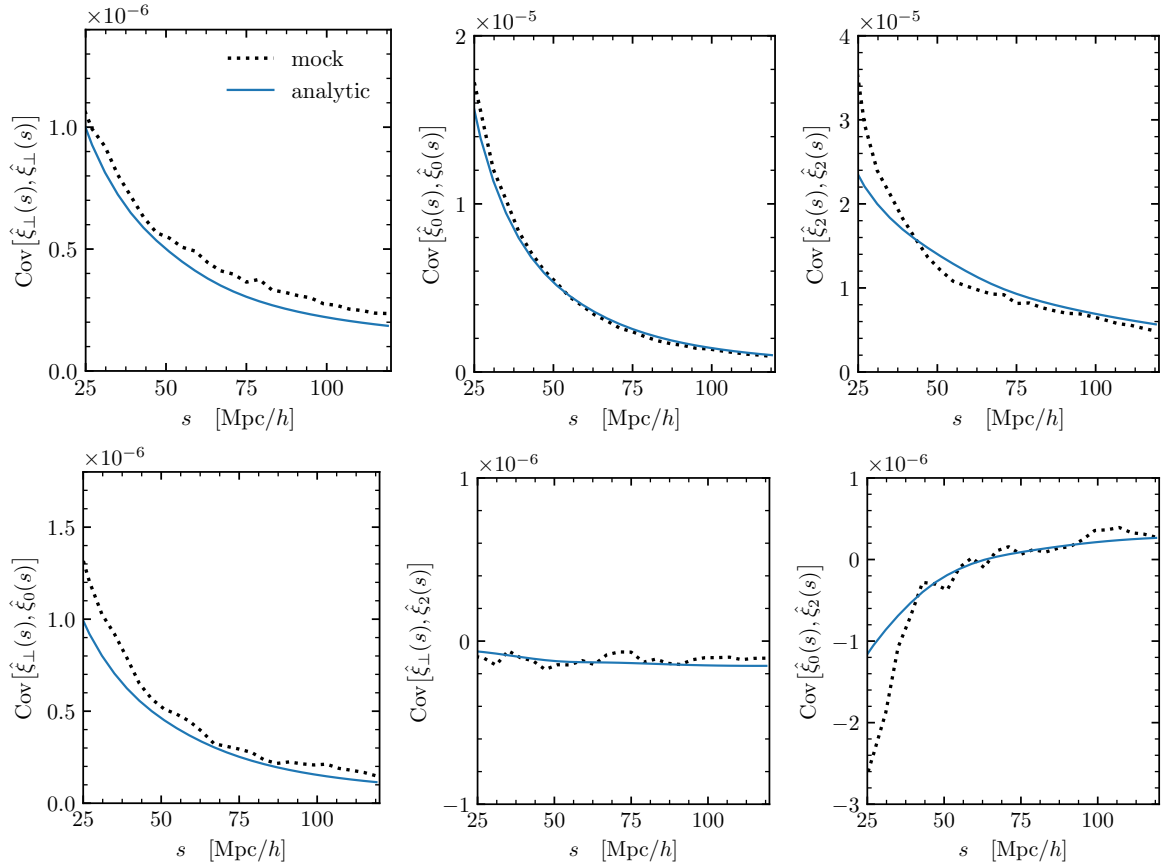
Figure 9. The diagonal of the joint covariance matrices of the projected correlation function and correlation function multipoles. We verify that the small-scale deviations in the ξ_ℓ covariances are due to the connected covariance.

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 following App. A.4, 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 $\ell_{\mathrm{max}} = 16$, 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 $\mathrm{Cov}^{\mathrm{disc}}$ to the sample covariance from mocks. Fig. 9 shows this comparison on the $s = s'$ diagonal of the auto covariance matrices. We find the analytic $\mathrm{Cov}^{\mathrm{disc}}$ 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 $\widehat{\mathrm{Cov}}^{\mathrm{conn,sm}}$ from Sec. 5.1. 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 $\mathrm{Cov}^{\mathrm{disc}}$ 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.
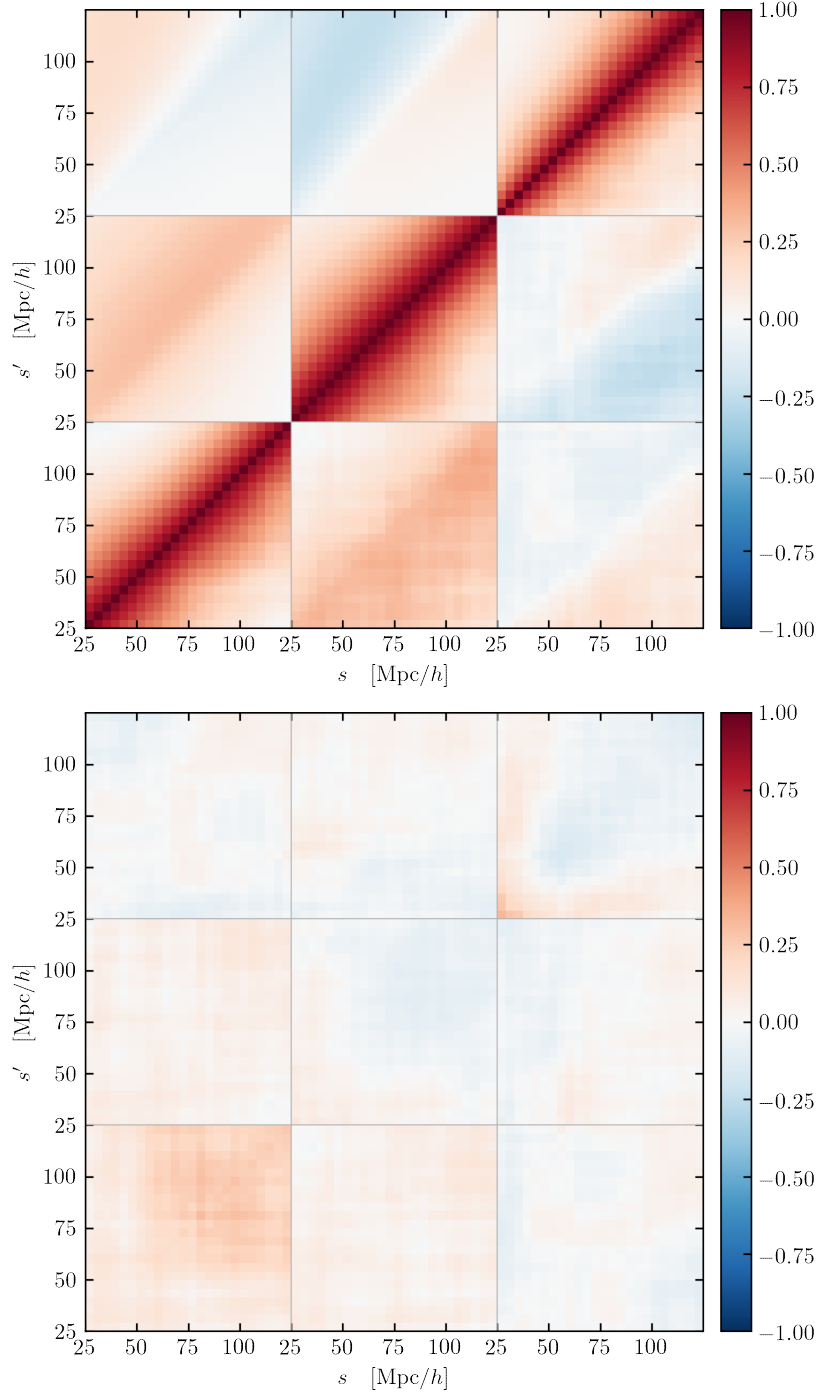
Figure 10. Correlation matrix of of the projected correlation function and correlation function multipoles. In the 3×3 blocks, from bottom to top and from left to right, we visualize the auto and cross correlations of $\hat{\xi}_\perp$, $\hat{\xi}_0$, $\hat{\xi}_2$. The upper figure compares the analytic 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.

