**Improved Master for the LSS: Fast and accurate analysis of the two point power spectra and correlation functions**

Sukhdeep Singh\textsuperscript{1,2,*}

\textsuperscript{1}McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
\textsuperscript{2}Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA

Accepted XXX. Received YYY; in original form ZZZ

**Abstract**

We review the methodology for measurements of two point functions of the cosmological observables, both power spectra and correlation functions. For pseudo-$C_\ell$ estimators, we will argue that the window weighted overdensity field can yield more optimal measurements as the window acts as an inverse noise weight, an effect that becomes more important for surveys with a variable selection function. We then discuss the impact of approximations made in the Master algorithm and suggest improvements, the iMaster algorithm, that uses the theoretical model to give unbiased results for arbitrarily complex windows provided that the model satisfies weak accuracy conditions. The methodology of iMaster algorithm is also generalized to the correlation functions to reconstruct the binned power spectra, for E/B mode separation, or to properly convolve the correlation functions to account for the scale cuts in the Fourier space model. We also show that the errors in the window estimation lead to both additive and multiplicative effects on the over density field. Accurate estimation of window power can be required up to scales of $\approx 2\ell_{\text{max}}$ or larger. Misestimation of the window power leads to biases in the measured power spectra which scale as $\delta C_\ell \approx M_W^{\ell \ell} \delta W_\ell$, where the $M_W^{\ell \ell}$ scales as $\sim (\ell + 1) C_\ell$ leading to effects that can be important at high $\ell$. While the notation in this paper is geared towards photometric galaxy surveys, the discussion is equally applicable to spectroscopic galaxy, intensity mapping and CMB surveys.

**Key words: cosmology: observations — large-scale structure of Universe — gravitational lensing: weak**

1 INTRODUCTION

The measurements of the large scale structure (LSS) in the universe provide important cosmological information about the evolution of the universe over time and also allow us to study the physical properties of its constituents, namely Dark matter, Dark energy, neutrinos and baryons (see Weinberg et al. 2013, for a review). Over the past two decades we have successfully measured the LSS using a number of probes, e.g. baryon acoustic oscillations and galaxy velocities (e.g. Alam et al. 2016; Neveux et al. 2020), weak gravitational lensing (e.g. Planck Collaboration et al. 2020; DES Collaboration et al. 2017; Singh et al. 2020; Heymans et al. 2021), galaxy clusters (e.g. Abbott et al. 2020). These measurements have yielded strong constraints on the cosmological models with precision of order 5-10%. There are also some tensions among the probes at $2-3\sigma$ significance level (DES Collaboration et al. 2017; Singh et al. 2020; Heymans et al. 2021; Di Valentino et al. 2020; Lange et al. 2021) that have been a source of general intrigue and excitement in the community and also require us to revisit many of the assumptions made in the analysis. With upcoming percent level measurements from the planned surveys, e.g. Rubin Observatory LSST-DESC (LSST Dark Energy Science Collaboration 2012), DESI (Levi et al. 2019), Roman space telescope (Dore et al. 2019), Simons Observatory (Ade et al. 2019), CMB-S4 (Abazajian et al. 2016), Spherex (Doré et al. 2014), it is going to become even more important for us to thoroughly understand the analysis in order to derive accurate inferences on cosmological models. In this paper we will review the methodology for the measurements of the two point statistics of the over density field from the LSS surveys.

The large scale structure measurements from the cosmological surveys are made by turning the data from the survey into the maps of over density field, which is the quan-
tity of interest in measuring the fluctuations in the matter density. Correlations of mean zero overdensity field have also been shown to be more optimal than the density field with non-zero mean (Landy & Szalay 1993; Singh et al. 2017). In this paper, by optimal we will usually mean minimizing some combination of bias and variance in the inferences. The maps of over density are then compressed using the summary statistics, especially the two point correlation functions or power spectra, which are a measure of the variance in the field. During the process of these measurements there are several questions that one needs to address to ensure the optimality of the measurement as well as the accuracy of the inferences derived from these measurements.

After the data is acquired and cleaned, the analysis begins with the process of map making (or making catalogs), whereby we bin the count of galaxies (or photons in case of intensity mapping) into pixels of a map. One of the questions that we need to address at this stage is the weighting to be applied to the pixels to ensure optimal analysis. The question of optimal weights for the case of gaussian random fields has been addressed by Feldman et al. (1994) giving us the famous FKP weights, which was later generalized by Hamilton (1997). While FKP weights (or quadratic estimator in general) has been shown to be optimal for the gaussian field, it is computationally expensive to use such weights in practice and instead the pseudo-$C_l$ like estimators (Wandelt et al. 2001) are preferred. In such estimators, it is fairly common to adopt uniform weighting, which is usually not optimal.

The next step in the analysis is to measure the power spectra or the correlation functions of the over density field. In this step, one of the major challenges is to account for the effects of the survey window. The survey window or the selection function depends on the survey geometry (mask) as well as the any observational selection effects and the weights applied on the maps as discussed earlier. The modeling of window effects on the pseudo-$C_l$ power spectra has been addressed in detail by Hivon et al. (2002) who introduced the now standard MASTER algorithm to deconvolve the effects of the window to reconstruct the power spectra of the underlying fields. Similarly, formalism for the measurement of the correlation functions was developed by Ng & Liu (1999). The window also affects the correlation functions in a very similar manner by acting as a weighting function and changing the effective scale of measurement (see e.g. appendix D of Singh et al. 2020).

Since proper modeling of the two point functions requires careful modeling of the window effects, one of the major challenges in the LSS measurements is to properly estimate the window. Biases in the window estimation directly propagate into both the power spectra and the correlation functions. The strategies to model window biases include the mode deprojection (Slosar et al. 2004; Leistedt et al. 2013; Elsner et al. 2017) where one can subtract out systematics or down weight the affected modes by including additional terms in the covariance. Similar strategies have also been applied for the case of the correlation functions, e.g. Ross et al. (2012).

The final step before running the inference is to implement the ‘scale cuts’ on the measurements to match the scales of the model. Typically the models we use to analyze the data are validated under certain assumptions and can only model a limited range of scales which is smaller than the range of scales probed by the current data. Applying these scale cuts carefully is necessary for the optimal analysis, i.e. to extract as much information as possible while avoiding the biases from the model outside its range of validity. Recently this issue of scale cuts has shown up in some apparent discrepancies between the analysis in the configuration space (correlation functions) and the Fourier space (power spectra) (see Hamana et al. 2020; Doux et al. 2021). Such issues are concerning as they complicate the interpretation of the inferences drawn from the analysis and need to be carefully addressed.

In this paper we will review the methodology described above, study the optimality of the weights applied on the over density field during the pseudo-$C_l$ like analysis and also test the assumptions made in the methods to model the window functions. Since modeling the window function is important and also very challenging, we will study the impact of the window mis-estimation on the two point functions (we will not discuss the methods to estimate the window, only the methods to account for window mis-estimation). Finally, we will generalize the MASTER algorithm to the correlation functions to study the impact of the scale cuts and devise new methods for more optimal scale cuts. We will use the notation commonly used in the analysis of the angular statistics. However, most of our discussion will be equally applicable to the analysis of the spectroscopic surveys, both galaxies and intensity mapping ones.

The simulations and power spectra calculations for window modeling are performed using the HEALPY (Górski et al. 2005) package, the matter power spectra is obtained using CAMB1 package and the computations of angular power spectra, coupling matrices, correlation functions and any other relevant calculations are performed using the SKYLENS package (Singh et al. in prep).

2 MAP MAKING

We begin with a brief the discussion of turning the catalogs of observables from the telescopes to the maps of the over density fields. For the case of galaxy counts (or intensity mapping), the observed number of galaxies at a position $\mathbf{x}$ can be described as

\[ n_g(\mathbf{x}) = \langle n_g(\mathbf{x}) \rangle (1 + \delta_g(\mathbf{x})), \]

(1)

where $\langle n_g(\mathbf{x}) \rangle$ is the ensemble average of $n_g(\mathbf{x})$ and can be thought of as the expected number of galaxies to be observed at position $\mathbf{x}$ ignoring the effects of noise and the over density field. $\delta_g(\mathbf{x})$ is the overdensity of galaxies at $\mathbf{x}$. We will also use the mean of $\bar{n}_g(\mathbf{x})$ over the whole survey,

\[ \bar{n}_g = \langle n_g(\mathbf{x}) \rangle. \]

(2)

Throughout this paper, $\langle \rangle$ implies the ensemble average of the quantity inside the brackets, which implies an average over many realizations of data and $X$ represents the sample mean of quantity $X$, where by sample mean we will usually imply mean over the given realization of the survey. Quantities with tophat, e.g. $\delta$, represent the measurements from data.

1 \url{http://camb.info}
From eq. (1), a simple way to define windowed over density field is

$$\tilde{\delta}_g(x) = \frac{\langle n_g(x) \rangle}{n_g} - 1 = \frac{n_g(x)}{W_g(x)n_g} - 1 = \delta_g(x),$$

(5)

where $$\tilde{\delta}_g(x)$$ is the selection function which accounts for the observational effects in the survey which modulate the observed density of the galaxies. In the second part of the equation we wrote as function of underlying contaminants, $$C_i$$. In general $$C_i$$ can be a non-linear function and a detailed discussion of window function is outside the scope of this work (see Ross et al. 2020; Rezaie et al. 2020; Everett et al. 2020, for some recent work). We will only focus on the effects of the window on statistics of interest, namely the two point correlations of the fields. We will discuss the effects of window miss-estimation in section 2.2 and in later sections on how it propagates into the two point correlations of the fields. In this work it is also assumed that the window function and the underlying over-density field are uncorrelated. This is in general not true as the window function is correlated with the underlying field due to observational effects such as blending and fiber collisions. Detailed discussion of such effects is left for the future work.

A more popular choice for defining over density maps is to remove the effects of window by dividing with $$W_g(\mathbf{x})$$ (or $$\tilde{W}_g(\mathbf{x})$$ as discussed above), in which case we get

$$\delta_g(x) = \frac{n_g(x)}{\langle n_g(x) \rangle} - 1 = \frac{n_g(x)}{W_g(x)n_g} - 1 = \delta_g(x),$$

(6)

which is free from the effects of window. This is not strictly true as the mask effects are still present and typically mask also needs to be modified to remove the pixels where $$W_g(x)$$ is small.

Many studies in the literature work with the galaxy catalogs instead of maps (see also discussion in section 4.1). In such a case, if we apply systematics weights on galaxies to correct for window effects and use uniform randoms, i.e. $$\langle n_R(x) \rangle/n_g = 1$$, where $$n_R$$ is the number of randoms, we get the estimator in eq. (5) (e.g. Ross et al. 2012; Alam et al. 2016; Elvin-Poole et al. 2018), i.e.,

$$\tilde{\gamma}_{i,j}(x) = \frac{n_g(x)}{W_g(x)n_g} - \frac{n_R(x)}{n_R} = \frac{n_g(x)}{W_g(x)n_g} - 1.$$

(7)

On the other hand, if the systematics weights are applied on the randoms, i.e. $$\langle n_R(x) \rangle/n_g = W_g(x)$$, then we obtain the estimator from eq. (3), i.e.,

$$\tilde{\gamma}_{i,j}(x) = \frac{n_g(x)}{W_g(x)n_g} - \frac{n_R(x)}{n_g}.$$

(8)

In eq. (3), Window ($$W_g(x)$$) acts as weight on the pixels in the map and turns out to be a nearly optimal way to apply the weights. In general, optimal weights for the power spectra estimation of a gaussian random field are (Feldman et al. 1994; Hamilton 1997)

$$W_{FKP}(x) \propto (C_i + N_i(x))^{-1},$$

(9)

where $$C_i$$ is the power spectra of the mode of interest and $$N_i(x) = 1/n_g(x)$$ is the noise power spectra. This is the inverse variance weighting where $$C_i$$ accounts for the sample (cosmic) variance and $$N_i$$ accounts for variance contributed by the noise. Note that the FKP weights are sometimes written as, $$w(x)/(1+n(x)C_i)$$. This definition of weights is valid when the weights are applied to galaxies, while the weights we define in eq. (8) are applied to the pixels of the over density maps.

