Galaxy–galaxy lensing estimators and their covariance properties

Sukhdeep Singh,[1]⋆ Rachel Mandelbaum,[1]⋆ Uroš Seljak,[2,3]⋆ Anže Slosar[4]
and Jose Vazquez Gonzalez[4]

[1]*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[2]*Department of Physics, University of California at Berkeley, Berkeley, CA 94705, USA*
[3]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[4]*Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA*

ABSTRACT

We study the covariance properties of real space correlation function estimators – primarily galaxy–shear correlations, or galaxy–galaxy lensing – using SDSS data for both shear catalogues and lenses (specifically the BOSS LOWZ sample). Using mock catalogues of lenses and sources, we disentangle the various contributions to the covariance matrix and compare them with a simple analytical model. We show that not subtracting the lensing measurement around random points from the measurement around the lens sample is equivalent to performing the measurement using the lens density field instead of the lens overdensity field. While the measurement using the lens density field is unbiased (in the absence of systematics), its error is significantly larger due to an additional term in the covariance. Therefore, this subtraction should be performed regardless of its beneficial effects on systematics. Comparing the error estimates from data and mocks for estimators that involve the overdensity, we find that the errors are dominated by the shape noise and lens clustering, which empirically estimated covariances (jackknife and standard deviation across mocks) that are consistent with theoretical estimates, and that both the connected parts of the four-point function and the supersample covariance can be neglected for the current levels of noise. While the trade-off between different terms in the covariance depends on the survey configuration (area, source number density), the diagnostics that we use in this work should be useful for future works to test their empirically determined covariances.

Key words: gravitational lensing: weak – galaxies: evolution – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

Galaxy–galaxy lensing, the measurement of the tangential shape distortion (shear) due to gravitational lensing by galaxies, has emerged as an important cosmological probe to study the dark matter distribution around galaxies and the growth of large-scale structure (Bartelmann & Schneider 2001; Weinberg et al. 2013). Since lensing is sensitive to all matter, galaxy–galaxy lensing provides a unique way to map the matter distribution around galaxies (or galaxy clusters) and has been measured to good precision by many surveys (e.g. Fischer et al. 2000; Hoekstra, Yee & Gladders 2004; Sheldon et al. 2004; Heymans et al. 2006; Mandelbaum et al. 2006; Leauthaud et al. 2012; van Uitert et al. 2012; Velander et al. 2014; Hudson et al. 2015; Viola et al. 2015; Buddendiek et al. 2016; Clampitt et al. 2017). Several studies have used galaxy–galaxy lens-ing to study the halo mass of galaxies and understand the galaxy–halo connection (e.g. Hoekstra et al. 2004; Heymans et al. 2006; Mandelbaum et al. 2006; Leauthaud et al. 2012; Tinker et al. 2012; van Uitert et al. 2012; Gillis et al. 2013; Velander et al. 2014; Hudson et al. 2015; Sifón et al. 2015; van Uitert et al. 2016). In combination with galaxy clustering, galaxy–galaxy lensing can also be used to recover the underlying matter correlation function, which can then be used to constrain cosmology (Seljak et al. 2005; Baldauf et al. 2010; Mandelbaum et al. 2013; More et al. 2015; Buddendiek et al. 2016; Kwan et al. 2017) and to test the theory of gravity (Zhang et al. 2007; Reyes et al. 2010; Blake et al. 2016).

With the increasing precision of cosmological surveys, there has been an increasing focus on estimating the covariances of the measurements more accurately as well, since the uncertainties in covariance matrices can lead to incorrect estimation of uncertainties in cosmological parameters (e.g. Hartlap, Simon & Schneider 2007; Dodelson & Schneider 2013; Taylor, Joachimi & Kitching 2013; Mohammed, Seljak & Vlah 2017). Estimating the non-Gaussian or connected part of the covariance matrix, which has two

⋆ E-mail: sukhdeep@cmu.edu (SS); rmandelb@andrew.cmu.edu (RM); useljak@berkeley.edu (US)

contributions, is especially challenging. The first contribution is due to mode couplings between small-scale (or in-survey) modes from the non-linear evolution of structure (Scoccimarro, Zaldarriaga & Hui 1999; Cooray & Hu 2001; Hu & White 2001; Mohammed, Seljak & Vlah 2017). The second term is the supersample variance contribution from the couplings of modes within the survey to the modes corresponding to length-scales that are larger than the survey size (e.g. Hu & Kravtsov 2003; Hamilton, Rimes & Scoccimarro 2006; Takada & Hu 2013; Li, Hu & Takada 2014; Schaan, Takada & Spergel 2014; Takada & Spergel 2014). Current prescriptions for estimating covariances include using numerical simulations (e.g. Dodelson & Schneider 2013; Manera et al. 2013; Li et al. 2014), using a physically motivated halo model (e.g. Takada & Hu 2013; Schaan et al. 2014) and using analytical estimates based on perturbation theory (e.g. Mohammed et al. 2017).