We also compare the shapes of the analytic and mock covariance by showing their

correlation matrices in Fig. 10. The top panel shows the two have very similar shapes, and their difference (normalized by the diagonal of the analytic Cov$^{\text{disc}}$ similar to Fig. 5) is $\lesssim 10\%$ except in the regions noted above.

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. 2.1, App. A.1) and correlation function (Sec. 2.4.4, App. A.4), and applies to the covariances for various probes including the multipoles (Sec. 2.2) and the wedges (Sec. 2.4.5) of 3D clustering, the angular (App. A.2) and the projected statistics of clustering and lensing (App. A.3), as well as their cross covariances (App. A.1, App. A.3).

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 [17, 19].

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 [72] 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. [73], 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. [74] 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. [74] are more complicated and need Monte-Carlo integration to evaluate. Ref. [61] 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.

Acknowledgments

A   Disconnected Covariance

In this appendix we present calculations of the disconnected auto and cross covariance of power spectra of various forms, including the general 3D power spectra, the multipoles, the angular power spectra.

A.1   Fully General 3D Case

As usual we model the tracer fields as a Poisson process that samples a *homogeneous* random field. Under such assumption, the cross correlation between two fields a and b is given by

$$
\left\langle \delta^a_{\rm data}(\boldsymbol{k})\delta^b_{\rm data}(-\boldsymbol{k}')\right\rangle = (2\pi)^3\,\delta^{\rm D}(\boldsymbol{k}-\boldsymbol{k}')P^{ab}(\boldsymbol{k}) + \int_{\boldsymbol{x}} \frac{\bar{n}_{ab}(\boldsymbol{x})}{\bar{n}_a(\boldsymbol{x})\bar{n}_b(\boldsymbol{x})}e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}
$$

$$
\left\langle \delta^a_{\rm rand}(\boldsymbol{k})\delta^b_{\rm rand}(-\boldsymbol{k}')\right\rangle = \int_{\boldsymbol{x}} \frac{\alpha\delta^{\rm K}_{ab}}{\bar{n}_a(\boldsymbol{x})}e^{-i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}
$$

$$
\left\langle \delta^a_{\rm data}(\boldsymbol{k})\delta^b_{\rm rand}(-\boldsymbol{k}')\right\rangle = \left\langle \delta^b_{\rm data}(\boldsymbol{k})\delta^a_{\rm rand}(-\boldsymbol{k}')\right\rangle = 0. \tag{A.1}
$$

For each field the subscript 'data' and 'rand' labels the data and random catalogs, respectively. And the overdensity field for analyses is usually the difference between the data and random overdensities

$$
\delta \equiv \delta_{\rm data} - \delta_{\rm rand}. \tag{A.2}
$$

The shot noise term is related to $\bar{n}_{ab}(\boldsymbol{x})$, the mean number density of the overlap between the two samples. Apparently $\bar{n}_{ab}$ reduces to $\bar{n}_a$ when $a = b$, and the cross shot noise is at most as big as the minimum of the auto shot noises of the two samples: $\bar{n}_{ab}/(\bar{n}_a\bar{n}_b) \leq \min(1/\bar{n}_a, 1/\bar{n}_b)$.

The windowed field in Fourier space is a convolution

$$
\delta^a_W(\boldsymbol{k}) = \int_{\boldsymbol{q}} \delta^a(\boldsymbol{k}-\boldsymbol{q})W^a(\boldsymbol{q}). \tag{A.3}
$$

The following estimator of 3D power spectrum measures the cross correlation between a and b fields

$$\hat{P}^{ab}(\mathbf{k}) = \frac{\delta_W^a(\mathbf{k})\delta_W^b(-\mathbf{k})}{\mathcal{W}^{ab}(0)} - P_{\text{shot}}^{ab}. \tag{A.4}$$

This is a Hermitian function that satisfies $\hat{P}^{ab}(\mathbf{k})^* = \hat{P}^{ab}(-\mathbf{k}) = \hat{P}^{ba}(\mathbf{k})$, and a real function when $a = b$. The auto and cross covariance functions of the estimators

$$\text{Cov}\big[\hat{P}^{ab}(\mathbf{k}), \hat{P}^{cd}(\mathbf{k}')\big] \equiv \big\langle \hat{P}^{ab}(\mathbf{k})\hat{P}^{cd}(\mathbf{k}')^* \big\rangle - \big\langle \hat{P}^{ab}(\mathbf{k}) \big\rangle \big\langle \hat{P}^{cd}(\mathbf{k}')^* \big\rangle \tag{A.5}$$

has the following disconnected part

$$\text{Cov}^{\text{disc}}\big[\hat{P}^{ab}(\mathbf{k}), \hat{P}^{cd}(\mathbf{k}')\big] = \frac{\big\langle \delta_W^a(\mathbf{k})\delta_W^c(-\mathbf{k}') \big\rangle \big\langle \delta_W^b(-\mathbf{k})\delta_W^d(\mathbf{k}') \big\rangle}{\mathcal{W}^{ab}(0)\mathcal{W}^{cd}(0)} + (c, \mathbf{k}' \leftrightarrow d, -\mathbf{k}'). \tag{A.6}$$

Both covariance are Hermitian forms that satisfy $\text{Cov}^{ab,cd}(\mathbf{k}, \mathbf{k}')^* = \text{Cov}^{cd,ab}(\mathbf{k}', \mathbf{k})$. So the full Hermitian symmetry including those of the estimators are: $a, \mathbf{k} \leftrightarrow b, -\mathbf{k}$; $c, \mathbf{k}' \leftrightarrow d, -\mathbf{k}'$; and $a, b, \mathbf{k} \leftrightarrow c, d, -\mathbf{k}'$.

Plugging (A.1) into the following 2-point function

$$\big\langle \delta_W^a(\mathbf{k})\delta_W^b(-\mathbf{k}') \big\rangle$$
$$= \int_{\mathbf{q}} P^{ab}(\mathbf{k} - \mathbf{q})W^a(\mathbf{q})W^b(\mathbf{k} - \mathbf{k}' - \mathbf{q}) + (1 + \alpha\delta_{ab}^{\text{K}})\int_{\mathbf{x}} \bar{n}_{ab}(\mathbf{x})w_a(\mathbf{x})w_b(\mathbf{x})e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}$$
$$= \int_{\mathbf{k}''} P^{ab}(\mathbf{k}'')W^a(\mathbf{k} - \mathbf{k}'')W^b(\mathbf{k}'' - \mathbf{k}') + (1 + \alpha\delta_{ab}^{\text{K}})\int_{\mathbf{x}} \bar{n}_{ab}(\mathbf{x})w_a(\mathbf{x})w_b(\mathbf{x})e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}$$
$$\approx \frac{P^{ab}(\mathbf{k}) + P^{ab}(\mathbf{k}')}{2}\mathcal{W}^{ab}(\mathbf{k} - \mathbf{k}') + \mathcal{S}^{ab}(\mathbf{k} - \mathbf{k}'), \tag{A.7}$$

where the curly mode mixing windows $\mathcal{W}$ and $\mathcal{S}$ are define as

$$\mathcal{W}^{ab}(\mathbf{q}) = \int_{\mathbf{x}} \mathcal{W}^{ab}(\mathbf{x})e^{-i\mathbf{q}\cdot\mathbf{x}} \equiv \int_{\mathbf{x}} W^a(\mathbf{x})W^b(\mathbf{x})e^{-i\mathbf{q}\cdot\mathbf{x}},$$
$$\mathcal{S}^{ab}(\mathbf{q}) = \int_{\mathbf{x}} \mathcal{S}^{ab}(\mathbf{x})e^{-i\mathbf{q}\cdot\mathbf{x}} \equiv (1 + \alpha\delta_{ab}^{\text{K}})\int_{\mathbf{x}} \bar{n}_{ab}(\mathbf{x})w_a(\mathbf{x})w_b(\mathbf{x})e^{-i\mathbf{q}\cdot\mathbf{x}}. \tag{A.8}$$

In the second equality of (A.7), we have used change of variables to rewrite it in the more symmetric way. Under the assumption that the power spectrum is smooth, which is true at large k and k', we can make the approximation $P(\mathbf{k}) \approx P(\mathbf{k}'') \approx P(\mathbf{k}')$ to obtain the last line. This is due to the fact that the window functions are only significant when $\mathbf{k} \approx \mathbf{k}'' \approx \mathbf{k}'$ to the extent of the Fourier-space window scale. By replacing $P(\mathbf{k}'')$ with its average at $\mathbf{k}$ and $\mathbf{k}'$, our approximation preserves the Hermitian symmetry $a, \mathbf{k} \leftrightarrow b, -\mathbf{k}'$.