In the noise dominated regime, ($$N_i \gg C_i$$), the FKP weights reduce to $$W_g$$ in eq. (4). In general the FKP weights depend on the power spectra mode being measured and for a tomographic survey with multiple redshift and $$\ell$$ bins, the proper use of FKP weight for optimal analysis can be computationally expensive. Hence pseudo-$$C_i$$ estimators with a single weighting scheme are usually preferred in practice. For such a case, FKP like weights can be defined by fixing $$C_i = C_0$$, where $$C_0$$ is power spectra at some fixed chosen $$\ell$$. The overdensity field from eq. (3) is then modified to

$$\tilde{\delta}_{g,FKP}(x) = \frac{n_g(x)}{n_g} - 1 = \frac{n_g(x)}{W_g(x)n_g} - 1 = \delta_g(x),$$

(10)

and

$$\tilde{\gamma}_{i,j,FKP}(x) = \frac{n_g(x)}{n_g} - \frac{n_R(x)}{n_g} = \frac{n_g(x)}{W_g(x)n_g} - 1 = \tilde{\gamma}_{i,j}(x).$$

(11)

These weights are similar to those in eq. (4), except that $$C_0$$ here modulates the weights at the higher end to prevent few pixels from having very large weights which can increase the cosmic variance.

The maps for other observables such as galaxy shear can be defined analogously to eq. (3) as

$$\tilde{\gamma}_{i,j}(x) = \frac{n_g(x)}{n_g} \gamma_{i,j}(x) = W_g(x) \gamma_{i,j}(x),$$

(12)

where the window $$W_g(x)$$ depends on the observed number of galaxies in the pixel (as opposed the expected number of galaxies in $$W_g$$)

$$W_g(x) = \frac{n_g(x)}{n_g} = 1 + F(C_i(x)) + \delta_g(x).$$

(13)

$$W_g(x)$$ is relatively easier to estimate compared to $$W_g$$ since it is determined by the position of the source galaxies. That being said, the effects of varying photometry on the shear estimation can be thought of as part of the window and hence the problems associated with the window estimation we discuss later in this paper are applicable to galaxy shear estimation as well. The FKP like weights can also be defined for shear noise and that using the noise scales as $$\sigma_p^2/n_g(x),$$ resulting in

$$\tilde{\gamma}_{i,j,FKP}(x) = \frac{n_g(x)}{n_g} \frac{\sigma_p^2 + n_g(x)C_0}{\sigma_p^2 + n_g(x)C_0} \gamma_{i,j}(x).$$

(14)

For the case of intensity mapping surveys, the window, $$W(x),$$ is same as the mean intensity term, $$\bar{I}$$ (up to a constant), that is sometimes used (e.g. Schaan & White 2021). In fact $$W(x)$$ is a generalization of the $$\bar{I}$$ term in Schaan & White (2021) as we allow for the effects such as foregrounds, detector calibration, etc. to vary over the positions $$\mathbf{x}$$. While Schaan & White (2021) suggested that cross correlations may be useful in determining $$\bar{I}$$ under certain assumptions, unfortunately the cross correlations in a general case may not be of much help with the window modeling. This is because different tracers/surveys have (at least partially) uncorrelated windows and hence the sensitivity of cross correlations to the window is very different from the sensitivity.
of the auto correlations. We will see an example of such an effect in section 3.

In this work, we will restrict ourselves to the pseudo-$C_\ell$ like analysis and will use the weighting scheme of eq. (4) and (10), which is more optimal in most cases than no weighting in (5) (see also discussion in section 2.1). It should be remembered that this weighting scheme increases the cosmic variance and in many practical applications weighting of eq. (9) and (12) can be more optimal.

### 2.1 Shot noise

Under the assumption that the sampling of the discreet tracers follow Poisson distribution, the noise in observed number of galaxies in each pixel is given by

$$\langle n^2_g(x) \rangle = \langle n_g(x) \rangle. \tag{13}$$

When the galaxies are assigned weights, $w$, the noise changes to (Bohm & Zech 2014, we assume weights are deterministic)

$$\langle n^2_{g,w}(x) \rangle = \langle n_g(x) \rangle w(x)^2. \tag{14}$$

The over density field of eq. (3) assigns weights of $1/n_g$ to each galaxy, in which case the noise is given by

$$\langle \delta^2_{N,W}(x) \rangle = \frac{\langle n_g(x) \rangle}{n_g^2} = w_g(x) \frac{1}{n_g}. \tag{15}$$

Averaging over the whole survey, we get

$$\langle \delta^2_{N,W} \rangle = w_g(x) \frac{1}{n_g}. \tag{16}$$

Notice that in eq. (15), two powers of noise, $\delta^2_N$, depend on a single power of $w_g$. Therefore the noise is effectively multiplied by the window given by $\sqrt{w_g(x)}$ (Feldman et al. 1994; Li et al. 2019), i.e.,

$$\sqrt{\langle \delta^2_{N,W}(x) \rangle} = \sqrt{w_g(x)} \frac{1}{\sqrt{n_g}}. \tag{17}$$

Similarly for the over density field in eq. (5), noise is given by

$$\langle \delta^2_N(x) \rangle = \frac{1}{\langle n_g(x) \rangle} = \frac{1}{w_g(x)} \frac{1}{n_g}. \tag{18}$$

Averaging over the whole survey, we get

$$\langle \delta^2_N \rangle = \frac{1}{w_g(x)} \frac{1}{n_g}. \tag{19}$$

With $w_g(x) = 1$ and $w_p(x) \in [0, \infty)$ by construction, it can be shown via Jensen’s inequality that (see also appendix A for an alternate proof)

$$\left[ \frac{1}{w_g(x)} \right] \geq w_g(x). \tag{20}$$

Thus the estimator in eq. (3) is in general has lower noise than the estimator of eq. (5). Both estimator give very similar signal when $w_g(x) = 1$. Qualitatively, this is because the power spectra measurement, pseudo-$C_\ell$, depends on the window power spectra which is the second power of window while the noise depends on the first power of window. Hence the choice of estimator has stronger effect on noise than on signal. We will also compute the response of the pseudo-$C_\ell$ power spectra estimator to the window power spectra in section 3.4 and show that the response is positive, implying that the estimator in eq. (3) will give larger signal in pseudo-$C_\ell$ since it will in general have higher window power. This further increases the signal to noise ratio of the estimator in eq. (3) with respect to the estimator in eq. (5).

The inequality in eq. (20) will get worse as the window becomes more complex ($w_g(x)$ distribution get wider). Some times arguments are made in favor of using eq. (5) to simplify the window modeling but instead it becomes more important to use eq. (3) (or eq. (9)) for more complex windows.

A corollary to eq. (20) is that with the estimator of eq. (5), using $1/n_g$ to model the noise effects in the covariance (analytical or mocks with overly simplified window) will lead to under estimation of the such effects during the analysis, unless the window effects are properly modeled. One simple method is to modify the effective number density of galaxies, but it can still leads to biases in the cross covariance terms of the form $N_tC_\ell$, where $N_t$ is the noise power spectra.

Also, we only considered the galaxy shot noise as source of variance in this section. In appendix B, we derive similar expression in presence of additional source of noise as well as weighted applied to galaxies. For the main part of this paper, we will continue using the galaxy shot noise only version of equations to keep the notation and discussion simpler. The results of the paper do not change for the case considered in appendix B.

While in this paper we only consider noise correlated at zero lag, it is worth remembering that in the general case correlated noise will also depend upon the window and the window for noise is different from the window for the underlying over density field. Thus window effects on noise need to be modeled separately, both for mean noise subtraction from the power spectra measurement as well as in the covariance of both power spectra and correlation functions. The code, SKYLENS, used in this paper is able to handle noise with any user input power spectra and window function.

### 2.2 Window bias

In eq (4), we defined window as a function of some underlying contaminants. Even if those contaminants are known, we still need to estimate the response of the selection function to these contaminants in order to estimate the window. This is a non-trivial task and usually there will be some errors in estimating these responses, in which case our estimates of window can be biased. Without the loss of generality, we can write the estimated window as

$$\tilde{W}_w(x) = (1 + m(x)) W_w(x), \tag{21}$$

where $W_w(x)$ is the true underlying window and $m(x)$ denotes the relative error in the estimated window.

With the erroneous window defined in eq. (21), our estimate of the over-density fields become

$$\tilde{\delta}_{g,w,m}(x) = n_{g}(x) - \frac{\langle n_g(x) \rangle}{n_g}.$$

where $\langle n_g(x) \rangle$ is the biased estimate of the expected number of galaxies. Writing the observed number of galaxies, $n_g(x)$ in terms of the true window and the over density field, we
\[ \tilde{\delta}_{g, w, m}(\mathbf{x}) = W_g(\mathbf{x}) \delta_g(\mathbf{x}) - \delta W_g(\mathbf{x}) \]
\[ = \frac{\tilde{W}_g(\mathbf{x})}{1 + m(\mathbf{x})} \delta_g(\mathbf{x}) - \delta W_g(\mathbf{x}). \quad (23) \]

Where in the second step we wrote the true window in terms of the biased estimate, since that is the window we use in modeling. When \( m \) is small, the multiplicative bias can be estimated as \( \sim -\tilde{W}_g(\mathbf{x}) m(\mathbf{x}) \).

For the over density field defined in eq. (5), we can write the effects of bias as
\[ \delta_g, n, m(\mathbf{x}) = \frac{\tilde{\delta}_g(\mathbf{x}) - m(\mathbf{x})}{1 + m(\mathbf{x})}. \quad (24) \]

Both estimators, \( \tilde{\delta}_{g, w, m} \), and \( \tilde{\delta}_g \) contain the multiplicative bias of \( O(m\delta) \) and an additive bias of \( O(m) \).

If \( mW_g \) is a bias (as opposed to random noise) and is small, we can assume the bias in the window is a linear combination of the underlying contaminants (e.g. Leistedt et al. 2013), \( C_i \),
\[ \delta W_g(\mathbf{x}) = m(\mathbf{x}) W_g(\mathbf{x}) = \sum \alpha_i C_i(\mathbf{x}). \quad (25) \]

Where \( \alpha_i \) are the unknown amplitudes. It is important to stress here that we are only ‘assuming’ the biases in the window function as linear function of contaminants and we do not assume the window function itself to be a linear function of the underlying contaminants. The additive biases can sometimes be constrained by cross-correlating the over density maps with the maps of systematics (Ross et al. 2012; Leistedt et al. 2013). We will study the effects of multiplicative bias in the section 3.4.

Finally it is worth mentioning that the multiplicative and additive biases in the shear estimation can also be thought of as part of the window in a similar vein as above and our discussion of the window systematics here and in later sections is applicable to all tracers.

## 3 POWER SPECTRA

Once the over density maps are generated, the next step in the standard two point analysis is to compute the power spectra or the correlation functions of the maps. In this section we will discuss the power spectra of the windowed maps, the algorithm to model the window effects and the methods to account for the uncertainties in the window estimation.

### 3.1 Pseudo-\( C_\ell \) power spectra

Since the observed over density field is multiplied by the window, its Fourier transform\(^2\) is convolved with the Fourier transform of the window. Computing the power spectra of this windowed field results in the pseudo-\( C_\ell \) estimator \( \hat{D}_\ell \), whose expected value is related to the true power spectra via a coupling matrix (Hivon et al. 2002),
\[ \hat{D}_\ell = M_{\ell\ell'} C_{\ell'} + M_{\ell\ell'} N_{\ell'} , \quad (26) \]
where \( M_{\ell\ell'} \) is the coupling matrix for the signal part and \( M_{\ell\ell'}^N \) is the coupling matrix for the noise in the observed field (as discussed in section 2.1.2 noise and signal have a different window). Here summation over \( \ell' \) is implied and throughout the paper we will use the Einstein summation convention to imply sum over repeated indices. To keep notation simpler, we will omit the the noise term from the equations. This will not affect our discussion since it is easy to generalize to the noise case.

The coupling matrices is given by
\[ M_{\ell, \ell'} = \frac{(2\ell' + 1)}{4\pi} \sum_{\ell''} W_{\ell''}(2\ell'' + 1) \begin{pmatrix} \ell & \ell' & \ell'' \\ s_1 & s_1 & 0 \end{pmatrix} \times \begin{pmatrix} \ell & \ell' & \ell'' \\ s_2 & s_2 & 0 \end{pmatrix} , \quad (27) \]
where where \( \ell, \ell' \) are as defined in eq. (26), \( W_{\ell''} \) is the power spectra (pseudo-\( C_\ell \)) of the window defined at \( \ell'' \) (cross power spectra of two windows for cross correlations), \( s_1, s_2 \) are the spins of the two tracers being correlated to obtain the \( \hat{D}_\ell \).