While many studies have explored this issue of covariance matrices for two-point functions in general, galaxy–galaxy lensing covariances have been relatively less well studied. When addressing this question, one must also address the question of which estimator is used for the measurement. Several estimators for the galaxy–galaxy lensing signal can be found in the literature. One estimator uses the average tangential shear of background galaxies with respect to the lens galaxies. Another estimator also includes the subtraction of tangential shear around random points, which has an expectation value of zero in the absence of systematics and which can be used to remove the impact of coherent additive shear systematics. Subtraction of the lensing shear around random points is often argued to be beneficial primarily due to the way it removes these additive systematic errors (e.g. Mandelbaum et al. 2005, 2013). However, another motivation for the latter estimator can be found in the work on optimal estimators of galaxy clustering: Landy & Szalay (1993) illustrated that the estimators that are constructed using mean-zero quantities (overdensity), while having an expectation value that is the same as the simple estimator, have better covariance properties.[1] For example, in the case of galaxy surveys, random points (R) that follow the area coverage of the lenses are typically used to estimate the mean of the galaxy field (D) in the presence of complicated survey masks. They are then used to convert the galaxy field into the normalized overdensity (mean-zero) field $(D - R)/R$, the autocorrelation of which is the standard Landy–Szalay (LS) estimator for galaxy clustering with improved covariance properties. Likewise, for galaxy–galaxy lensing, estimating the mean tangential shear around lens galaxies corresponds to correlating the galaxy density field (nonzero mean) with the shear field, while subtraction of the mean tangential shear around random points results in correlating the mean-zero galaxy overdensity with the shear.

In general, most galaxy–galaxy lensing studies either compute the covariance matrices analytically assuming shape noise and measurement noise only (see e.g. Viola et al. 2015), or use the jackknife method, which has the advantage that it includes all observational effects, though it is noisier and also limits the scales that can be used in the analysis (see e.g. Blake et al. 2016; Hildebrandt et al. 2017, for comparison of theoretical and jackknife covariance). It is also not clear how well the jackknife method can capture the supersample covariance, though since galaxy–galaxy lensing is dominated by shape noise in current generation surveys, supersample covariance is expected to be subdominant. Recently Shirasaki et al. (2017, see also Blake et al. 2016) did a detailed study of the galaxy–galaxy lensing covariance matrix using realistic *N*-body and ray tracing simulations. In version 1 of their paper, they found that once the scales are of similar order as the jackknife division size, the jackknife method overestimates the errors compared to errors obtained from the standard deviation across different simulation realizations, even in the presence of shape noise. This overestimation in jackknife errors was interpreted as increased contribution from supersample covariance, since the jackknife method has effectively divided the survey into several small survey realizations, and the supersample covariance grows with the square of the mass variance within the survey volume (Takada & Hu 2013), which can scale differently from the usual inverse-volume scaling of the covariance terms. The analysis by Shirasaki et al. (2017) applies to the galaxy–galaxy lensing estimators without subtracting the measurement around the randoms lens sample, as in e.g. Leauthaud et al. (2012), Viola et al. (2015), Hudson et al. (2015), van Uitert et al. (2016) and Blake et al. (2016) [Blake et al. (2016) subtracted the measurement around random points from the signal, but this was not done for the covariance estimation]. Other galaxy–galaxy lensing studies subtract out the signal around the randoms (e.g. Sheldon et al. 2004; Mandelbaum et al. 2005, 2006; Clampitt et al. 2017; Kwan et al. 2017). In an updated version of their paper, Shirasaki et al. (2017) show that after subtracting out the measurement around randoms, the covariance decreases and the covariance from the jackknife method is consistent with the covariance obtained using different mock realizations (for scales smaller than the size of the jackknife regions).