With (A.7) we can easily derive the following disconnected covariance

$$\text{Cov}^{\text{disc}}\big[\hat{P}^{ab}(\mathbf{k}), \hat{P}^{cd}(\mathbf{k}')\big] \approx \frac{1}{\mathcal{W}^{ab}(0)\mathcal{W}^{cd}(0)}\Bigg\{ \left[\frac{P^{ac}(\mathbf{k}) + P^{ac}(\mathbf{k}')}{2}\mathcal{W}^{ac}(\mathbf{k} - \mathbf{k}') + \mathcal{S}^{ac}(\mathbf{k} - \mathbf{k}')\right]$$
$$\times \left[\frac{P^{bd}(\mathbf{k}) + P^{bd}(\mathbf{k}')}{2}\mathcal{W}^{bd}(\mathbf{k} - \mathbf{k}') + \mathcal{S}^{bd}(\mathbf{k} - \mathbf{k}')\right]^*$$
$$+ \left[\frac{P^{ad}(\mathbf{k}) + P^{ad}(\mathbf{k}')^*}{2}\mathcal{W}^{ad}(\mathbf{k} + \mathbf{k}') + \mathcal{S}^{ad}(\mathbf{k} + \mathbf{k}')\right]$$
$$\times \left[\frac{P^{bc}(\mathbf{k}) + P^{bc}(\mathbf{k}')^*}{2}\mathcal{W}^{bc}(\mathbf{k} + \mathbf{k}') + \mathcal{S}^{bc}(\mathbf{k} + \mathbf{k}')\right]^*\Bigg\}. \tag{A.9}$$

Furthermore because we have assumed $P(\boldsymbol{k}) \approx P(\boldsymbol{k}')$, in the very same approximation we can combine certain terms by letting $P(\boldsymbol{k})P(\boldsymbol{k}) \approx P(\boldsymbol{k})P(\boldsymbol{k}') \approx P(\boldsymbol{k}')P(\boldsymbol{k}')$ to simplify the expression while still preserving its Hermitian symmetries

$$
\begin{aligned}
\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}^{ab}(\boldsymbol{k}), \hat{P}^{cd}(\boldsymbol{k}')\big] \approx \frac{1}{\mathcal{W}^{ab}(0)\mathcal{W}^{cd}(0)} & \left\{ \frac{P^{ac}(\boldsymbol{k})P^{bd}(\boldsymbol{k}')^* + P^{ac}(\boldsymbol{k}')P^{bd}(\boldsymbol{k})^*}{2} \mathcal{Q}_{\mathcal{W}}^{ac,bd}(\boldsymbol{k}-\boldsymbol{k}') \right. \\
+ \frac{P^{ac}(\boldsymbol{k}) + P^{ac}(\boldsymbol{k}')}{2}\mathcal{Q}_{\times}^{ac,bd}(\boldsymbol{k}-\boldsymbol{k}') &+ \frac{P^{bd}(\boldsymbol{k})^* + P^{bd}(\boldsymbol{k}')^*}{2}\mathcal{Q}_{\times}^{bd,ac}(\boldsymbol{k}-\boldsymbol{k}')^* + \mathcal{Q}_{\mathcal{S}}^{ac,bd}(\boldsymbol{k}-\boldsymbol{k}') \\
+ \frac{P^{ad}(\boldsymbol{k})P^{bc}(\boldsymbol{k}') + P^{ad}(\boldsymbol{k}')^*P^{bc}(\boldsymbol{k})^*}{2} & \mathcal{Q}_{\mathcal{W}}^{ad,bc}(\boldsymbol{k}+\boldsymbol{k}') \\
+ \frac{P^{ad}(\boldsymbol{k}) + P^{ad}(\boldsymbol{k}')^*}{2}\mathcal{Q}_{\times}^{ad,bc}(\boldsymbol{k}+\boldsymbol{k}') &+ \left. \frac{P^{bc}(\boldsymbol{k})^* + P^{bc}(\boldsymbol{k}')}{2}\mathcal{Q}_{\times}^{bc,ad}(\boldsymbol{k}+\boldsymbol{k}')^* + \mathcal{Q}_{\mathcal{S}}^{ad,bc}(\boldsymbol{k}+\boldsymbol{k}') \right\},
\end{aligned}
\tag{A.10}
$$

where we have further defined the following auto and cross power spectra of $\mathcal{W}$ and $\mathcal{S}$

$$
\mathcal{Q}_{\mathcal{W}}^{ab,cd}(\boldsymbol{q}) \equiv \mathcal{W}^{ab}(\boldsymbol{q})\mathcal{W}^{cd}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{W}}^{ab,cd}(\boldsymbol{s})e^{-i\boldsymbol{q}\cdot\boldsymbol{s}},
$$

$$
\mathcal{Q}_{\mathcal{S}}^{ab,cd}(\boldsymbol{q}) \equiv \mathcal{S}^{ab}(\boldsymbol{q})\mathcal{S}^{cd}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\mathcal{S}}^{ab,cd}(\boldsymbol{s})e^{-i\boldsymbol{q}\cdot\boldsymbol{s}},
$$

$$
\mathcal{Q}_{\times}^{ab,cd}(\boldsymbol{q}) \equiv \mathcal{W}^{ab}(\boldsymbol{q})\mathcal{S}^{cd}(\boldsymbol{q})^* = \int_{\boldsymbol{s}} \mathcal{Q}_{\times}^{ab,cd}(\boldsymbol{s})e^{-i\boldsymbol{q}\cdot\boldsymbol{s}},
\tag{A.11}
$$

with $\mathcal{Q}(\boldsymbol{s})$ being the auto and cross correlation functions of $\mathcal{W}$ and $\mathcal{S}$, e.g.

$$
\mathcal{Q}_{\times}^{ab,cd}(\boldsymbol{s}) = \int_{\boldsymbol{x}} \mathcal{W}^{ab}(\boldsymbol{x}+\boldsymbol{s})\mathcal{S}^{cd}(\boldsymbol{x}).
\tag{A.12}
$$

The different $\mathcal{Q}$ functions completely captures the mode mixing effect in the disconnected covariance, therefore is the key to its evaluation. Since the survey window is usually conveniently represented by a synthetic random catalog, one can measure the $\mathcal{Q}$ window as either correlation functions or power spectra of the random catalog. We show how to quantify the $\mathcal{Q}$ window and then the disconnected covariance in each specific case in the following subsections.

A.2 Angular Case

We consider the angular power spectrum and its disconnected covariance in the flat sky limit. The angular field is related to the 3D field by an integral along the line of sight

$$
\delta_W(\boldsymbol{\theta}) = \int_{\chi} \delta(\boldsymbol{x})W(\boldsymbol{x}),
\tag{A.13}
$$

where $\boldsymbol{x} \equiv (\boldsymbol{x}_\perp, x_\parallel) \equiv (\chi\boldsymbol{\theta}, \chi)$. The 3D window function is usually decomposable into angular and radial distributions $W(\boldsymbol{x}) \equiv W(\boldsymbol{\theta})W(\chi) = \bar{n}(\boldsymbol{\theta})w(\boldsymbol{\theta})\bar{n}(\chi)w(\chi)$. In Fourier space the angular field becomes

$$
\delta_W(\boldsymbol{\ell}) \equiv \int_{\boldsymbol{\theta}} \delta_W(\boldsymbol{\theta})e^{-i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} = \int_{\chi,\boldsymbol{\ell}'} \delta\Big(\frac{\boldsymbol{\ell}-\boldsymbol{\ell}'}{\chi}, \chi\Big)\frac{W(\boldsymbol{\ell}')W(\chi)}{\chi^2}
\tag{A.14}
$$

Note that here $\delta(\boldsymbol{k}_\perp, \chi)$ is only the 2D Fourier transform of a transverse slice of $\delta(\boldsymbol{x})$, so that it has the dimension of area instead of volume.