For the case of spin-2 quantities, e.g. galaxy shear or CMB polarization, the power spectra of \( E \) and \( B \) modes is given by,
\[ D_{\ell}^{EE} = M^+ C_{\ell}^{EE} + M^- C_{\ell}^{BB} \quad (28) \]
\[ D_{\ell}^{BB} = M^+ C_{\ell}^{BB} + M^- C_{\ell}^{EE} \quad (29) \]

In the second equality we wrote \( E/B \) pseudo-power spectra in terms of true \( E/B \) power spectra using spin-2 coupling matrices, given by
\[ M_{\ell, \ell'} = \frac{(2\ell' + 1)}{4\pi} \sum_{\ell''} W_{\ell''}(2\ell'' + 1) \begin{pmatrix} 1 \pm (1 - \ell \ell' \ell'' - 1) \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} \times \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} . \quad (30) \]

In this section we will assume that the \( B \text{-mode power spectra is zero} \), i.e. \( C_{\ell}^{BB} = 0 \) and will therefore only use \( D_{\ell}^{EE} = M^+ C_{\ell}^{EE} \) and \( D_{\ell}^{BB} = M^- C_{\ell}^{EE} \). We will discuss the more general case in section 4.2.1. Hereafter, we will also drop the superscripts \( \pm, EE, BB \) on the \( M \) and power spectra unless necessary for clarity.

In general \( M_{\ell, \ell'} \) is a rectangular matrix with \( \ell_{\text{max}}, \ell'_{\text{max}} \gg \ell_{\text{max}} \). In practice, the modern survey windows are large enough (narrow in Fourier space) such that with \( \ell_{\text{max}} \sim O(1000 - 5000) \) and \( C_{\ell} \propto \ell^{-2} \), \( \ell_{\text{max}} \) approximation is sufficient for accuracy of up to few percent near \( \ell_{\text{max}} \) (the examples shown later in this section satisfy this assumption). With such an approximation, we need to estimate the window up to \( \ell' \sim 2\ell_{\text{max}} \). These approximation however may not work if the window has large enough power out to very high \( \ell' \) or in the case when the power spectra does not fall fast enough with \( \ell \) (e.g. noise spectrum). In such a case \( \ell_{\text{max}} \gg \ell_{\text{max}} \) maybe necessary or one may have to resort to trickery such as apodizing the window in order to tame the coupling matrix. We will discuss an example of such effects in a later section when we reconstruct power spectra from correlation functions.

### 3.2 Master algorithm

As discussed in the previous section, one of the challenges in the pseudo-\( C_\ell \) analysis is that \( M_{\ell\ell'} \) is an \( O(\ell_{\text{max}}^2) \) matrix
where $\ell_{\text{max}} \sim 1000 - 5000$ or even larger. While computation of $M_{\ell'}$ can be easily handled on the modern day computers, it is still desirable to reduce the dimensionality of the $M_{\ell'}$ for the purpose of sampling during the inference. De-convolving unbinned $D_\ell$ to reconstruct $C_\ell$ by inverting eq. (26) is also not straightforward in the presence of large noise (remember noise has a different coupling matrix and noise contributions scale as $M^{-1}[\langle M^\dagger N_\ell \rangle - \langle M^\dagger N_\ell \rangle]$). Thus it is generally desirable to bin the noisy $D_\ell$ measurements to reduce noise effects. Hivon et al. (2002) presented the MASTER algorithm which allows us to work with binned quantities, reducing the complexity of the problem to $O(N_b^2)$, where $N_b$ is the number of bins.

In the MASTER algorithm, we simply write the eq. (26) in terms of the binned quantities,

$$\hat{D}_{\ell_b} = M_{\ell_b \ell_b} C_{\ell_b} + N_{\ell_b},$$

(31)

The binned $M_{\ell_b \ell_b}$ is given by

$$M_{\ell_b \ell_b} = P_{\ell_b \ell_b} M_{\ell'\ell'} Q_{\ell_b \ell_b},$$

(32)

where $P_{\ell_b \ell_b}$ and $Q_{\ell_b \ell_b}$ are binning and inverse binning operations respectively and can be analytically written under the assumption that $\ell(\ell + 1)C_\ell \sim$ constant,

$$P_{\ell_b \ell_b} = \begin{cases} \frac{1}{\Delta \ell_b} \ell(\ell + 1), & \ell \in b \\ 0 & \text{otherwise} \end{cases},$$

(33)

$$Q_{\ell_b \ell_b} = \begin{cases} \frac{\Delta \ell_b}{\ell(\ell + 1)}, & \ell \in b \\ 0 & \text{otherwise} \end{cases},$$

(34)

where $\ell \in b$ is true when $\ell$ belongs to the bin centered on $\ell_b$ and $\Delta \ell_b$ is the bin size. In some implementations of the MASTER algorithm, $C_\ell$ is assumed to be constant within the bin (e.g. Alonso et al. 2019), in which case $P$ and $Q$ take values of 0 or 1. We will refer to such approximation as cMASTER.

The eq. (31) can be inverted to reconstruct the power spectra from the pseudo-$C_\ell$

$$\hat{C}_{\ell_b} = M_{\ell_b \ell_b}^{-1} \hat{D}_{\ell_b},$$

(35)

where we have taken the pseudo-inverse of the binned coupling matrix.

In fig. 1, we see the comparison of the $D_\ell$ predictions using both MASTER and cMASTER algorithms. In fig. 1 as well as in most other demonstrations later in the paper, we will employ two choices of the window, $W_1$ and $W_2$. $W_1$ is chosen to be a more realistic and is similar to the galaxy shear window where the sampling is determined by galaxy positions and hence window power spectra is similar to galaxy power spectra. For $W_2$, we increase the complexity of window further by raising the power spectra of window further at some scales (see appendix F for further details). $W_2$ acts as a stress test for our methodology and comparison with $W_1$ also allows us to see where and how the biases are introduced by different methods.

In fig. 1 we observe that as the complexity of the window increases, both algorithms result in biased predictions. This is because coupling matrix becomes broader with more complex window and since $D_\ell$ is effectively a weighted sum of $C_\ell$, a broader coupling matrix increases the impact of biases introduced by the approximations made in the definition of binning operators $P$ and $Q$. Furthermore, cMASTER has larger biases since it makes worse approximations in $P$ and $Q$. These biases also propagate to the deconvolved power spectra (see fig. 2) and necessitate that some corrections be applied to the theory power spectrum before it can be compared with the biased estimators of either $D_\ell$ or deconvolved power spectra (see discussion in Alonso et al. 2019), i.e. we define:

$$C_{\ell_b \ell_b} = M_{\ell_b \ell_b}^{-1} B_D M_{\ell' \ell'} C_{\ell' \ell'} = M_{\ell_b \ell_b}^{-1} B_D M_{\ell' \ell'} C_{\ell' \ell'},$$

(36)

$B_D$ is the binning operator similar to $P$ and is defined in next section. In the second equality we defined $M_{\ell_b \ell_b}^{-1} B_D M_{\ell' \ell'}$. Using such an estimator is not strictly necessary, as one can simply work with the binned pseudo power spectra instead, i.e.

$$D_{\ell_b} = M_{\ell_b \ell_b}^{-1} C_{\ell' \ell'},$$

(37)

where $M_{\ell_b \ell_b}^{-1} B_D M_{\ell' \ell'}$. This is a simple convolution+binning operation and will not be branded as MASTER algorithm in this paper.

In the next section we discuss the proper expression for $P$ and $Q$ which lead to recovery of unbiased results which can be compared directly with theory power spectra.

### 3.3 iMASTER

As discussed in the previous section, the biases in the MASTER algorithm are sourced by the assumptions made in implementing the effects of binning. Now we derive the MASTER algorithm without making such assumptions.
The binned version of eq. (26) can be written as

\[
B_{D,\ell_0} D_{\ell} = B_{D,\ell_0} M_{\ell} B_{C,\ell_0}^{-1} B_{C,\ell_0}^* C_{\ell} + B_{D,\ell_0} M_{\ell}^2 B_{C,\ell_0}^{-1} B_{C,\ell_0}^* C_{\ell_0} N_{\ell_0}
\]

(38)

where \( B_{D,\ell_0} D_{\ell} \) is the binning operator. We introduced separate binning operators, \( B_{C} \), \( B_{D} \) and \( B_{N} \), which operate on \( C_{\ell}, D_{\ell}, N_{\ell} \) since the coupling matrix is not symmetric in general and \( \ell, \ell' \) can have different range. For examples shown throughout this work we will make the approximation that \( \ell_{\text{max}} \approx \ell_{\text{max}} \) (coupling matrix is not too wide) in which case \( B_{D} = B_{C} \) can be used. We will simply assume \( B_{D} = B_{C} = B \) and omit the subscripts hereafter unless required for clarity. Furthermore we assume noise is estimated and subtracted before binning and will neglect the effects of \( B_{N} \) as well. We will not assume any particular form for the binning operator and will keep our discussion general so that our method does not depend on the binning operator unlike the standard Master algorithm. We will discuss some particular forms of binning operator later in this section and the form of the binning operator used in examples of this paper is defined in eq. (45).

Binning data in general leads to loss of information and hence the binning operation can not be inverted. Thus \( B^{-1} \) in general is not defined. However, if we have good model for the underlying signal, then it is possible to obtain \( B^{-1} \). In Hivon et al. (2002) model for \( C_{\ell} \) was assumed to be \( \ell (\ell + 1) C_{\ell} \sim \text{constant} \) which allowed for an analytical expression for \( Q_{\ell}, t_{\ell} \) which is the inverse of binning operator \( P_{\ell i} \). However, since we hopefully have a better model for power spectra, the assumptions about \( C_{\ell} \) are not necessary and we can simply write the inverse of binning operator as

\[
B_{C,\ell_0}^{-1} = \begin{cases} C_{\ell_0}^{-1}, & \ell \in b \\ 0 & \text{otherwise} \end{cases}
\]

(39)

where \( C_{\ell_0} = B C_{\ell} \) is obtained after binning the model \( C_{\ell} \) and can be described as power spectra at some effective center of the bin, \( t_{\ell_0} \). The proper choice of \( t_{\ell_0} \) is the effective \( \ell_0 \) at which the binned \( C_{\ell_0} \) is measured and is given by

\[
t_{\ell_0} = \frac{B_{C,\ell_0} \ell C_{\ell}}{B_{C,\ell_0} \ell}
\]

(40)

The Master binned coupling matrix is then

\[
M_{\ell_0,\ell_0'} = B_{D,\ell_0} M_{\ell_0} B_{C,\ell_0'}^{-1} C_{\ell_0'}
\]

(41)

Fig. 2a shows the comparison of \( D_{\ell} \) obtained using the iMaster algorithm with that of 1000 gaussian simulations, similar to fig. 1. The model predictions are consistent with those of simulations to well within 1%, even for the case of more complex window \( W_2 \). Figs. 2b-2d show the comparison of the \( C_{\ell} \) reconstructed from the \( D_{\ell} \) of simulations by convolving with the (pseudo) inverse of binned coupling matrix as described in eq. (35). Here again we observe that the iMaster yield unbiased results to within 1% while both MASTER and cMASTER yield biased results as in the case of \( D_{\ell} \) predictions in fig. 1.

It is important to stress here that the \( C_{\ell} \) from the iMaster algorithm can be directly compared with the unbinned prediction computed at \( t_{\ell_0} \), i.e. the binning is performed analytically. This algorithm also does not require any additional correction to be applied unlike the method in eq. (36) (Alonso et al. 2019). Furthermore, since the algorithm allows for cleaner reconstruction of \( C_{\ell} \), it is also more optimal in the analysis with scale cuts as it prevents the mixing of information between different scales in a much cleaner way. For example, in eq. (38) if we have scale cuts of form of \( \ell < \ell_{\text{max}}, \text{iMaster} \) prevents the information from \( \ell \ell > \ell_{\text{max}} \) modes leaking into our analysis. At the same time the algorithm also captures the information about \( \ell < \ell_{\text{max}} \) modes which had leaked into high \( \ell \) modes (\( \ell > \ell_{\text{max}} \) of the pseudo-\( C_{\ell} \) estimator (this requires \( \ell_{\text{max}} > \ell_{\text{max}} \) in eq. (38) and also using different binning operators \( B_{D} \) and \( B_{C} \)). These are the primary advantages of the Master algorithm.