In this work, we explore the covariance properties of these two galaxy–galaxy lensing estimators both in the presence and the absence of systematic errors. We show that there is a theoretical reason to believe that the estimator with the mean shear around random points subtracted should have more optimal covariance properties, and we explore the impact of this difference in practice for one particular survey. Aside from the issue of removing systematics, we demonstrate the correlated noise term between measurements around galaxies and randoms, which results in more optimal variance properties after subtracting the shear around random points. We also study the differences in the covariance matrices obtained from the jackknife method and standard deviations across several mock realizations, similar to Shirasaki et al. (2017). We demonstrate several methods of empirically estimating specific covariance contributions, and interpret the results of those methods in terms of which galaxy–galaxy lensing covariance terms they include.

This work is organized as follows. In Section 2, we briefly review the theoretical formalism and estimators, and in Section 3 we present the data used. Results are presented in Section 4, and we conclude in Section 5. In Appendix A, we derive the expressions for covariance when cross-correlating non-zero mean quantities and in Appendix B we present comparisons of different estimators in the case of galaxy clustering measurements.

Throughout this work, we use the Planck 2015 cosmological parameters (Planck Collaboration XIII 2016) with $\Omega_m = 0.309$, $n_s = 0.967$, $A_s = 2.142 \times 10^{-9}$, $\sigma_8 = 0.82$. Theory predictions are computed using the linear theory + halofit (Smith et al. 2003; Takahashi et al. 2012) power spectrum generated with the CAMB (Lewis & Bridle 2002) software. We use $h = 1$ when computing distances and hence our $\Delta\Sigma$ measurements are in units of $h\,M_\odot\,pc^{-2}$.

---

[1] Sometimes the reverse claim is made in the literature – i.e. that the estimator with the signal around random points subtracted has increased variance. This claim is typically made in cases where not enough random points are used, in which case there is indeed some added variance. Our argument that the covariance properties of this estimator are superior is true in the limit of infinite random points: they are nothing other than a Monte Carlo method to determine the survey volume and hence the mean density. We explore this issue in more realistic cases in this work.

2 FORMALISM AND METHODOLOGY

2.1 Galaxy lensing

Here, we briefly review the formalism of galaxy–galaxy lensing. For a general review of gravitational lensing, we refer the reader to Bartelmann & Schneider (2001), Weinberg et al. (2013) and Kilbinger (2015).

In galaxy–galaxy lensing, we measure the projected surface mass density Σ around the lens galaxies. In the case of a spherically symmetric lens, we can write the convergence and shear as

$$\kappa(r_p) = \frac{\Sigma(r_p)}{\Sigma_c} \tag{1}$$

$$\gamma_t(r_p) = \frac{\bar{\Sigma}(<r_p) - \Sigma(r_p)}{\Sigma_c}. \tag{2}$$

$\bar{\Sigma}(<r_p)$ is the mean surface mass density within the transverse separation r_p, and the critical surface density is defined as

$$\Sigma_c = \frac{c^2}{4\pi G} \frac{f_k(\chi_s)}{(1+z_l)f_k(\chi_l)f_k(\chi_s - \chi_l)}, \tag{3}$$

where $f_k(\chi)$ is the transverse comoving distance ($f_k(\chi) = \chi$ in a flat universe). $1 + z_l$ converts the c^2/G factor to comoving space.

We can write Σ in terms of the projected galaxy–matter correlation function as

$$\Sigma(r_p) = \bar{\rho}_m \int d\Pi \, \xi_{\rm gm}(r_p, \Pi) = \bar{\rho}_m w_{\rm gm}(r_p), \tag{4}$$

where Π denotes the line-of-sight separation from the halo centre, and we have ignored the effects of lensing window function, which depends on Π. Nominally the definition for Σ should include a factor of $1 + \xi_{\rm gm}$ within the integral (rather than just $\xi_{\rm gm}$), but the constant term does not contribute to γ_t because it gets removed by subtraction of the $\bar{\Sigma}(<r_p)$ term. In the linear bias regime, the galaxy–matter projected correlation function can be derived from the matter power spectrum as

$$w_{\rm gm}(r_p) = b_g r_{\rm cc} \int dz\, W(z) \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} \, P_{\delta\delta}(\boldsymbol{k}, z) e^{i(r_p \cdot \boldsymbol{k})}, \tag{5}$$

where b_g is the galaxy bias and $r_{\rm cc}$ is the galaxy–matter cross-correlation coefficient, both of which are assumed to be independent of redshift in this equation. $P_{\delta\delta}(k, z)$ is the matter power spectrum (linear+halofit) at redshift z. To lowest order, lensing measurements are not affected by redshift space distortions, and hence we do not include any corrections for them. The weight function $W(z)$ depends on the redshift distribution of the source galaxies and on the weights used in the estimators when measuring the signal (see Section 2.2). We explicitly include these weights when computing the effective redshift $z_{\rm eff}$ for the theory calculations.