The flat-sky Limber approximation [75–77] states that in the $k_\perp |\chi - \chi'| \gg 1$ limit, the correlation between two density slices is given by

$$\langle \delta^a(\boldsymbol{k}_\perp, \chi) \delta^b(-\boldsymbol{k}'_\perp, \chi') \rangle \equiv \int_{\boldsymbol{x}_\perp, \boldsymbol{x}'_\perp} \langle \delta^a(\boldsymbol{x}) \delta^b(\boldsymbol{x}') \rangle e^{-i(\boldsymbol{k}_\perp \cdot \boldsymbol{x}_\perp - \boldsymbol{k}'_\perp \cdot \boldsymbol{x}'_\perp)}$$

$$\approx \delta^{\mathrm{D}}(\chi - \chi') \Big[(2\pi)^2 \, \delta^{\mathrm{D}}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp) P^{ab}(\boldsymbol{k}_\perp; z) + \int_{\boldsymbol{x}_\perp} \frac{(1 + \alpha \delta^{\mathrm{K}}_{ab}) \bar{n}_{ab}(\boldsymbol{x})}{\bar{n}_a(\boldsymbol{x}) \bar{n}_b(\boldsymbol{x})} e^{-i(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp) \cdot \boldsymbol{x}_\perp} \Big], \quad \text{(A.15)}$$

where $P(\boldsymbol{k}; z)$ is the 3D power spectrum at redshift $z(\chi)$.

Using the Limber approximation one can derive the angular power spectrum of a windowed field in a similar way that leads to (A.7)

$$\langle \delta_W(\boldsymbol{\ell}) \delta_W(-\boldsymbol{\ell}') \rangle$$

$$\approx \int_{\chi, \boldsymbol{\ell}''} P\Big(\frac{\boldsymbol{\ell}''}{\chi}; z \Big) \frac{W(\boldsymbol{\ell} - \boldsymbol{\ell}'') W(\boldsymbol{\ell}'' - \boldsymbol{\ell}') W(\chi)^2}{\chi^2} + (1 + \alpha) \int_{\chi, \boldsymbol{\theta}} \frac{\bar{n}(\boldsymbol{x}) w(\boldsymbol{x})^2}{\chi^2} e^{-i(\boldsymbol{\ell} - \boldsymbol{\ell}') \cdot \boldsymbol{\theta}}$$

$$= \int_{\boldsymbol{\ell}''} C(\boldsymbol{\ell}'') W(\boldsymbol{\ell} - \boldsymbol{\ell}'') W(\boldsymbol{\ell}'' - \boldsymbol{\ell}') + \mathcal{S}(\boldsymbol{\ell} - \boldsymbol{\ell}') \int_\chi \frac{\bar{n}(\chi) w(\chi)^2}{\chi^2}$$

$$\approx \frac{C(\boldsymbol{\ell}) + C(\boldsymbol{\ell}')}{2} \mathcal{W}(\boldsymbol{\ell} - \boldsymbol{\ell}') + \mathcal{S}(\boldsymbol{\ell} - \boldsymbol{\ell}') \int_\chi \frac{\bar{n}(\chi) w(\chi)^2}{\chi^2} \quad \text{(A.16)}$$

where $C(\boldsymbol{\ell})$ is the unmasked angular power spectrum in the flat-sky limit

$$C(\boldsymbol{\ell}) = \int_\chi P\Big(\frac{\boldsymbol{\ell}}{\chi}; z \Big) \frac{W(\chi)^2}{\chi^2}, \quad \text{(A.17)}$$

and similar to (A.8) the angular mode-mixing windows are

$$\mathcal{W}(\boldsymbol{\ell}) \equiv \int_{\boldsymbol{\theta}} W(\boldsymbol{\theta})^2 e^{-i\boldsymbol{\ell} \cdot \boldsymbol{\theta}},$$

$$\mathcal{S}(\boldsymbol{\ell}) \equiv (1 + \alpha) \int_{\boldsymbol{\theta}} \bar{n}(\boldsymbol{\theta}) w(\boldsymbol{\theta})^2 e^{-i\boldsymbol{\ell} \cdot \boldsymbol{\theta}}. \quad \text{(A.18)}$$

Likewise their correlation function $\mathcal{Q}_{\mathcal{W}}$, $\mathcal{Q}_{\mathcal{S}}$, and $\mathcal{Q}_\times$ can be defined analogous to (A.11). Notice that even though we have assigned the same symbols to both the 3D and angular window factors, the specific one in use should be clear from the context. From (A.16) one can show that the following $C(\boldsymbol{\ell})$ estimator

$$\hat{C}(\boldsymbol{\ell}) = \frac{|\delta_W(\boldsymbol{\ell})|^2}{\mathcal{W}_0} - C_{\mathrm{shot}} \quad \text{(A.19)}$$

is unbiased at high ℓ and the shot noise is

$$C_{\mathrm{shot}} = \int_\chi \frac{\bar{n}(\chi) w(\chi)^2}{\chi^2} \frac{\mathcal{S}_0}{\mathcal{W}_0} = (1 + \alpha) \int_\chi \frac{\bar{n}(\chi) w(\chi)^2}{\chi^2} \frac{\int_{\boldsymbol{\theta}} \bar{n}(\boldsymbol{\theta}) w(\boldsymbol{\theta})^2}{\int_{\boldsymbol{\theta}} \bar{n}(\boldsymbol{\theta})^2 w(\boldsymbol{\theta})^2}. \quad \text{(A.20)}$$

Because of the similarity of structure here to that of the 3D case, the disconnected covariance of the angle-averaged angular power spectrum resembles that of the power spectrum

multipoles. So analogous to (2.24) we can write down immediately

$$
\begin{aligned}
\mathrm{Cov}^{\mathrm{disc}}\big[\hat{C}(\ell), \hat{C}(\ell')\big] &= \int_{\hat{\boldsymbol{\ell}}, \hat{\boldsymbol{\ell}}'} \mathrm{Cov}^{\mathrm{disc}}\big[\hat{C}(\boldsymbol{\ell}), \hat{C}(\boldsymbol{\ell}')\big] \approx 2 \int_{\hat{\boldsymbol{\ell}}, \hat{\boldsymbol{\ell}}'} \bigg\{ C(\ell)C(\ell') \frac{|\mathcal{W}(\boldsymbol{\ell} - \boldsymbol{\ell}')|^2}{\mathcal{W}_0^2} \\
&\quad + \big[C(\ell) + C(\ell')\big] C_{\mathrm{shot}} \frac{\Re\big[\mathcal{W}(\boldsymbol{\ell} - \boldsymbol{\ell}')\mathcal{S}(\boldsymbol{\ell} - \boldsymbol{\ell}')^*\big]}{\mathcal{W}_0 \mathcal{S}_0} + C_{\mathrm{shot}}^2 \frac{|\mathcal{S}(\boldsymbol{\ell} - \boldsymbol{\ell}')|^2}{\mathcal{S}_0^2} \bigg\} \\
&= \frac{2}{\mathcal{W}_0^2} \bigg\{ C(\ell)C(\ell') \int 2\pi\theta\mathrm{d}\theta\, \mathcal{Q}_{\mathcal{W}}(\theta) J_0(\ell\theta) J_0(\ell'\theta) \\
&\quad + \big[C(\ell) + C(\ell')\big] \int 2\pi\theta\mathrm{d}\theta\, \mathcal{Q}_\times(\theta) J_0(\ell\theta) J_0(\ell'\theta) + \int 2\pi\theta\mathrm{d}\theta\, \mathcal{Q}_{\mathcal{S}}(\theta) J_0(\ell\theta) J_0(\ell'\theta) \bigg\}, \quad \text{(A.21)}
\end{aligned}
$$

in deriving which we have used the Jacobi-Anger expansion (B.4).