Due to \( B^{-1} \), the estimator is now dependent on the underlying theoretical model. If one feels uncomfortable with such dependence, a half binned version, \( M_{\ell,\ell'} = B_{D} M_{\ell,\ell'} \) (see eq. (37)) can be used, with larger computational costs (both time and system memory). As discussed earlier, iMaster is also more optimal in terms of scale cuts and preventing mixing of information between different scales. It may be advantageous to use iMaster in the form similar to eq. (36), i.e.,

\[
C_{\ell_0,\ell_0'} = M_{\ell_0,\ell_0'} B_{D} D_{\ell_0} M_{\ell_0,\ell_0'} C_{\ell_0} = M_{\ell_0,\ell_0'} B_{D} M_{\ell_0,\ell_0'} C_{\ell_0}.
\]

(42)

where \( M_{\ell_0,\ell_0'} \) is computed at fiducial cosmology and is kept fixed. To reiterate, one of the advantages of iMaster is that \( C_{\ell_0,\ell_0'} = C_{\ell_0} \) and we can use the un-binned theory to compare with the measurements. Only in the case where the model varies strongly enough such that variations in iMaster are significant compared to noise (while such variations are found to be small in our examples, similar tests should be performed at the time of analysis), using the form of eq. (42) will be necessary, where theory is properly binned using a fixed Master computed with fiducial model. It should also be noted that the model dependence of iMaster is weak, as \( B^{-1} \) is only sensitive to the slope of the power spectra within the bin. As long as the ratio does not vary significantly the model dependence should not introduce any significant biases. In fig. 2, the \( \chi^2 \) per bin for different curves obtained using iMaster is always less than 2/25 and for noiseless simulations (not shown) is less than 0.5/25. Therefore the model dependence introduced by iMaster is not much different from the model dependence from other components in the analysis, e.g. converting distances to redshifts or computing covariance matrices at fixed cosmology. If the model dependence of iMaster does matter in the inference problem, i.e. the bias due to assumptions in \( B^{-1} \) is larger than the uncertainties, it is also likely to be an indicative of a problem with binning, namely that bins are too wide and it maybe better to move to narrower bins to capture more information. It is not optimal to use overly broad bins with iMaster. If using narrow bins is not possible, then obtaining model using eq. (42) may be necessary.

Now we discuss some particular forms of the binning operators. It should be remembered that in deriving the iMaster algorithm we did not assume any particular form of the binning operators and any sensible choice of binning will work.

For noisy measurements with a given covariance, the binning operator can be defined using the optimal estimator of the mean \( D_{\ell} \) within the bin (assuming gaussian distribu-
Note: The text seems to be discussing comparisons of different observables (Pseudo-\(C_\ell\), Galaxy-galaxy, Galaxy-shear, and Shear-shear) using different window functions (W1 and W2) and comparing them with theoretical predictions. The key points include:

- The effective scale \(\ell\) can be directly compared with underlying binned theoretical models or equivalent unbinned power spectra to perform an unbiased comparison between theory and data.

- The iMaster and Master algorithms give unbiased power spectra (within 1% error) while Master and cMaster algorithms give biased results on scales where the window has large power.

- Comparisons of \(D_\ell\) and \(C_\ell\) of different observables, obtained using different window functions, W1 and W2 (see appendix F).

- The binning operator can be obtained at the same time as computing the covariance with the effective number of modes.

- The optimal binning operator in eq. (43) depends on the unbiased covariance of the power spectra. The binning can be obtained at the same time as computing covariance with little additional computing cost. \(B\) also depends on the particular power spectra being considered and hence each cross and auto correlation in a tomographic analysis will require different \(B\). If such properties of \(B\) are not desirable, an approximation to \(B\) can be used where each \(\ell\) is weighted by the effective number of modes (ignoring the effects of noise).
Simulations and theory calculations shown in this work use the binning operators from eq. (45).

### 3.4 Window bias

In this section we discuss the response of the pseudo-$C_\ell$ estimator to the errors in estimating the window functions. We begin by noticing that the $D_\ell$ is symmetric in response to the window and the underlying over density field and therefore we can write (see appendix C):

$$D_\ell = M_{\ell,\ell}^W W_{\ell,\ell},$$

(46)

where $M_{\ell,\ell}^W$ is the response of $D_\ell$ to the window and depends on the $C_\ell$. Relative error in $D_\ell$ due to errors in window power spectra can be written as

$$\frac{\delta \log D_\ell}{\partial W_{\ell,\ell}} = \frac{1}{D_\ell} M_{\ell,\ell}^W.$$

(47)

$\frac{1}{D_\ell} M_{\ell,\ell}^W$ can be used to define the upper limits on the bias in window power spectra $\delta W_{\ell,\ell}$ given an upper limit on the percentage bias $D_\ell$ that is acceptable.

Given the multiplicative and additive errors in the window as defined in section 2.2, we can also write the bias in $D_\ell$ as

$$\delta D_\ell = M_{\ell,\ell}^W m_{\ell,\ell} + a_\ell,$$

(48)

where $m_{\ell,\ell}$ is the power spectra of the multiplicative bias $(m(x)W(x))$ and $a_\ell$ is the power spectra of the additive bias. It is important to note that multiplicative and additive biases do not necessarily have to be present together. For example, from eq. (21), it is possible that the estimated and true windows have the very similar power spectra but the two windows are not very correlated. In such a case, window errors will behave like noise, resulting in additive bias $a_\ell$ but negligible multiplicative bias. Efficacy of window weighting will also be lost in such a case.

For window biases which may be described by eq. (25), under the assumption that the contaminant maps are not correlated, we can write (Ross et al. 2012; Leistedt et al. 2013; Elsner et al. 2017)

$$a_\ell = \sum_i a_i D_{\ell,i},$$

(49)

$D_{\ell,i}$ is the power spectra of the contaminant maps. $a_\ell$ can be estimated by cross correlating the estimated over density map (which includes systematics) with contaminant map. Cov is the joint covariance of the auto and cross correlations and eq. (51) denotes the variance of $a_\ell$. The method is only valid under the assumptions of linear expansion in eq. (25) and that the contaminant maps are independent. If the contaminant maps are not independent, one can use a PCA like method to project the contaminant maps into maps of orthogonal linear combinations which are uncorrelated and then define $a_\ell$ and cross correlations over the new set of maps.

The cross correlation in eq. (50) is not sensitive to the $m_{\ell,\ell}$ as it shows up in the third order term, whose expectation is zero when window and the underlying overdensity field are uncorrelated. If the corrections from the above procedure are reliable and are applied to the window, it may reduce the effects of multiplicative bias as well.

For the case of multiplicative bias, fig. 3a shows the $\frac{1}{M_{\ell,\ell}^W}$ matrix. At low $\ell$ the matrix has nearly scale independent effects along the column, especially at $\ell = 0$. This accounts for the effects of the misestimation of the largest scales of the window, primarily sourced by the mask, and includes the effects of the mean of window or the $f_{sky}$ factor. The diagonal of the matrix is sub-dominant at low $\ell$ but it increases as $\sim 2\ell + 1$, becoming large at high $\ell$. Typically when modeling the window in power spectra such high $\ell$ effects are ignored. However, in fig. 3b, we see that this effect leads to rapidly increasing biases at higher $\ell$, which dominate the information during the cosmological inference. Thus it is important to carefully model window out to high $\ell$ and properly account for any uncertainties in such modeling. Current modeling, e.g. Ross et al. (2012, 2020), uses HEALPix maps with resolution of order $N_{side} = 512$, which may not be enough in the case of moderately complex windows even for $\ell_{max} \sim 1000$ (roughly similar to scale cuts in current weak lensing $3 \times 2$ analysis) and in the upcoming analysis such with LSST where we may wish to extend the analysis to $\ell_{max} \gtrsim 5000$, a much more careful characterization of the window will be needed.

To account for uncertainties in window effects, one can use some parametric function for $m_{\ell,\ell}$ which can be added to the $D_\ell$ model using eq. (48) and its parameters can be marginalized over. We can also use the analytical marginalization (Bridle et al. 2002), in which case the the contributions from window uncertainties can be added to the covariance of the $D_{\ell,i}$ and the additional contribution to the covariance can be written as

$$\text{Cov}^W(D_{\ell,i}, D_{\ell,f}) \approx M_{\ell,\ell} W m_{\ell,\ell} T M^W_T + a_\ell a_\ell^T,$$

(52)

where $a_\ell$ and its uncertainty can be obtained from eqs. (49)–(51). In general, if we have an estimation of the covariance of the window power spectra, $W_{\ell,i}$, we can write

$$\text{Cov}^W(D_{\ell,i}, D_{\ell,f}) \approx M_{\ell,\ell} W \text{Cov}(W_{\ell,i}, W_{\ell,f}) M^W_T + \text{Cov}(W_{\ell,i}, W_{\ell,f}).$$

(53)

This expression is an approximation since a full covariance with the window and the over density field will require an expansion of 8-point function, which is tedious and computationally expensive even for the case of gaussian field. A detailed study of the window uncertainties in a realistic setting of LSS windows for the case of DESI survey will be presented in an upcoming work (Karim, Rezai & Singh in prep).

There are also cases where the window is well estimated but the multiplicative error can still have significant contributions to the covariance. One such case is when estimating the jackknife covariance, where each jackknife sample has a slightly different window and these seemingly minor differences can bias the covariance (Yu, Singh et al. in prep).

Finally, as discussed in section 2, it is also tempting to conclude that we can use cross correlations between different tracer/surveys to ‘self calibrate’ the effects of window.

MNRA 000, 1–77 (2021)
However, this in general does not work unless the windows between tracer/surveys are perfectly correlated. An example of such an effect can be observed on figs. 1 and 2, where galaxies and shear have uncorrelated window by design and hence the cross correlation, \( g_{\gamma} \), is almost completely free of the window effects except for the very large scales (low \( \ell \)), where windows are partially correlated due to the mask effects. Thus it is in general not possible to use the cross correlation, \( g_{\gamma} \), to understand the window effects in auto correlations, \( g_{\gamma} \gamma \) in our case.

4 CORRELATION FUNCTIONS

In this section we discuss the two point function measurements in the configuration space, namely the correlation functions. We begin with brief introduction of correlation functions and will discuss the estimators, effects of windows, the methods to apply scale cuts and the generalization of the iMASTER algorithm to reconstruct the power spectra from the correlation functions.

The curved sky correlation function can be written as the Hankel transform of the power spectra,

\[
\xi(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} -s_{1}d_{t,s_{2}}(\theta) C_{t} = \mathcal{H}_{s_{1}s_{2}}C_{t},
\]

where \( s_{1}d_{t,s_{2}} \) is the wigner-d matrix, with \( s_{1},s_{2} \) being the spins of tracers being correlated and in the second equality we have written the sum as a matrix multiplication, with \( \mathcal{H} \) being the Hankel transform operator. The sum is over all \( \ell \) and therefore computing \( \xi \) to an arbitrary accuracy can be expensive. Fortunately, \( C_{t} \) drops with \( \ell \sim \ell^{-2} \), therefore very high \( \ell \) modes contribute very little and the summation can be truncated at \( \ell \sim a\pi/\theta \), where \( a \sim 5-10 \) is a suitably chosen constant. We will see in section 4.2 that one can also work with the binned quantities similar to the discussion for the pseudo-\( C_{t} \) to further speed up the calculation. But first we discuss the estimator used to measure the correlation functions.

4.1 Landy-Szalay estimator

The correlation functions measure the excess probability of finding galaxies around other galaxies at a given separation. These are measured by counting the excess number of pairs of galaxies at a given separation relative to a distribution of randoms. The commonly employed Landy-Szalay estimator (Landy & Szalay 1993) is given by

\[
\xi(\theta_{b}) = \frac{B_{\theta}(DD - DR - RD + RR)}{B_{\theta}RR},
\]

where \( B_{\theta} \) is the binning operator, \( D \) denotes the galaxies, \( R \) denotes the randoms that correspond to the galaxy sample and different two points, e.g. \( B_{\theta}DD \), denote the count of auto and cross pairs of galaxies and randoms with separation \( \theta \) that falls within the bin \( \theta_{b} \). It can be shown that when randoms follow the window function, LS is an optimal estimator to compute the two point correlation functions (Landy & Szalay 1993; Singh et al. 2017).

We now show that the LS estimator is equivalent to correlating the over density maps as defined in section 2. The pairs counts can be written as integral over the galaxy density field, i.e.

\[
DD(\theta) = \int d\theta' n_{g}(\theta)n_{g}(\theta' + \theta).
\]
\[ \xi(\theta) = \frac{\int \varphi d\theta' \sin(\theta')}{\int \varphi d\theta' \sin(\theta')} \] is the integral over the volume element in \( \theta' \). From this we can show that the eq. (55) is equivalent to
\[ \xi(\theta) = \frac{B_0 \int \varphi (n_g(\theta') - n_R(\theta') + n_R(\theta' + \theta))}{B_0 \int \varphi (n_R(\theta') - n_R(\theta' + \theta)) \cdot \theta'}. \]

Using the definition of \( \delta \) from eq. (3) and (7), we get
\[ \xi(\theta) = \frac{B_0 \int W(\theta') \delta_b(\theta') W(\theta' + \theta) \delta_b(\theta' + \theta)}{B_0 \int W(\theta') \delta_b(\theta')} \cdot \theta', \]
where \( \delta_b(\theta') \) is the correlation function of window. From eq. (58), the LS estimator is same as correlating the over-density field, as long as the pixel size of the map is small enough to measure the minimum scale of interest. Computing correlation functions via pixels can be cheaper when we have more than one galaxy (or more accurately, one random point) per pixel on average. Further gains can also be made by taking advantage of the regularized nature of the pixel grid. Since use of randoms is simply a monte-carlo method to account for window effects in the estimator, direct correlation of maps will also be free from additional noise introduced by the randoms.

As discussed in section 2, when computing correlation functions from galaxy catalogs, usually one of following two approaches is adopted. In the first case, randoms are uniformly distributed on the sky and the galaxies are weighted by \( \delta(\theta') \). This approach is commonly used (e.g. Ross et al. 2012; Alam et al. 2016; Elvin-Poole et al. 2018) and is equivalent to using the over density map as defined in eq. (5). In the second method, the randoms are weighted by the window, i.e. \( \int n_R(\theta') - n_R(\theta' + \theta) \cdot \theta' \). This approach is equivalent to correlating the maps as defined in eq. (3). In terms of optimality, same arguments as presented in section 2 applies to the question of weighing galaxies vs randoms, i.e. in general it is better to apply weights to randoms.

It is also worth noting that the \( B_0 \) \( \delta(\theta') \) term in the denominator is not strictly necessary if the window effects are properly accounted for in the modeling. The use of this term in the estimator can also lead to the sub-optimal results as it up-weights the noisiest modes where \( B_0 \) \( \delta(\theta') \) is low. It can be more optimal to instead use a simple normalization constant, e.g. \( A_W = \int \varphi(\theta')d\theta' \) (or \( A_W = 1 \)). In such a case, it can be easily shown using the equations in appendix E1 (replace \( w(\theta') \) with \( \delta(\theta') \)) that the correlation function and the pseudo-\( \hat{C}_t \) estimators are identical when the measurements are performed over the full range of scales, \( \theta \in [0, \pi] \) and \( \ell \in [0, \infty] \). In the following section we relax this condition and study the relations between correlation functions and power spectra over a limited range of scales.

### 4.2 iMaster for correlation functions

In the previous section, we have studied the correlation function estimators and showed their equivalence to the power spectra estimators. Now we will study the relations between correlation functions and power spectra for the practical cases where the measurements and the models are defined over a limited range of scales. We will generalize the iMaster algorithm to reconstruct power spectra from the correlation function measurements done over a limited \( \theta \) range and will also derive the expressions for properly implementing scale cuts on the correlation functions when the model is defined in Fourier space.

From eq. (58) and appendix D, the correlation function for a windowed field is given by
\[ \xi(\theta) = \frac{B_0 \xi_W(\theta) \theta_B \ell C_t}{B_0 \xi_W(\theta)}. \]

We can use the trick from section 3.3 and bin both \( \xi \) and \( C_t \)
\[ \xi(\theta_B) = \frac{1}{B_0 \xi_W(\theta)} B_0 \xi_W(\theta) \theta_B \ell C_t \cdot \ell, \]
where \( B_0 \) is binning operator in \( \theta \), acting on \( \xi \) and \( C_t \) is the binning operator in \( \ell \), acting on \( C_t \). Therefore we can write the operation in terms of binned quantities, where the binned Hankel transform is given by
\[ \hat{H}_\theta, \ell C_t = \frac{1}{B_0 \xi_W(\theta)} B_0 \xi_W(\theta) \hat{H}_\theta, \ell C_t \cdot \ell. \]

As discussed in previous section 4.1, the \( B_0 \xi_W(\theta) \) term in the denominator of eq. (59) can be replaced with a constant \( A_W \), in which case the binned Hankel transform changes to,
\[ \hat{H}_\theta, \ell C_t = \frac{1}{A_W} B_0 \xi_W(\theta) \hat{H}_\theta, \ell C_t \cdot \ell. \]

\( \theta_B \) is the effective \( \theta \) at which the correlation function is measured and is given by
\[ \theta_B = \frac{B_0 \xi_W(\theta) \xi(\theta)}{B_0 \xi_W(\theta) \xi(\theta)}. \]

An aside, \( \hat{H}_\theta, \ell C_t \) can be used to transform \( C_t \) to binned correlation function in \( O(N_{\ell B}^2) \) time instead of \( O(N_{\ell T}^2) \), where \( N_{\ell B} \gg N_{\ell T} \), thereby speeding up the calculations without compromising the accuracy.

Finally, we can invert Hankel transform to reconstruct the binned power spectra as
\[ C_\ell = \hat{H}_\ell, \theta_B \cdot \xi(\theta_B). \]

In general, in the presence of scale cuts in \( \theta \), the inversion of \( \hat{H}_\ell, \theta_B \) can be unstable. It is better to define inverse of \( \hat{H}_\ell, \theta_B \) from the definition of the inverse transform, where
\[ C_\ell = 2\pi \int d\theta \sin(\theta) \hat{H}_\ell, \theta_B \cdot \xi(\theta). \]

Replacing the integral with sum over discrete values of \( \theta \), we can write
\[ C_\ell^{-1} = 2\pi \Delta \theta \sin(\theta) \hat{H}_\ell, \theta_B \cdot \xi(\theta). \]

which can then be binned to obtain
\[ \hat{H}_\ell, \theta_B = B_\ell^{-1} C_\ell^{-1} B_\ell^{-1}. \]

Notice that in this process we are reconstructing the true power spectra and not the pseudo-\( \hat{C}_t \) as long as the correlation functions are measured over the full range, \( \theta \in [0, \pi] \).

The effects of the survey window are absorbed in the \( \hat{H}_\ell, \theta_B \), via the \( B_\ell^{-1} \) operation. \( B_\ell^{-1} \) is defined similar to eq. (39) using the ratio \( \xi(\theta)/\xi(\theta_0) \).

In practice, the correlation functions are usually measured over a limited range of scales, in which case we define
\[ \xi_{\text{cut}}(\theta) = \xi(\theta) w(\theta), \]
where $w(\theta)$ is a weight function (usually a top hat function with the limits of $\theta_{\text{min}}, \theta_{\text{max}}$) that applies the scale cuts. Inverting $\xi_{\text{cut}}$, we will get
\[
H_{\xi,\theta_{\text{cut}}}^{-1}(\theta) = M_{\ell,\ell'} C_{\ell,\ell'}^{-1},
\]  
(69)
where $M_{\ell,\ell'}$ is the coupling matrix as defined in eq. (27) and eq (41), this time with $w(\theta)$ acting as window (see appendix E for derivation). Note that when $\theta$ range is small, the coupling matrix can be rather broad in which case separate binning operators on $B_D$ and $B_C$ may be required similar to eq (41).

In figure 4, we see the $D_\ell$ obtained by the inverse Hankel transform of the correlation function. For a wide range, $\theta \in [0.1, 600]$ arcminutes, the $D_\ell$ from inverse Hankel transform of correlation function and from convolving $C_\ell$ are consistent. However, for larger $\theta_{\text{min}} = 10$ arcminutes, the results from convolving $C_\ell$ are biased. This is because the coupling matrix $M_{\ell,\ell'}$ is very broad and the range $\ell \in [0, 3000]$ used here for $C_\ell$ calculations is not enough. This is demonstration of the case where we require $\ell', \ell'' \gg \ell$ when computing the coupling matrix (see discussion after eq. (27)). The figure also shows an attempt to reduce the complexity of the coupling matrix by apodizing the scale cut window $w(\theta)$. Here a simple procedure was adopted where the $w_\ell$ is multiplied with a cosine function which goes from one to zero for $\ell \in [100, 1000]$ and then $w_\ell$ is transformed back into $w(\theta)$. This apodized window effectively brings back some power from $\theta < 10$ arcminutes and helps in partially reducing the bias. In practical applications, one will have to experiment with a few different apodization schemes depending on the measurements being performed and scales being used to obtain good results.

From eq. (69), the power spectra can be reconstructed from correlation functions as
\[
C_{\ell}^{\ell'} = M_{\ell,\ell'}^{-1} H_{\xi,\theta_{\text{cut}}}^{-1}(\theta) = M_{\ell,\theta_{\text{cut}}}^{-1} C_{\ell,\ell'}^{-1},
\]  
(70)
where we defined
\[
M_{\ell,\theta_{\text{cut}}}^{-1} = M_{\ell,\ell'}^{-1} H_{\xi,\theta_{\text{cut}}}^{-1}.
\]  
(71)
The covariance of reconstructed power spectra is
\[
\text{Cov}(C_{\ell_1}, C_{\ell_2}) = M_{\theta_1,\theta_2} \text{Cov}(\xi_{\theta_1}, \xi_{\theta_2}) M_{\theta_2,\theta_1}^T.
\]  
(72)
Note that unlike existing methods in the literature, e.g. Joachimi et al. (2021), the MASTER method we discuss here returns the unbiased $C_\ell$ power spectra at the effective bin centers, $\ell$, and removes any effects of mode mixing from the correlation function measurements done over a limited range of scales. The method also accounts for the window effects which has not been done before in literature to the best of my knowledge. As in section 3.3, the results from eq. (70) and (78) can directly be compared with the model $C_\ell$ predictions and do not require any additional operations such as binning or any corrections to be applied.

In fig. 5, we see the comparison of the reconstructed power spectra from both correlation function and the direct pseudo-$C_\ell$ power spectra (all calculations are analytical). Correlation functions are computed over a wide range, $\theta \in [0.01, 1200]$ arcminutes, to lower the complexity of the coupling matrix. Both correlation functions and pseudo-$C_\ell$ give consistent results to well within 1%. The right panel of the figure also shows the comparison of the signal to noise ratio $(S/N)$ as a function of $\ell_{\text{max}}$. To obtain $S/N$ of reconstructed $C_\ell$, we inverted the analytical gaussian covariances of pseudo-$C_\ell$ (Efstathiou 2004) and correlation functions (see appendix A of Singh et al. 2017) using the relation in eq. (72). The $S/N$ is defined as
\[
S/N = \sqrt{C_\ell \text{Cov}^{-1} C_\ell}.
\]  
(73)
Recently there has been some discussion about the apparent discrepancies or low correlation in the cosmological analysis from correlation functions and power spectra (e.g. Doux et al. 2021; Hamana et al. 2020). These discrepancies are known to be caused primarily by effects of the scale cuts imposed on the estimators, in addition to some small effects caused by approximations made in covariances and the estimators (see section 4.1). From eq. (D9) and appendix E1, it can be shown that when full range of scales is used, the correlation function and power spectra have same information, i.e. up to the normalization factor of window in correlation functions (the $RR$ term in the denominator), the correlation function and pseudo-$C_\ell$ are the same. Because of the impact of scale cuts, it is sometimes claimed that correlation functions and power spectra provide complementary information. This is true in the technical sense as the coupling matrix becomes complex with scale cuts in $\theta$ (see fig. 4), leading to mixing of information from larger $\ell$ range and the information from different $\ell$ modes is complementary (assuming they are independent). However, such ‘complementarity’ is not desirable if we do not have a good model for the subset of $\ell$ modes (which is usually the motivation for scale cuts). Mixing of information from such scales only complicates the interpretation of the full posteriors from the analysis, even if the analysis is shown to be ‘unbiased’ under some tests. To further understand the potential implications, consider the fact that the size of scatter shown in fig.
17 of Hamana et al. (2020) and fig. 9 of Doux et al. (2021) is comparable to the statistical uncertainties on parameters and to the magnitude of tensions observed between some of the weak lensing measurements and the predictions from Planck cosmology.

In general, if possible, one should keep data and model in the same space to avoid the complexities of transforming to the Fourier counterpart. Fig. 5 shows that when treated consistently, the correlation functions and power spectra carry similar information and the differences are primarily due to the impact of limited \( \ell \) range and the window correction term in the correlation function when the scale cuts are imposed, which we need to implement carefully.

We have addressed the implications of scale cuts imposed on the model in the Fourier space in section 4.2.2. Before that, we now address the issue of reconstructing lensing E/B modes from the cosmic shear correlation functions.

### 4.2.1 E/B mode reconstruction

For cosmic shear measurements, we typically measure two sets of correlation functions, \( \xi_+ \) and \( \xi_- \), which can be written in terms of the underlying EE and BB power spectra as

\[
\xi_{\pm}(\theta_b) = H_{\pm,\theta_b,\ell_b}(C_{\ell}^{EE, \pm} + C_{\ell}^{BB, \pm}).
\]

where \( H_{\pm,\theta_b,\ell_b} \) are Hankel transform operators as defined in eq (67) with spin-2 wigner-d matrices, \( 2d_{\ell,\pm,2} \).