As in the multipole case, given the $\mathcal{Q}$ measured from the survey random catalog and the $C(\ell)$ constrained from the data, we can compute the disconnected covariance of $\hat{C}(\ell)$ with this equation. Its evaluation again includes three integrals involving two zeroth order Bessel functions J_0. We show their solution in Sec. 3.

A.3 Projected and Projected×Multipole Cases

When cross correlate a field with full 3-dimensional information, e.g. 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 $\boldsymbol{x}_\perp$ using the radial information of the 3D field. Compared to the angular projection in A.2, in such case each field are effectively projected as

$$
\delta_W(\boldsymbol{x}_\perp) = \int_\chi \delta(\boldsymbol{x}) W(\boldsymbol{x}), \tag{A.22}
$$

which in Fourier space reads

$$
\delta_W(\boldsymbol{k}_\perp) = \int_{\boldsymbol{q}} \delta(\boldsymbol{k}_\perp - \boldsymbol{q}) W(\boldsymbol{q}) = \int_{\chi, \boldsymbol{\ell}} \delta\Big(\boldsymbol{k}_\perp - \frac{\boldsymbol{\ell}}{\chi}, \chi\Big) W(\boldsymbol{\ell}) W(\chi). \tag{A.23}
$$

The projected field is indeed the $k_\parallel = 0$ component of the corresponding 3D field, therefore having the same equations for 2-point estimator, shot noise, covariance, etc. Therefore according to (A.7) the cross correlation between a projected and a 3D field is

$$
\langle \delta_W^a(\boldsymbol{k}_\perp) \delta_W^b(-\boldsymbol{k}') \rangle \approx \frac{P^{ab}(\boldsymbol{k}_\perp) + P^{ab}(\boldsymbol{k}')}{2} \mathcal{W}^{ab}(\boldsymbol{k}_\perp - \boldsymbol{k}') + \mathcal{S}^{ab}(\boldsymbol{k}_\perp - \boldsymbol{k}'), \tag{A.24}
$$

with $\mathcal{W}$ and $\mathcal{S}$ defined as in (A.8). The auto correlation of two projected fields then reduces to a special case of the above equation when $k_\parallel' = 0$.

As before, the above cross correlation forms the basis to compute the disconnected cross covariance between a projected power spectrum and multipoles

$$
\begin{aligned}
&\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}_\perp^a(k), \hat{P}_{\ell'}^b(k')\big] \\
&= \frac{2(2\ell' + 1)}{\mathcal{W}^a(0)\mathcal{W}^b(0)} \int_{\hat{\boldsymbol{k}}, \hat{\boldsymbol{k}}'} 2\,\delta^{\mathrm{D}}(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}})\, \mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}) \big|\langle \delta_W^a(\boldsymbol{k}) \delta_W^b(-\boldsymbol{k}') \rangle\big|^2 \\
&\approx \sum_{\ell=0}^{\ell_{\max}} \frac{2(2\ell + 1)(2\ell' + 1)\,\mathcal{L}_\ell(0)}{\mathcal{W}^a(0)\mathcal{W}^b(0)} \int_{\hat{\boldsymbol{k}}, \hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}})\, \mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}' \cdot \hat{\boldsymbol{n}}) \big|\langle \delta_W^a(\boldsymbol{k}) \delta_W^b(-\boldsymbol{k}') \rangle\big|^2. \tag{A.25}
\end{aligned}
$$

Since both power spectrum and the window are smooth functions in angular direction, we can expand the Dirac delta above in multipoles using (B.2) truncated at some $\ell_{\max}$. This reduce the above expression to the multipole case that we have already solved, as shown in (2.24), (2.25), (2.26), and (2.27).

One may want to compute the disconnected auto covariance of a projected power spectrum in a similar way by expanding both Dirac deltas as follows

$$
\begin{aligned}
&\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}_\perp^a(k), \hat{P}_\perp^b(k')\big] \\
&\approx \sum_{\ell\ell'} \frac{2(2\ell+1)(2\ell'+1)\,\mathcal{L}_\ell(0)\,\mathcal{L}_{\ell'}(0)}{\mathcal{W}^a(0)\mathcal{W}^b(0)} \int_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'} \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})\,\mathcal{L}_{\ell'}(\hat{\boldsymbol{k}}'\cdot\hat{\boldsymbol{n}})\big|\langle\delta_W^a(\boldsymbol{k})\delta_W^b(-\boldsymbol{k}')\rangle\big|^2.
\end{aligned}
$$

However, this turns out to converge much more slowly with $\ell_{\max}$ than the equation before. Instead, for the projected auto covariance we limit the computation only in the transverse direction as in (A.21)

$$
\begin{aligned}
\mathrm{Cov}^{\mathrm{disc}}\big[\hat{P}_\perp^a(k), \hat{P}_\perp^b(k')\big] &= \frac{2}{\mathcal{W}^a(0)\mathcal{W}^b(0)} \int_{\hat{\boldsymbol{k}}_\perp,\hat{\boldsymbol{k}}'_\perp} \big|\langle\delta_W^a(\boldsymbol{k}_\perp)\delta_W^b(-\boldsymbol{k}'_\perp)\rangle\big|^2 \\
&\approx \frac{2}{\mathcal{W}^a(0)\mathcal{W}^b(0)} \Bigg\{ P_\perp^{ab}(k)P_\perp^{ab}(k') \int 2\pi s\mathrm{d}s\, \mathcal{Q}_{\mathcal{W}\perp}^{ab,ab}(s)J_0(ks)J_0(k's) \\
&\qquad + \big[P_\perp^{ab}(k) + P_\perp^{ab}(k')\big] \int 2\pi s\mathrm{d}s\, \mathcal{Q}_{\times\perp}^{ab,ab}(s)J_0(ks)J_0(k's) \\
&\qquad + \int 2\pi s\mathrm{d}s\, \mathcal{Q}_{\mathcal{S}\perp}^{ab,ab}(s)J_0(ks)J_0(k's) \Bigg\}.
\end{aligned}
\tag{A.26}
$$

Here $\boldsymbol{k}\equiv\boldsymbol{k}_\perp$ on the first line. The projected power spectrum $P_\perp$ can be expressed as a linear combination of multipoles $P_\perp(k) = \sum_\ell P_\ell(k)\,\mathcal{L}_\ell(0) \approx P_0(k) - P_2(k)/2 + 3P_4(k)/8$. And the projected $\mathcal{Q}$ factors in the above equation have their radial separation integrated and their direction of transverse separation averaged

$$
\mathcal{Q}_\perp^{ab,cd}(s_\perp) = \int_{s_\parallel,\hat{\boldsymbol{s}}_\perp} \mathcal{Q}^{ab,cd}(\boldsymbol{s}).
\tag{A.27}
$$

A.4 Correlation Function Case

Sec. 2.4.4 shows that the Fourier transform relates the 3D correlation function and power spectrum, and thus their covariances. Here we derive similar relations for the more practical bases, including multipole, angular and projected cases. For multipoles, the Fourier transform connecting two representations becomes the Hankel transform. Apart from that, additional polar angle dependence in $\mathcal{Q}_W$ couples to that in the power spectrum to give

$$
\begin{aligned}
\xi_\ell(s) &= i^\ell \int \frac{k^2\mathrm{d}k}{2\pi^2} P_\ell(k)j_\ell(ks), \\
\hat{\xi}_\ell(s) &= (2\ell+1)\mathcal{W}_0 \sum_{\ell_1\ell_2} \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 \mathcal{Q}_{W\ell_2}^{-1}(s)\Big[i^{\ell_1} \int \frac{k^2\mathrm{d}k}{2\pi^2}\hat{P}_{\ell_1}(k)j_{\ell_1}(ks)\Big],
\end{aligned}
\tag{A.28}
$$

where $\mathcal{Q}_{W\ell}^{-1}$ is the multipole moments of $1/\mathcal{Q}_W(\boldsymbol{s})$, i.e.

$$
\mathcal{Q}_{W\ell}^{-1}(s) \equiv (2\ell+1) \int_{\hat{\boldsymbol{s}}} \frac{1}{\mathcal{Q}_W(\boldsymbol{s})}\,\mathcal{L}_\ell(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}).
\tag{A.29}
$$

Assuming smooth angular variation, $\mathcal{Q}_{W\ell}^{-1}(s)$ can be computed given $\mathcal{Q}_{W\ell}(s)$ and we show this procedure in App. B.1.