Using eq. (70) we can also do a clean E/B mode separation for the case of cosmic shear, by converting \( \xi_\pm \) to pseudo-\( \xi \) as

\[
D_{\ell_b}^\pm = H_{\pm,\theta_b,\ell_b}^{-1} \xi_{\pm}(\theta_b).
\]

Here we assume that the \( \xi_\pm \) are measured over the same \( \theta \) range. Now we can obtain E/B pseudo-power spectra as

\[
\begin{align*}
D_{\ell_b}^{EE, \pm} &= \frac{1}{2}(D_{\ell_b}^+ + D_{\ell_b}^-) \\
D_{\ell_b}^{BB, \pm} &= \frac{1}{2}(D_{\ell_b}^+ - D_{\ell_b}^-).
\end{align*}
\]

\( D_{\ell_b}^{EE, BB, \pm} \) and the coupling matrices are defined in section 3.1. From \( D_{\ell_b}^{EE, BB, \pm} \), we can reconstruct the underlying E/B power spectra as

\[
\hat{C}_{\ell_b}^{EE} = (M_{\ell_b}^{EE})^{-1} \hat{D}_{\ell_b}^{EE} - (M_{\ell_b}^{EE})^{-1} M_{\ell_b}^{BB} C_{\ell_b}^{BB}
\]

\[
\hat{C}_{\ell_b}^{BB} = (M_{\ell_b}^{BB})^{-1} \hat{D}_{\ell_b}^{BB} - (M_{\ell_b}^{BB})^{-1} M_{\ell_b}^{EE} C_{\ell_b}^{EE}.
\]

M are binned coupling matrices, with the superscript, e.g. +EE, referring to the power spectra used in defining the \( B^{-1} \) operator. In eq. (78) and (79), we have also made a distinction between the quantities measured from data, denoted with \( \hat{\cdot} \), and the ones estimated from theory. We are using the model \( C_{\ell_b}^{EE, BB} \) to subtract out the leakage contribution of the form, \( M^{-1} C_{\ell_b} \). This is likely to be sufficient for most applications as \( (M^{+})^{-1} M^{-1} \) is typically small (\( \lesssim 10^{-2} \)) for broad windows. When noise is sub-dominant, it may be desirable to replace the \( C_{\ell} \) from model with the ones from data in order to cancel sample variance, e.g. in the case where we are attempting to detect and reconstruct small \( B \) mode power spectra. Under such a scenario, we can use an iterative method where the initial estimates of \( C_{\ell_b}^{EE, BB} \) from...
Many of the cosmological models are written and validated in the Fourier space and thus have a well defined cuts in the Fourier space based on their scales of validity. This implies truncation in the summation in eq. (54). Similar to the issues we addressed in the case of power spectra, the truncation in the Fourier counterpart should result in a convolution on the correlation functions of the form (see appendix E for derivation),
\[ \xi_{\ell-\text{cut}}(\theta) = H_{\theta,\theta} C_{\ell} b_{\ell} = b(\theta, \theta') \otimes \xi(\theta'), \]
where \( b_{\ell} \) is the truncation function in the Fourier space and in the second equation we wrote the computed correlation function as a convolution between true underlying correlation function and the Hankel transform of \( b_{\ell} \), given by \( \xi_{\ell-\text{cut}}(\theta) = B b(\theta, \theta') B_0^{-1} \).

Therefore in the case of a theory cut off defined in Fourier space, we can simply convolve the measured correlation function with the \( b(\theta, \theta') \) to impose the ‘scale cuts’ before running the inference chain. Using our binning trick from previous sections, we can define the binned version of the coupling matrix,
\[ b(\theta_{\text{b}}, \theta_{\text{b}}') = B b(\theta, \theta') B_0^{-1}. \]

Also note that \( b_{\ell} \) is operationally same as an isotropic pixel/beam smoothing applied on over density maps with the beam function given by \( \sqrt{b_{\ell}} \). Thus this method can be used to account for such smoothing effects as well. We can also account for correct factors of \( b_{\ell} \) in the covariance and cross covariance calculations by considering each tracer to be smoothed by its own ‘beam’ given by \( \sqrt{b_{\ell}} \), which can be different for different tracers (also remember that shot noise in covariance will no longer scale as \( 1/N_{\text{pairs}} \) after smoothing).

4.2.2 Model cuts in Fourier space

While \( b_{\ell} \) is typically chosen to be top hat function, it can lead to convolution over rather large scales in \( \ell \). A choice of with more compact \( b(\theta, \theta') \) can be given by a function that drops more smoothly from one to zero, such as
\[ b_{\ell,\text{cos}} = \begin{cases} \frac{1}{2} \cos \left( \frac{\ell - \ell_{\text{cut,min}}}{\ell_{\text{cut,max}} - \ell_{\text{cut,min}}} \right), & \ell_{\text{cut,min}} < \ell < \ell_{\text{cut,max}} \\ 0, & \ell \geq \ell_{\text{cut,max}} \end{cases} \]

where \( b_{\ell,\text{cos}} \) use a cosine function to smoothly truncate the power spectra to zero between \( \ell_{\text{cut,max}} - \ell_{\text{cut,min}} \). A larger separation between \( \ell_{\text{cut,max}} \) and \( \ell_{\text{cut,min}} \) will lead to narrower convolution in \( \theta \) space.

More frequently, the cutoff in theoretical models are defined in the comoving \( (k) \) space, in which case the \( b_{\ell} \) can be written as
\[ b_{\ell, k,\text{cut}} = \frac{C_{\ell, k,\text{cut}}}{C_{\ell}}. \]

where \( C_{\ell, k,\text{cut}} \) is computed with cut off in \( k \) and \( C_{\ell} \) is true underlying power spectra without any cutoff. This requires us to have some estimate of \( C_{\ell} \) and in practice it will likely be better to use some combination of \( b_{\ell, k,\text{cut}} \) and \( b_{\ell,\text{cos}} \), i.e. \( b_{\ell,\text{cos}} \times b_{\ell, k,\text{cut}} \), for more accurate results.
In such a case, it is easiest to work with the reconstructed power spectra via iMASTER, as in eq. (70) or in eq. (35) (with $M$ from eq. (41)) if one is working with the pseudo-$C_l$ measurements. For E/B power spectra, the procedure in section 4.2.1 should be followed.

It can also be computationally efficient to combine the correlation function and pseudo-$C_l$ estimators for the reconstruction of power spectra on all scales. The pair counting correlation functions estimators are fast to implement at small scales (total number of pairs is small) while the pseudo-$C_l$ estimators are faster at large scales (Fourier transforms can be computed on coarser grids/maps). Reconstructing $C_l$ from both estimators and then combining them with minimum variance weighting can result in faster reconstruction of $C_l$ over a broad range of modes.

4.3 Window errors

As discussed in section 4.1, the correlation function measured using the LS estimator is given by

$$\xi(\theta) \approx \frac{B_0 \xi_W(\theta) \xi(\theta)}{B_0 \xi_W(\theta)}.$$  \hspace{1cm} (88)

where $\xi_W(\theta)$ is the correlation function of the window. Note that the binning operation is applied separately to the numerator and the denominator. If $\xi_W(\theta)$ has a strong gradient within the bin, it can introduce a bias, namely that the effective scale of the measurement $\theta_b$ (eq. (63)) will be different from the bin center computed without accounting for the window effect.

Since the Landy-Szalay estimator uses the estimated window (randoms follow the estimated window function), which maybe different from the true underlying window, correlation functions also suffer from the window bias. The numerator of eq. (88) has the correlation function of the true window ($W$) while the denominator has the correlation function of the estimated window($\tilde{W}$). In presence of these biases we get

$$\xi(\theta_b) = \frac{B_0 \xi_W(\theta_b) \xi(\theta)}{B_0(1 + m_b(\theta))\xi_W(\theta)} + a(\theta_b),$$  \hspace{1cm} (89)

where $a(\theta_b)$ is the correlation function of the additive bias and we introduced $m_b$ to refer to the multiplicative bias in the correlation function of the window. As in the power spectra, $a(\theta_b)$ is easily accounted for by using the correlations method (e.g. Ross et al. 2012), but $m(\theta)$ is usually harder to estimate and can introduce significant biases, especially if it has a strong scale dependence. Since correlation functions and pseudo-$C_l$ are same, we can use the same methods as described in section 3.4 to understand these window biases and add corrections terms to the data/model or to the covariance.

5 CONCLUSION

In this paper we reviewed the formalism for the measurement and modeling of two point functions of the cosmological tracers, starting from the process of making maps from the catalogs to the measurements of the power spectra and the correlation functions and addressed several of the issues related to the importance of survey window, biases in its
modeling as well as the issues relevant to imposing scale cuts.

In section 2, we derived the expressions for the over density field and the expressions for shot noise contributions to the two different estimators. We showed that the window effectively acts as the inverse noise weight and hence window weighted estimator has in general lower noise compared to the estimator in which the window is removed. In the noise dominated regime, this estimator is nearly equivalent to the suboptimal estimator does not in any way help with the problem of window modeling (even in the absence of biases, one still need to model the mask which is not much easier). In section 3.4 we studied the impact of these window biases on the estimation of the pseudo-$C_l$ power spectra. Window biases lead to both additive and multiplicative biases in the power spectra. Additive biases have received some attention in the literature (e.g. Ross et al. 2012; Leistedt et al. 2013). Multiplicative biases while higher order, can be equally concerning for measurements with O(1%) precision requirements. We also saw that these biases can be more important at smaller scales (high $\ell$) which contribute a large fraction of the cosmological information.

It is also worth stressing that these issues are equally applicable to tracers such as galaxy shear. In the case of shear, we normally think of multiplicative and additive biases on the ensemble bases. However, if these biases vary with the photometric conditions of the survey, then they can also be thought of as part of the window and can be modeled using the methods in this paper and elsewhere in the literature.

In section 3 we studied the estimator for the measurement of the power spectra, namely the pseudo-$C_l$ estimator and the algorithms used in the modeling. We discussed the pesky issues involved in estimating the coupling matrices which can get quite complex for a non trivial window or for a power spectra that falls slowly or is nearly flat. For such cases the window needs to be estimated accurately out to very high $\ell$, consistent with our discussion on systematics. In terms of algorithms, we saw that the standard MASTER algorithm can be biased due to incorrect assumptions about the scaling of the power spectra and our improved iMASTER algorithm corrects for those biases by using the correct power spectra from the model. This algorithm is more powerful than the existing algorithms (e.g. Alonso et al. 2019) as it allows direct comparison with unbinned theoretical models computed at effective bin centers, $\ell_b$ (i.e. binning is performed analytically) and does not require any corrections to be applied. Furthermore, the algorithm is also more optimal as it allows for cleaner extraction of information for a given set of scales by undoing the effects of mode mixing in the pseudo-$C_l$ estimator. We list the steps involved in iMASTER computation in the algorithm 1.

In section 4, we generalized the iMASTER algorithm to the correlation functions. Using this algorithm it is possible to reconstruct the power spectra from the correlation functions, including the E/B mode separation, though it can be hard (coupling matrix is complex) if the range of scales at which correlation function is measured is limited. Similar to the case of pseudo-$C_l$ mentioned earlier, power spectra modes reconstructed from correlation functions via iMASTER can be directly compared with unbinned theoretical models computed at effective bin centers (i.e. binning+window corrections are performed within the algorithm) and do not require any corrections to be applied. Steps involved in reconstruction of power spectra from correlation functions are listed in the algorithm 2. We also developed the proper method to convolve the correlation func-

Algorithm 1 iMASTER for power spectra

**Input:** Measured power spectra $D_l$; Window power spectra $W_l$; Model power spectra $C_l$; Binning operators $B_D$, $B_C$.

1. Subtract the noise from $D_l$, i.e. $\hat{D}_l = D_l - N_l$.
2. Bin $\hat{D}_l$ to obtain $\bar{D}_l = B_D \hat{D}_l$.
3. Bin $C_l$, to obtain $\bar{C}_l = B_C \bar{C}_l$.
4. Obtain $\bar{C}_l$, eq. (40).
5. Obtain $B_C^{-1}$, eq. (39).
6. Compute the coupling matrices, $M_{t_1, t_2}$, eq. (27).
7. Bin the coupling matrices to obtain $\bar{M}_{t_1, t_2}$, eq. (41).
8. Compute the pseudo-inverse of $\bar{M}_{t_1, t_2}$ and obtain $\bar{C}_l = \bar{M}_{t_1, t_2} \bar{D}_l$.

For E/B mode separation, use eq. (78) and (79) as discussed in section 4.2.1. Use iterative method if needed.

**Output:** Binned power spectra $\bar{C}_l$ and the effective $\ell$ values for the bin $\ell_b$.

Algorithm 2 iMASTER for correlation functions

**Input:** Measured correlation functions $\xi(\theta_i)$; Window correlation functions, $\xi_W(\theta)$ and $\xi_W(\theta)$; Model power spectra $C_l$; Binning operators $B_D$, $B_C$ ($B_D$ if required); $w(\theta)$ defining the scale cuts.

**Output:** Binned power spectra $\bar{C}_l$.

1. Compute the Hankel transform operator and its inverse.
2. Compute the binned theory $\bar{\xi}(\theta_i)$ using eq. (59) and obtain $B_C^{-1}$.
3. Bin the inverse Hankel transform operator, eq. (67).
4. Obtain pseudo-$C_l$ power spectra, eq. (69).
5. Obtain $W_l$ using inverse Hankel transform of $w(\theta)$.
6. Go to step 3. of the algorithm 1: iMASTER for power spectra.

We also discussed the impact of the window estimation on the various estimators. In terms of the over density field, both estimators we discussed in section 2 are affected similarly by the biases in window estimation and using the sub-optimal estimator does not in any way help with the problem of window modeling (even in the absence of biases, one still need to model the mask which is not much easier).
tions in order to account for the scale cuts on the model in the Fourier space. This method prevents the leakage of information from the scales in Fourier space that are not properly modeled.

The iMASTER algorithm is also useful in speeding up the computations (during sampling) as after the initial setup, the computational complexity is reduced to $O(N_{\text{bin}})$ instead of $O(\ell_{\text{max}}^3)$. That being said, the performance of these algorithms becomes even more important in the memory management during a large analysis, such as an LSST-like $3 \times 2$ analysis, where the amount of peak memory requirement for the analysis decreases from $O(N_{\text{corr}} \times \ell_{\text{max}}')$ to $O(N_{\text{corr}} \times N_{\text{bin}}')$, where $N_{\text{corr}} \sim O(100)$ is the number of correlation pairs. For these large analysis, it will become imperative to use such an algorithm for computationally fast and efficient sampling of the large parameter spaces.

### AVAILABILITY OF DATA & CODE

No new data were generated or analyzed in support of this research.

The code used for the computations in this paper is available at [https://github.com/sukhdeep2/Skylens_public/tree/imaster_paper/](https://github.com/sukhdeep2/Skylens_public/tree/imaster_paper/).

### ACKNOWLEDGEMENTS

I would like to thank Yin Li for many stimulating discussions that have helped me in this work and for providing detailed feedback on this manuscript. I also thank James Sullivan, Tanveer Karim and Mehdi Rezaie for helpful discussions and feedback on this work. I also thank the anonymous referee for several helpful comments and questions that helped in improving the presentation and explanations in the paper.

I am supported by the McWilliams fellowship at the Carnegie Mellon university. Part of this work was done at University of California, Berkeley, where I was supported by a postdoctoral fellowship at the Berkeley center for cosmological physics.

### REFERENCES

Abazajian K. N., et al., 2016, arXiv e-prints, p. arXiv:1610.02743
Abbott T. M. C., et al., 2020, Phys.Rev.D, 102, 023509
Ade P., et al., 2019, J. Cosmology Astropart. Phys., 2019, 056
Alam S., et al., 2016, preprint, (arXiv:1607.03155)
Alonso D., Sanchez J., Slosar A., LSST Dark Energy Science Collaboration 2019, MNRAS, 484, 4127
Baddour N., 2014, SpringerPlus, 3, 246
Baldauf T., Smith R. E., Seljak U., Mandelbaum R., 2010, Phys.Rev.D, 81, 063531
Banados E., et al., 2020, arXiv e-prints, p. arXiv:2008.11284
Battye R. A., Kehagias A., 2003, Phys. Rev. D, 67, 083003
Bridle S. L., Crittenden R., Melchiorri A., Hobson M. P., Kneissl R., Lasenby A. N., 2002, MNRAS, 335, 1193
DES Collaboration et al., 2017, preprint, (arXiv:1703.00973)
Di Valentino E., et al., 2020, arXiv e-prints, p. arXiv:2008.11284
Doré O., et al., 2014, arXiv e-prints, p. arXiv:1412.4872
Dore O., et al., 2019, BAAS, 51, 341
Douc C., et al., 2021, MNRAS, 502, A129
Elmsfield F., Leistedt B., Peiris H. V., 2017, MNRAS, 465, 1847
Elvin-Poole J., et al., 2018, Phys.Rev.D, 98, 042006
Everett S., et al., 2020, arXiv e-prints, p. arXiv:2012.12825
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hamana T., et al., 2020, PASJ, 72, 16
Hamilton A. J. S., 1997, MNRAS, 289, 285
Heymans C., et al., 2021, A&A, 646, A140
Hivon E., Górski K. M., Netterfield C. B., Challinor P. B., Prunet S., Hansen F., 2002, ApJ, 567, 2
Joachimi B., et al., 2021, A&A, 646, A129
Li Y., Singh S., Yu B., Feng Y., Seljak U., 2019, J. Cosmology Astropart. Phys., 2019, 016
Leistedt B., Peiris H. V., Mortlock D. J., Benoit-Lévêque A., Pontzen A., 2013, MNRAS, 435, 1857
Levi M., et al., 2019, in Bulletin of the American Astronomical Society, p. 57 (arXiv:1907.10698)
Liang J. U., Leauthaud A., Singh S., Guo H., Zhou R., Smith T. L., Cyr-Racine F.-Y., 2021, MNRAS, 502, 2074
Leistedt B., Peiris H. V., Mortlock D. J., Benoit-Lévêque A., Pontzen A., 2013, MNRAS, 435, 1857
Levi M., et al., 2019, in Bulletin of the American Astronomical Society, p. 57 (arXiv:1907.10698)
Li Y., Singh S., Yu B., Feng Y., Seljak U., 2019, J. Cosmology Astropart. Phys., 2019, 016
Leistedt B., Peiris H. V., Mortlock D. J., Benoit-Lévêque A., Pontzen A., 2013, MNRAS, 435, 1857
Levi M., et al., 2019, in Bulletin of the American Astronomical Society, p. 57 (arXiv:1907.10698)
Ng K.-W., Liu G.-C., 1999, International Journal of Modern Physics D, 8, 61
Planck Collaboration et al., 2020, A&A, 641, A8
Rezaie M., Seo H.-J., Ross A. J., Bunescu R. C., 2020, MNRAS, 495, 1613
Ross A. J., et al., 2012, MNRAS, 424, 564
Ross A. J., et al., 2020, MNRAS, 498, 2354
Schaan E., White M., 2021, arXiv e-prints, p. arXiv:2103.01971
Schneider P., Eifler T., Krause E., 2010, A&A, 520, A116
Singh S., Mandelbaum R., Seljak U., Slosar A., Vazquez Gonzalez J., 2017, MNRAS, 471, 3827
Singh S., Mandelbaum R., Seljak U., Rodríguez-Torres S., Slosar A., 2020, MNRAS, 491, 51
Slosar A., Seljak U., Makarov A., 2004, Phys. Rev. D, 69, 123503
Wandelt B. D., Hivon E., Górski K. M., 2001, Phys.Rev.D, 64, 083003
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys.Rep., 530, 87
APPENDIX A: COMPARISON OF WINDOW WEIGHTING ON NOISE

In this appendix we prove the claim from section 2.1 that \( \frac{1}{W(x)} \geq \bar{W}(x) \).

We begin by noting that by \( W(x) \in [0, \infty) \) and \( \bar{W}(x) = 1 \). For the case \( W(x) \in [0, \infty) \), proof is trivial as \( \frac{1}{W(x)} \rightarrow \infty \). Hence we will focus on the case \( W(x) \in (0, \infty) \) (i.e. 0 is excluded).

We can write \( W(x) \) in terms of a mean zero variable as

\[
W(x) = 1 + w(x),
\]

where \( \bar{W}(x) = 1 + w(x) = 1, \quad \bar{w}(x) = 0. \) (A1)

The mean of \( 1/W(x) \) is then

\[
\frac{1}{\bar{W}_g(x)} = \left[ \frac{1}{1 + w(x)} \right].
\]

Defining \( y = w(x) \) and rewriting we get

\[
\frac{1}{\bar{W}_g(x)} = \int_{-1}^{\infty} \frac{1}{1 + y} P(y) dy = \int_{-1}^{1} \frac{1}{1 + y} P(y) dy + \int_{1}^{\infty} \frac{1}{1 + y} P(y) dy.
\]

\( P(y) \) is the probability distribution and in the second step we split the integral into two ranges, \( y \in (-1, 1) \) and \( y \in [1, \infty) \). Using the Taylor series in the first integral, we get

\[
\frac{1}{\bar{W}_g(x)} = \int_{-1}^{1} \frac{1 + \sum_{i=1}^{\infty} (-1)^i y^i}{1 + y} P(y) dy + \int_{1}^{\infty} \frac{1}{1 + y} P(y) dy
\]

(A5)

\[
= \int_{-1}^{1} dy P(y)(1 - y) + \int_{-1}^{1} dy P(y)(y^2 - y^3) + \int_{1}^{\infty} \frac{1}{1 + y} P(y) dy.
\]

(A6)

Now we use the fact that \( \langle y \rangle = 0 \) and \( \int_{-1}^{1} P(y) = 1 \), to change the limits of the first integral from \( y \in (-1, 1) \) to \( y \in [1, \infty) \)

\[
\frac{1}{\bar{W}_g(x)} = 1 + \int_{-1}^{\infty} dy P(y)(-1 + y) + \int_{-1}^{1} dy P(y)(y^2 - y^3) + \int_{1}^{\infty} \frac{1}{1 + y} P(y) dy
\]

(A8)

\[
= 1 + \int_{-1}^{1} dy P(y)(y^2 - y^3) + \int_{1}^{\infty} \frac{y^2}{1 + y} P(y) dy \geq 1.
\]

(A9)

Notice that the quantities inside square brackets, \([ \cdot ]\), are always positive and the last integral is also non negative, thus proving that \( \frac{1}{\bar{W}_g(x)} \geq 1 \)

APPENDIX B: MORE GENERAL WEIGHTING

In this appendix we derive the window and the noise effects in estimator of eq. (3) when more general weighting schemes are used.

Typically in a LSS survey, the galaxies are assigned weights which may depend on some intrinsic property of galaxy, we call such weights \( w_{0,i} \), and another weight dependent on the variance, \( w_{\sigma,i} \), such that the total weight is given by

\[
w_i = w_{0,i} w_{\sigma,i}
\]

(B1)

The inverse variance weight is usually written as

\[
w_{\sigma,i} = \frac{N}{\text{Var}_i} = \frac{N}{C + \sigma_{m,i}^2},
\]

(B2)

where \( \sigma_{m,i}^2 \) is contributed by the measurement noise and \( C \) is the sampling noise in galaxy field. \( C = 1 \) for galaxies and \( C = \sigma_{v,i}^2 \), i.e. shape noise, for shear. \( N \) is the normalization of the weights. The observed effective number of galaxies in a pixel are then

\[
n_g(x) = \sum_i w_i = \langle n_g(x) \rangle \bar{w}_i(x) (1 + \delta(x)).
\]

(B3)

\( \bar{w}_i(x) \) is mean of weights within the pixel. The windowed overdensity field is

\[
\delta_{g,W}(x) = \frac{n_g(x)}{n_g w_i} \frac{\langle n_g(x) \rangle}{\bar{n}_g \bar{w}_i},
\]

(B4)
where $\bar{w}_i$ is the sample mean of all the weights and is usually normalized to be 1. The window in this case is, \begin{equation}
abla W_g(x) = \frac{n_s(x)}{n_g \bar{w}_i} \omega_i(x).
\end{equation}

The variance is given as \begin{equation}
\delta_N^2(x) = \frac{1}{n_g^2 \bar{w}_i^2} \sum_i w_i^2 \text{Var}_{s_i} = \frac{N}{n_g^2 \bar{w}_i^2} \sum_i w_0^2, w_{v,i}.
\end{equation}

Assuming that $w_{0,i}$ and $w_{v,i}$ are uncorrelated, we can write the sum as \begin{equation}
\sum_i w_0^2, w_{v,i} = \langle n_s(x) \rangle w_{v,i}(x) w_0^2, i(n_g(x)),
\end{equation}
to obtain \begin{equation}
\langle \delta_N^2(x) \rangle = N \frac{\langle n_s(x) \rangle w_{v,i}(x) w_0^2, i(n_g(x))}{n_g \bar{w}_i \omega_0, i(x)}.
\end{equation}

Averaging over the survey we get \begin{equation}
\langle \delta_N^2 \rangle = N \frac{\nabla W_g}{n_g \bar{w}_i \omega_0, i \omega_0, i}.
\end{equation}

Sometimes shot noise is described in terms of effective number density of galaxies, where \begin{equation}
\frac{1}{n_g^{eff}} = \frac{1}{n_g \bar{w}_i \omega_0, i \omega_0, i}.
\end{equation}

Notice that in the absence of $w_{0,i}$ eq. (B9) is equivalent to eq. (15) (in eq. (15) $N = 1$).

In general the additional weights do change the window as well as the dependence of noise in the window. Since we only worked with shot noise in this paper and subtracted out the correct noise from pseudo-$C_l$ measurements, this does not affect the results in the main part of this paper. However, these results highlight the dependence of noise on the window and weighting and such dependencies will need to be carefully modeled both for noise modeling and covariance matrix calculations.