Therefore the disconnected covariance of the correlation function multipoles is

$$\text{Cov}^{\text{disc}}\big[\hat{\xi}_\ell(s), \hat{\xi}_{\ell'}(s')\big] = (2\ell+1)(2\ell'+1)\mathcal{W}_0^2 \sum_{\ell_1\ell_2\ell_3\ell_4} \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} \ell' & \ell_3 & \ell_4 \\ 0 & 0 & 0 \end{pmatrix}^2$$

$$\mathcal{Q}_{W\ell_2}^{-1}(s)\mathcal{Q}_{W\ell_4}^{-1}(s)\, i^{\ell_1-\ell_3} \int \frac{k^2\mathrm{d}k}{2\pi^2}\frac{k'^2\mathrm{d}k'}{2\pi^2} \text{Cov}^{\text{disc}}\big[\hat{P}_{\ell_1}(k), \hat{P}_{\ell_3}(k')\big] j_{\ell_1}(ks)j_{\ell_3}(k's'). \quad \text{(A.30)}$$

Once we have computed $\text{Cov}^{\text{disc}}\big[\hat{P}_\ell(k), \hat{P}_{\ell'}(k')\big]$ following Sec. 2.2, we can evaluate the above equation by calling `mcfit` to Hankel transform k into s, and k' into s', respectively.

The flat-sky angular correlation function is related to the corresponding power spectrum (App. A.2) by the J_0 Hankel transform

$$w(\theta) = \int \frac{\ell\mathrm{d}\ell}{2\pi}\, C(\ell) J_0(\ell\theta),$$

$$\hat{w}(\theta) = \frac{\mathcal{W}_0}{\mathcal{Q}_W(\theta)} \int \frac{\ell\mathrm{d}\ell}{2\pi}\, \hat{C}(\ell) J_0(\ell\theta). \quad \text{(A.31)}$$

Thus the disconnected covariance of the angular correlation function is

$$\text{Cov}^{\text{disc}}\big[\hat{w}(\theta), \hat{w}(\theta')\big] = \frac{\mathcal{W}_0^2}{\mathcal{Q}_W(\theta)\mathcal{Q}_W(\theta')} \int \frac{\ell\mathrm{d}\ell}{2\pi}\frac{\ell'\mathrm{d}\ell'}{2\pi} \text{Cov}^{\text{disc}}\big[\hat{C}(\ell), \hat{C}(\ell')\big] J_0(\ell\theta) J_0(\ell'\theta'). \quad \text{(A.32)}$$

Again this is straightforward to compute with `mcfit`, given $\text{Cov}^{\text{disc}}\big[\hat{C}(\ell), \hat{C}(\ell')\big]$ from Sec. A.2.

The same relations in the above two equations also connect the projected correlation function and power spectrum (App. A.3), if we replace $w \to \xi_\perp$, $C \to P_\perp$, $\theta \to s$, and $\ell \to k$. Here $\xi_\perp$ is the dimensionless projected correlation functions. Combining the results on multipole and projected cases, we can write down the disconnected covariance of projected×multipole correlation functions

$$\text{Cov}^{\text{disc}}\big[\hat{\xi}_\perp(s), \hat{\xi}_{\ell'}(s')\big] = \frac{(2\ell'+1)\mathcal{W}_0^2}{\mathcal{Q}_{W\perp}(s)} \sum_{\ell_3\ell_4} \begin{pmatrix} \ell' & \ell_3 & \ell_4 \\ 0 & 0 & 0 \end{pmatrix}^2$$

$$\mathcal{Q}_{W\ell_4}^{-1}(s')\, i^{-\ell_3} \int \frac{k\mathrm{d}k}{2\pi}\frac{k'^2\mathrm{d}k'}{2\pi^2} \text{Cov}^{\text{disc}}\big[\hat{P}_\perp(k), \hat{P}_{\ell_3}(k')\big] J_0(ks)j_{\ell_3}(k's'). \quad \text{(A.33)}$$

We verify this equation with galaxy mocks in Sec. 5.2.

B   Useful Relations

The product of two Legendre polynomials can be expanded in itself as

$$\mathcal{L}_{\ell_1}(\mu)\,\mathcal{L}_{\ell_2}(\mu) = \sum_{\ell_3=|\ell_1-\ell_2|}^{\ell_1+\ell_2} (2\ell_3+1) \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \mathcal{L}_{\ell_3}(\mu), \quad \text{(B.1)}$$

where the 2×3 matrix enclosed by parentheses denotes the Wigner $3j$ symbol. The completeness relation of the Legendre polynomials is

$$\delta^{\text{D}}(\mu-\mu') = \sum_{\ell=0}^{\infty} \frac{2l+1}{2} \mathcal{L}_\ell(\mu)\,\mathcal{L}_\ell(\mu'). \quad \text{(B.2)}$$

The plane wave expansion equates the kernel of Fourier transform in 3D with a series in spherical Bessel function j_ℓ and Legendre polynomial

$$e^{i\boldsymbol{k}\cdot\boldsymbol{s}} = \sum_{\ell=0}^{\infty}(2\ell+1)i^\ell j_\ell(ks)\,\mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{s}}), \tag{B.3}$$

and in 2D with a series in Bessel function J_n

$$e^{i\boldsymbol{\ell}\cdot\boldsymbol{\theta}} = \sum_{n=-\infty}^{\infty} i^{|n|}J_{|n|}(\ell\theta)e^{in\phi} = J_0(\ell\theta) + 2\sum_{n=1}^{\infty} i^n J_n(\ell\theta)\cos(n\phi), \tag{B.4}$$

where $\cos\phi \equiv \hat{\boldsymbol{\ell}}\cdot\hat{\boldsymbol{\theta}}$.

The addition theorem for spherical harmonics relates the Legendre polynomial with sum of products of spherical harmonics

$$\mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{s}}) = \frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell} Y_\ell^m(\hat{\boldsymbol{k}})Y_\ell^m(\hat{\boldsymbol{s}})^*, \tag{B.5}$$

with the spherical harmonics satisfying the following orthonormal relation

$$\int_{\hat{\boldsymbol{v}}} Y_\ell^m(\hat{\boldsymbol{v}})Y_{\ell'}^{m'}(\hat{\boldsymbol{v}})^* = \frac{\delta^{\mathrm{K}}_{\ell\ell'}\delta^{\mathrm{K}}_{mm'}}{4\pi}, \tag{B.6}$$

where an extra 4π factor appears due to our shorthand notation $\int_{\hat{\boldsymbol{v}}}$ (footnote 3) that averages rather than integrates over the solid angle.

One can prove the following relation by using (B.3), (B.5), and (B.6)

$$\int_{\hat{\boldsymbol{k}}} \mathcal{L}_\ell(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{n}})e^{-i\boldsymbol{k}\cdot\boldsymbol{s}} = (-i)^\ell j_\ell(ks)\,\mathcal{L}_\ell(\hat{\boldsymbol{s}}\cdot\hat{\boldsymbol{n}}). \tag{B.7}$$

B.1 Inversion of Multipoles

The redshift-space correlation function has a μ-dependent denominator. To account for this in the covariance of $\xi_\ell(s)$, in one approach we can compute the multipole moments of the inverse of the denominator, whose multipoles are already known. Let a_ℓ be the multipole moments of $a(\mu)$ and we want to compute b_ℓ of $b(\mu)$ that satisfy

$$1 \equiv a(\mu)b(\mu) = \sum_{\ell\ell'}a_\ell b_{\ell'}\,\mathcal{L}_\ell(\mu)\mathcal{L}_{\ell'}(\mu) = \sum_{\ell\ell'\ell''}a_\ell b_{\ell'}(2\ell''+1)\begin{pmatrix}\ell & \ell' & \ell'' \\ 0 & 0 & 0\end{pmatrix}^2 \mathcal{L}_{\ell''}(\mu). \tag{B.8}$$

The multipole coupling on the right hand side should result in only the monopole on the left hand side, leading to the following linear system

$$\sum_{\ell\ell'}a_\ell b_{\ell'}\begin{pmatrix}\ell & \ell' & \ell'' \\ 0 & 0 & 0\end{pmatrix}^2 = \delta^{\mathrm{K}}_{\ell''0}, \tag{B.9}$$

from which we can solve for b_ℓ.

References

[1] C. Blake, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom et al., *The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9*, MNRAS **415** (Aug., 2011) 2876–2891, [1104.2948].

[2] F. Beutler, H.-J. Seo, A. J. Ross, P. McDonald, S. Saito, A. S. Bolton et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space*, MNRAS **464** (Jan., 2017) 3409–3430, [1607.03149].

[3] B. Reid, S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, R. Tojeiro et al., *SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues*, MNRAS **455** (Jan., 2016) 1553–1573, [1509.06529].

[4] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell et al., *The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant*, MNRAS **416** (Oct., 2011) 3017–3032, [1106.3366].

[5] L. Anderson, É. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton 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*, MNRAS **441** (June, 2014) 24–62, [1312.4877].

[6] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, MNRAS **470** (Sept., 2017) 2617–2652, [1607.03155].

[7] M. A. Troxel, N. MacCrann, J. Zuntz, T. F. Eifler, E. Krause, S. Dodelson et al., *Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear*, Phys. Rev. D **98** (Aug., 2018) 043528, [1708.01538].

[8] S. Joudaki, C. Blake, A. Johnson, A. Amon, M. Asgari, A. Choi et al., *KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering*, MNRAS **474** (Mar., 2018) 4894–4924, [1707.06627].

[9] S. Singh, S. Alam, R. Mandelbaum, U. Seljak, S. Rodriguez-Torres and S. Ho, *Probing gravity with a joint analysis of galaxy and CMB lensing and SDSS spectroscopy*, MNRAS (Oct., 2018) , [1803.08915].

[10] C. Hikage, M. Oguri, T. Hamana, S. More, R. Mandelbaum, M. Takada et al., *Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data*, ArXiv e-prints (Sept., 2018) , [1809.09148].

[11] A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko et al., *A 2.4% Determination of the Local Value of the Hubble Constant*, ApJ **826** (July, 2016) 56, [1604.01424].

[12] V. Bonvin, F. Courbin, S. H. Suyu, P. J. Marshall, C. E. Rusu, D. Sluse 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*, MNRAS **465** (Mar., 2017) 4914–4930, [1607.01790].

[13] LSST Dark Energy Science Collaboration, *Large Synoptic Survey Telescope: Dark Energy Science Collaboration*, ArXiv e-prints (Nov., 2012) , [1211.0310].

[14] L. Amendola, S. Appleby, D. Bacon, T. Baker, M. Baldi, N. Bartolo et al., *Cosmology and Fundamental Physics with the Euclid Satellite*, *Living Reviews in Relativity* **16** (Sept., 2013) 6, [1206.1225].

[15] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen et al., *The DESI Experiment Part I: Science,Targeting, and Survey Design*, ArXiv e-prints (Oct., 2016) , [1611.00036].

[16] O. Doré, C. Hirata, Y. Wang, D. Weinberg, I. Baronchelli, A. Benson et al., *WFIRST Science Investigation Team "Cosmology with the High Latitude Survey" Annual Report 2017*, ArXiv e-prints (Apr., 2018) , [1804.03628].

[17] J. Hartlap, P. Simon and P. Schneider, *Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix*, A&A **464** (Mar., 2007) 399–404, [arXiv:astro-ph/0608064].

[18] S. Dodelson and M. D. Schneider, *The effect of covariance estimator error on cosmological parameter constraints*, Phys. Rev. D **88** (Sept., 2013) 063537, [1304.2593].

[19] E. Sellentin and A. F. Heavens, *Parameter inference with estimated covariance matrices*, MNRAS **456** (Feb., 2016) L132–L136, [1511.05969].

[20] A. Taylor, B. Joachimi and T. Kitching, *Putting the precision in precision cosmology: How accurate should your data covariance matrix be?*, MNRAS **432** (July, 2013) 1928–1946, [1212.4359].

[21] M. Lippich, A. G. Sánchez, M. Colavincenzo, E. Sefusatti, P. Monaco, L. Blot et al., *Comparing approximate methods for mock catalogues and covariance matrices I: correlation function*, ArXiv e-prints (June, 2018) , [1806.09477].

[22] L. Blot, M. Crocce, E. Sefusatti, M. Lippich, A. G. Sánchez, M. Colavincenzo et al., *Comparing approximate methods for mock catalogues and covariance matrices II: Power spectrum multipoles*, ArXiv e-prints (June, 2018) , [1806.09497].

[23] F. Baumgarten and C.-H. Chuang, *Robustness of the covariance matrix for galaxy clustering measurements*, MNRAS **480** (Oct., 2018) 2535–2543, [1802.04462].

[24] P. Norberg, C. M. Baugh, E. Gaztañaga and D. J. Croton, *Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics*, MNRAS **396** (June, 2009) 19–38, [0810.1885].

[25] H. A. Feldman, N. Kaiser and J. A. Peacock, *Power-spectrum analysis of three-dimensional redshift surveys*, ApJ **426** (May, 1994) 23–37, [arXiv:astro-ph/9304022].

[26] A. Meiksin and M. White, *The growth of correlations in the matter power spectrum*, MNRAS **308** (Oct., 1999) 1179–1184, [astro-ph/9812129].

[27] R. Scoccimarro, M. Zaldarriaga and L. Hui, *Power Spectrum Correlations Induced by Nonlinear Clustering*, ApJ **527** (Dec., 1999) 1–15, [astro-ph/9901099].

[28] A. Cooray and W. Hu, *Power Spectrum Covariance of Weak Gravitational Lensing*, ApJ **554** (June, 2001) 56–66, [astro-ph/0012087].

[29] J. Harnois-Déraps and U.-L. Pen, *Non-Gaussian error bars in galaxy surveys - I*, MNRAS **423** (July, 2012) 2288–2307, [1109.5746].

[30] D. Bertolini, K. Schutz, M. P. Solon, J. R. Walsh and K. M. Zurek, *Non-Gaussian covariance of the matter power spectrum in the effective field theory of large scale structure*, Phys. Rev. D **93** (June, 2016) 123505, [1512.07630].

[31] I. Mohammed, U. Seljak and Z. Vlah, *Perturbative approach to covariance matrix of the matter power spectrum*, MNRAS **466** (Apr., 2017) 780–797, [1607.00043].

[32] A. Barreira and F. Schmidt, *Response approach to the matter power spectrum covariance*, J. Cosmology Astropart. Phys. **11** (Nov., 2017) 051, [1705.01092].

[33] F. Lacasa, *Covariance of the galaxy angular power spectrum with the halo model*, A&A **615** (July, 2018) A1, [1711.07372].

[34] A. J. S. Hamilton, C. D. Rimes and R. Scoccimarro, *On measuring the covariance matrix of the non-linear power spectrum from simulations*, MNRAS **371** (Sept., 2006) 1188–1204, [astro-ph/0511416].

[35] W. Hu and A. V. Kravtsov, *Sample Variance Considerations for Cluster Surveys*, ApJ **584** (Feb., 2003) 702–715, [`astro-ph/0203169`].

[36] M. Takada and B. Jain, *The impact of non-Gaussian errors on weak lensing surveys*, MNRAS **395** (June, 2009) 2065–2086, [`0810.4170`].