APPENDIX C: WINDOW COUPLING MATRIX

In this appendix we derive the response of the pseudo-$C_l$ power spectra to window power spectra.

From Hivon et al. (2002), the pseudo-$C_l$ power spectra is given as \begin{equation}
D_l = \sum_{\ell' \ell''} C_{\ell'} \left( \frac{2\ell' + 1}{4\pi} \right) \sum_{\ell''} W_{\ell''}(2\ell'' + 1) \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ s_1 & -s_1 & 0 \\ s_2 & -s_2 & 0 \end{array} \right) \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ s_1 & -s_1 & 0 \end{array} \right),
\end{equation}
\begin{equation}
D_l = M_{\ell \ell'} C_{\ell'},
\end{equation}
where in second equation we used the definition of the coupling matrix from eq. (27).

In eq. (C3), we can switch the order of summation over $\ell'$ and $\ell''$ to write \begin{equation}
D_l = \sum_{\ell'} W_{\ell'} \left( \frac{2\ell' + 1}{4\pi} \right) \sum_{\ell''} C_{\ell'} (2\ell' + 1) \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ s_1 & -s_1 & 0 \end{array} \right) \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ s_2 & -s_2 & 0 \end{array} \right),
\end{equation}
\begin{equation}
D_l = M_{\ell \ell'}^W W_{\ell'},
\end{equation}
where \begin{equation}
M_{\ell \ell'}^W = \left( \frac{2\ell' + 1}{4\pi} \right) \sum_{\ell''} C_{\ell'} (2\ell' + 1) \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ s_1 & -s_1 & 0 \end{array} \right) \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ s_2 & -s_2 & 0 \end{array} \right).
\end{equation}

We can also use the symmetries of wigner-3j symbols to change the ordering of $\ell'$, $\ell''$ if desired.

From eq. (C3) and (C4), we notice that the the pseudo-$C_l$ power spectra is symmetric in its response to the window and the underlying density field we wish to study. This highlights the importance of modeling window properly and we used the eq. (C4) to study the impact the multiplicative biases in window (biased $W_{\ell''}$) have on $D_l$.

APPENDIX D: CORRELATION FUNCTIONS

Here we derive the expressions for correlation functions in the presence of a survey window. We closely follow the expressions in Ng & Liu (1999).
We begin by writing the two point correlation function in presence of windows as
\[
\langle \delta \delta \rangle (\theta) = \frac{1}{\xi_W(\theta)} \int d^2 \delta \delta W_1(\theta) W_2(\theta),
\]
\[
= \frac{1}{\xi_W(\theta)} \int d^2 \theta \sum_{l_1,m_1} \delta_{l_1,m_1} \delta_{l_2,m_2} W_1(l_1, m_1) W_2(l_2, m_2) Y_{l_1,m_1}(\theta) Y_{l_2,m_2}(\theta + \theta),
\]
(1)

where in the second equation we simply wrote the \( \delta_i \) and \( W_i \) in terms of their spherical harmonic transforms (equivalent of Fourier Transform on a sphere).

Noting that \( \delta_{l_1,m_1} \delta_{l_2,m_2} = C_{l_1} \delta_D(l_1, l_2) \delta_D(m_1, m_2) \), we get

\[
\langle \delta \delta \rangle (\theta) = \frac{1}{\xi_W(\theta)} \int d^2 \theta \sum_{l_1,m_1,l_2,m_2} C_{l_1} \delta_D(l_1, l_2) \delta_D(m_1, m_2) W_1(l_1, m_1) W_2(l_2, m_2) Y_{l_1,m_1}(\theta) Y_{l_2,m_2}(\theta + \theta),
\]
(2)

Where in the second step I carried out the sums over \( \delta_D \). Now we use the spherical harmonics identity (Ng & Liu 1999),

\[
\sum_{l_1} Y_{l_1,m_1}(\theta) Y_{l_1,m_1}(\theta + \theta) = \sqrt{\frac{2l_1+1}{4\pi}} (-1)^{l_1-l_2} Y_{l_1, m_1}(\theta)
\]
to obtain

\[
\langle \delta \delta \rangle (\theta) = \frac{1}{\xi_W(\theta)} \int d^2 \theta \sum_{l_1,m_1,l_2,m_2} C_{l_1} \delta_D(l_1, l_2) \delta_D(m_1, m_2) W_1(l_1, m_1) W_2(l_2, m_2) Y_{l_1,m_1}(\theta) Y_{l_2,m_2}(\theta + \theta),
\]
(3)

where I omitted the \( (-1)^{l_1-l_2} \) factor since \( l_1 \) and \( l_2 \) are even numbers for the tracers we use. Rearranging the terms, we get

\[
\langle \delta \delta \rangle (\theta) = \frac{1}{\xi_W(\theta)} \sum_{l_1} \frac{2l_1+1}{4\pi} Y_{l_1,1}(\theta) C_{l_1} \int d^2 \theta \sum_{l_3,m_3} W_1(l_1, m_3) W_2(l_3, m_3) Y_{l_1,m_1}(\theta) Y_{l_3,m_3}(\theta + \theta),
\]
(4)

The term inside the integral is simply the correlation function of the window, denoted by \( \xi_W(\theta) \).

Using

\[
-1 Y_{l_1,1}(\theta) = \sqrt{\frac{2l_1+1}{4\pi}} Y_{l_1,1}(\theta) e^{-i\phi},
\]
(5)

and carrying out the angular integrals (see discussion in Ng & Liu 1999), we obtain

\[
\langle \delta \delta \rangle (\theta) = \frac{\xi_W(\theta)}{\xi_W(\theta)} \sum_{l_1} \frac{2l_1+1}{4\pi} Y_{l_1,1}(\theta) C_{l_1}.
\]
(6)

We used the full sky averaging expressions from eq. 7.3 of Ng & Liu (1999), to simplify the window correlation function, i.e.

\[
\xi_W(\theta) = \sum_{l_1} \frac{2l_1+1}{4\pi} W_{l_1,0}(\theta) = \sum_{l_1} \frac{2l_1+1}{4\pi} |d_0(\cos \theta)| W_{l_1} = \xi_W(\theta).
\]
(7)

\( W_{l_1} \) is the pseudo-\( C_l \) power spectra of the window.

While the window effect appears to cancel in eq. (6), as discussed in section 4.1, the binning operator act separately on numerator and denominator, hence the window effects do not fully cancel. Therefore the correlation functions are still sensitive to the window and here we have shown that the window effects separate out from the density field such that the estimator is only sensitive to the correlation function of the window.

APPENDIX E: SCALE CUTS

**E1** \( D_{\ell} \) reconstruction from \( \xi \)

Here we show that the inverse Hankel transform of correlation functions measured over limited range of scales leads to pseudo-\( C_{\ell} \) power spectra. This proof also shows the equivalence of correlation function and pseudo-\( C_{\ell} \) power spectra when window correlation function \( \xi_W(\theta) \) is used in place of \( w(\theta) \) in the equations below.

For scale cuts on correlation functions, we have

\[
D_{\ell} = 2 \pi \int d\theta \sin(\theta) \xi(\theta) w(\theta) \frac{4\pi}{2\ell+1} \xi_W(\theta).
\]

(E1)

**APPENDIX E: SCALE CUTS**

**E1** \( D_{\ell} \) reconstruction from \( \xi \)

Here we show that the inverse Hankel transform of correlation functions measured over limited range of scales leads to pseudo-\( C_{\ell} \) power spectra. This proof also shows the equivalence of correlation function and pseudo-\( C_{\ell} \) power spectra when window correlation function \( \xi_W(\theta) \) is used in place of \( w(\theta) \) in the equations below.

For scale cuts on correlation functions, we have

\[
D_{\ell} = 2 \pi \int d\theta \sin(\theta) \xi(\theta) w(\theta) \frac{4\pi}{2\ell+1} \xi_W(\theta).
\]

(E1)
This appendix describes the simulations and the assumptions about data samples used in the examples presented in the paper.

APPENDIX F: SIMULATIONS

Hence showing that eq. (E6) is equivalent to eq. (E5).

\[ D_\ell = \sum_{l_1, l_2} C_{l_1} w_{l_2} \sqrt{\frac{2 \ell_1 + 1}{4\pi}} \sqrt{\frac{2 \ell_2 + 1}{4\pi}} \int d\theta \sin(\theta - z) Y_{\ell_1, s_1} Y_{\ell_2, s_2}(\theta) \],

(E2)

\[ D_\ell = \sum_{l_1, l_2} C_{l_1} w_{l_2} \sqrt{\frac{4\pi}{2 \ell + 1}} \frac{(2 \ell + 1)(2 \ell + 1)}{4\pi} \frac{2 \ell + 1}{4\pi} \left( \ell \ell \ell \ell \ell \ell - \ell \ell \ell \ell \ell \ell \right) \],

(E3)

\[ D_\ell = \sum_{l_1, l_2} C_{l_1} w_{l_2} \frac{(2 \ell + 1)(2 \ell + 1)}{4\pi} \left( \ell \ell \ell \ell \ell \ell \ell - \ell \ell \ell \ell \ell \ell \ell \right) \],

(E4)

where in the second equation we wrote the integral over three spherical harmonics in terms of the wigner-3j function and the last equation is very similar to eq. (C3), with \( s_{1w} = s_{2w} = 0 \). As stated earlier, we can replace \( w(\ell) \) with the correlation function of the window \( \xi_w(\ell) \) in which case \( w(\ell_2) \) gets replaced with the window power spectra \( W(\ell_2) \) and we obtain the expressions identical to eq. (C3), thus showing that the power spectra and correlation function estimators are identical when full range of scales in \( \ell \), \( \theta \) is used.

E2 Modeling \( \xi \) with cuts on \( C_\ell \)

Here we derive the expressions for convolution operator acting on the correlation functions to account for the model cuts in the Fourier space.

We wish to derive the method to correct \( \xi \) for model cuts in the \( \ell \) space, i.e.

\[ \xi_{\text{cut}}(\theta) = \sum_{\ell} b_\ell C_\ell \sqrt{\frac{2 \ell + 1}{4\pi}} Y_{\ell, s_2}(\theta). \]  

(E5)

\( b_\ell \) is the function that applies the cuts in \( \ell \).

Following Baddour (2014), we make an ansatz that

\[ \xi_{\text{cut}}(\theta) = \int_0^{2\pi} d\phi_1 \int d\theta_1 \sin(\theta_1) \xi(\theta_1) b(\theta, \theta_1), \]  

(E6)

where

\[ b(\theta, \theta_1) = \sum_{\ell} b_\ell \sqrt{\frac{2 \ell + 1}{4\pi}} Y_{\ell, s_2}(\theta_1) Y_{\ell, s_2}(\theta_1). \]  

(E7)

Plugging \( b(\theta, \theta_1) \) back we get

\[ \xi_{\text{cut}}(\theta) = \int d\phi_1 \int d\theta_1 \sin(\theta_1) \sum_{\ell} b_\ell \sqrt{\frac{2 \ell + 1}{4\pi}} Y_{\ell, s_2}(\theta) Y_{\ell, s_2}(\theta_1) \sum_{\ell_1} C_{\ell_1} \sqrt{\frac{2 \ell_1 + 1}{4\pi}} Y_{\ell_1, s_2}(\theta_1), \]  

(E8)

\[ = \sum_{\ell} b_\ell C_\ell \sqrt{\frac{2 \ell + 1}{4\pi}} Y_{\ell, s_2}(\theta). \]  

(E9)

Hence showing that eq. (E6) is equivalent to eq. (E5).

APPENDIX F: SIMULATIONS

This appendix describes the simulations and the assumptions about data samples used in the examples presented in the paper.

For shear, we used a LSST shear sample properties, with \( n_g = 26 \) arcminutes\(^{-2} \), \( f_{sky} = 0.3 \), shape noise \( \sigma_g = 0.26 \) per component. All the galaxies are assumed to be in a narrow redshift bin at \( z = 1 \). For the galaxy sample, we used \( n_g = 10 \) arcminutes\(^{-2} \), \( f_{sky} = 0.3 \), \( b_g = 1 \). Magnification and intrinsic alignments are set to be zero.

Fig. F1 shows the two different window functions used in the examples presented in main part of the paper.
Figure F1. Window functions used for the simulations as well as other calculations using windows in the main part of the paper. W₁ is obtained using same Cₗ as the galaxy sample (Cₗ,gg) and represents a realistic window for the case galaxy shear. W₂ represents a more complex window with an additional gaussian power spectra (Cₗ,gg + Cₗ,gaussian), where Cₗ,gaussian peaks at ℓ ≈ 200 and has the width of σ ≈ 50. Both windows have constants added such that the minimum value is 0. c) The power spectra of two windows.