[37] M. Takada and W. Hu, *Power spectrum super-sample covariance*, Phys. Rev. D **87** (June, 2013) 123504, [`1302.6994`].

[38] Y. Li, W. Hu and M. Takada, *Super-sample covariance in simulations*, Phys. Rev. D **89** (Apr., 2014) 083519, [`1401.0385`].

[39] Y. Li, W. Hu and M. Takada, *Super-sample signal*, Phys. Rev. D **90** (Nov., 2014) 103530, [`1408.1081`].

[40] K. Akitsu, M. Takada and Y. Li, *Large-scale tidal effect on redshift-space power spectrum in a finite-volume survey*, Phys. Rev. D **95** (Apr., 2017) 083522, [`1611.04723`].

[41] Y. Li, M. Schmittfull and U. Seljak, *Galaxy power-spectrum responses and redshift-space super-sample effect*, J. Cosmology Astropart. Phys. **2** (Feb., 2018) 022, [`1711.00018`].

[42] A. Barreira, E. Krause and F. Schmidt, *Complete super-sample lensing covariance in the response approach*, J. Cosmology Astropart. Phys. **6** (June, 2018) 015, [`1711.07467`].

[43] J. D. Cohn, *Power spectrum and correlation function errors: Poisson vs. Gaussian shot noise*, New A **11** (Jan., 2006) 226–239, [`astro-ph/0503285`].

[44] R. E. Smith, *Covariance of cross-correlations: towards efficient measures for large-scale structure*, MNRAS **400** (Dec., 2009) 851–865, [`0810.1960`].

[45] A. C. Pope and I. Szapudi, *Shrinkage estimation of the power spectrum covariance matrix*, MNRAS **389** (Sept., 2008) 766–774, [`0711.2509`].

[46] R. O'Connell, D. Eisenstein, M. Vargas, S. Ho and N. Padmanabhan, *Large covariance matrices: smooth models from the two-point correlation function*, MNRAS **462** (Nov., 2016) 2681–2694, [`1510.01740`].

[47] D. W. Pearson and L. Samushia, *Estimating the power spectrum covariance matrix with fewer mock samples*, MNRAS **457** (Mar., 2016) 993–999, [`1509.00064`].

[48] A. Hall and A. Taylor, *A Bayesian method for combining theoretical and simulated covariance matrices for large-scale structure surveys*, ArXiv e-prints (July, 2018) , [`1807.06875`].

[49] R. O'Connell and D. J. Eisenstein, *Large Covariance Matrices: Accurate Models Without Mocks*, ArXiv e-prints (Aug., 2018) , [`1808.05978`].

[50] M. Tegmark, *Measuring Cosmological Parameters with Galaxy Surveys*, Physical Review Letters **79** (Nov., 1997) 3806–3809, [`astro-ph/9706198`].

[51] K. Yamamoto, M. Nakamichi, A. Kamino, B. A. Bassett and H. Nishioka, *A Measurement of the Quadrupole Power Spectrum in the Clustering of the 2dF QSO Survey*, PASJ **58** (Feb., 2006) 93–102, [`astro-ph/0505115`].

[52] N. Hand, Y. Li, Z. Slepian and U. Seljak, *An optimal FFT-based anisotropic power spectrum estimator*, J. Cosmology Astropart. Phys. **7** (July, 2017) 002, [`1704.02357`].

[53] N. S. Sugiyama, M. Shiraishi and T. Okumura, *Limits on statistical anisotropy from BOSS DR12 galaxies using bipolar spherical harmonics*, MNRAS **473** (Jan., 2018) 2737–2752, [`1704.02868`].

[54] D. Bianchi, H. Gil-Marín, R. Ruggeri and W. J. Percival, *Measuring line-of-sight-dependent Fourier-space clustering using FFTs*, MNRAS **453** (Oct., 2015) L11–L15, [`1505.05341`].

[55] R. Scoccimarro, *Fast estimators for redshift-space clustering*, Phys. Rev. D **92** (Oct., 2015) 083532, [`1506.02729`].

[56] Z. Slepian and D. J. Eisenstein, *Accelerating the two-point and three-point galaxy correlation functions using Fourier transforms*, MNRAS **455** (Jan., 2016) L31–L35, [1506.04746].

[57] M. Sinha and L. Garrison, "Corrfunc: Blazing fast correlation functions on the CPU." Astrophysics Source Code Library, Mar., 2017.

[58] N. Hand, U. Seljak, F. Beutler and Z. Vlah, *Extending the modeling of the anisotropic galaxy power spectrum to $k = 0.4\ hMpc^{-1}$*, J. Cosmology Astropart. Phys. **10** (Oct., 2017) 009, [1706.02362].

[59] M. J. Wilson, J. A. Peacock, A. N. Taylor and S. de la Torre, *Rapid modelling of the redshift-space power spectrum multipoles for a masked density field*, MNRAS **464** (Jan., 2017) 3121–3130, [1511.07799].

[60] Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters*, A&A **594** (Sept., 2016) A11, [1507.02704].

[61] S. Singh, R. Mandelbaum, U. Seljak, A. Slosar and J. Vazquez Gonzalez, *Galaxy-galaxy lensing estimators and their covariance properties*, MNRAS **471** (Nov., 2017) 3827–3844, [1611.00752].

[62] F. Simpson, C. Heymans, D. Parkinson, C. Blake, M. Kilbinger, J. Benjamin et al., *CFHTLenS: testing the laws of gravity with tomographic weak lensing and redshift-space distortions*, MNRAS **429** (Mar., 2013) 2249–2263, [1212.3339].

[63] S. D. Landy and A. S. Szalay, *Bias and variance of angular correlation functions*, ApJ **412** (July, 1993) 64–71.

[64] V. Assassi, M. Simonović and M. Zaldarriaga, *Efficient evaluation of angular power spectra and bispectra*, J. Cosmology Astropart. Phys. **11** (Nov., 2017) 054, [1705.05022].

[65] H. S. Grasshorn Gebhardt and D. Jeong, *Fast and accurate computation of projected two-point functions*, Phys. Rev. D **97** (Jan., 2018) 023504, [1709.02401].

[66] J. D. Talman, *Numerical Fourier and Bessel Transforms in Logarithmic Variables*, Journal of Computational Physics **29** (Oct., 1978) 35–48.

[67] A. J. S. Hamilton, *Uncorrelated modes of the non-linear power spectrum*, MNRAS **312** (Feb., 2000) 257–284, [astro-ph/9905191].

[68] Y. Li, "Multiplicatively convolutional fast integral transforms." https://github.com/eelregit/mcfit.

[69] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release*, MNRAS **456** (Mar., 2016) 4156–4173, [1509.06400].

[70] Z. Slepian and D. J. Eisenstein, *Computing the three-point correlation function of galaxies in O(N^2) time*, MNRAS **454** (Dec., 2015) 4142–4158, [1506.02040].

[71] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak et al., *nbodykit: an open-source, massively parallel toolkit for large-scale structure*, The Astronomical Journal **156** (2018) 160.

[72] J. N. Grieb, A. G. Sánchez, S. Salazar-Albornoz and C. Dalla Vecchia, *Gaussian covariance matrices for anisotropic galaxy clustering measurements*, MNRAS **457** (Apr., 2016) 1577–1592, [1509.04293].

[73] C. Howlett and W. J. Percival, *Galaxy two-point covariance matrix estimation for next generation surveys*, MNRAS **472** (Dec., 2017) 4935–4952, [1709.03057].

[74] C. Blake, P. Carter and J. Koda, *Power spectrum multipoles on the curved sky: an application to the 6-degree Field Galaxy Survey*, MNRAS **479** (Oct., 2018) 5168–5183, [1801.04969].

[75] D. N. Limber, *The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.*, ApJ **117** (Jan., 1953) 134.

[76] N. Kaiser, *Weak gravitational lensing of distant galaxies*, ApJ **388** (Apr., 1992) 272–286.

[77] M. Loverde and N. Afshordi, *Extended Limber approximation*, Phys. Rev. D **78** (Dec., 2008) 123506, [0809.5112].