2.2 Estimator

Our observable quantity for the galaxy–galaxy lensing measurement is $\Delta\Sigma$, which is estimated in bins of r_p as

$$\widehat{\Delta\Sigma}_{\rm gR}(r_p) = \frac{\sum_{\rm ls} w_{\rm ls} \gamma_t^{\rm (ls)} \Sigma_c^{\rm (ls)}}{\sum_{\rm Rs} w_{\rm Rs}} - \frac{\sum_{\rm Rs} w_{\rm Rs} \gamma_t^{\rm (Rs)} \Sigma_c^{\rm (Rs)}}{\sum_{\rm Rs} w_{\rm Rs}}. \tag{6}$$

The summation is over all lens–source (ls) pairs. γ_t is the tangential shear measured in the lens–source frame. Σ_c is the geometric factor defined in equation (3), and the optimal weight $w_{\rm ls}$ for each lens–source pair ($w_{\rm Rs}$ is defined analogously for random–source pairs) is defined as (see Mandelbaum et al. 2005)

$$w_{\rm ls} = \frac{\Sigma_c^{-2}}{\sigma_\gamma^2 + \sigma_{\rm SN}^2}. \tag{7}$$

The Σ_c^{-2} enters the inverse variance weight because we defined the $\Delta\Sigma$ in equation (6) as the maximum likelihood estimator (Sheldon et al. 2004). Note that the denominator in equation (6) has a sum over weights $w_{\rm Rs}$, measured by using random lenses rather than lens galaxies. Division by $\sum_{\rm Rs} w_{\rm Rs}$ rather than $\sum_{\rm ls} w_{\rm ls}$ corrects for the dilution of the shear signal by source galaxies that are physically associated with the lens but appear to be behind the lens due to photo-z scatter. These galaxies do not contribute any shear but are counted in the total weights (sum over $w_{\rm ls}$). The correction factor for this effect $\sum_{\rm ls} w_{\rm ls}/\sum_{\rm Rs} w_{\rm Rs}$ (properly normalized to account for different number of random and real lenses) is usually called the boost factor (Sheldon et al. 2004; Mandelbaum et al. 2005) and is $\sim$1 for the scales we use in this work $r_p \gtrsim 1\,h^{-1}$ Mpc. Finally, we subtract the shear signal measured around the random points to remove any systematics that may contribute a spurious shear signal at large scales, and to construct a more optimal estimator. Throughout this paper, the subscript 'gR' is used to indicate that the measurement around random points is subtracted from the measurement around the lenses:

$$\widehat{\Delta\Sigma}_{\rm gR}(r_p) = \widehat{\Delta\Sigma}_g(r_p) - \widehat{\Delta\Sigma}_R(r_p). \tag{8}$$

One of the main goals of this paper is to test how the subtraction of the signal measured around random points impacts the covariance matrix of the final measurement. Hence, we will study the signals measured around galaxies and randoms separately as well. We will refer to the signal measured around galaxies by $\widehat{\Delta\Sigma}_g$ and around random points by $\widehat{\Delta\Sigma}_R$. The ratio of the number of random points used to the number of lens galaxies is N_R:

$$N_R = \frac{\text{Number of random lenses}}{\text{Number of lens galaxies}}. \tag{9}$$

In case of $N_R = 0$, $\widehat{\Delta\Sigma}_{\rm gR} \equiv \widehat{\Delta\Sigma}_g$.

To estimate jackknife errors, we use 100 approximately equal area ($\sim$10 deg on a side) jackknife regions to obtain the jackknife mean and errors for each r_p bin.

2.3 Covariance: theoretical expectations

As is derived in Appendix A, the covariance for $\Delta\Sigma_g$ is given by

$$\text{Cov}(\Delta\Sigma_g)(|\boldsymbol{r}_{p,i}|, |\boldsymbol{r}_{p,j}|)$$
$$= \left[\frac{\mathcal{A}_W(\boldsymbol{r}_{p,i} - \boldsymbol{r}_{p,j})}{\mathcal{A}_W(\boldsymbol{r}_{p,i})\mathcal{A}_W(\boldsymbol{r}_{p,j})} \frac{1}{L_W} \int \frac{dk\,k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) \right.$$
$$\times \left. \left(\Sigma_c^2 \left(P_{\rm gg}(k) + N_g \right) \left(P_{\kappa\kappa}(k) + N_\gamma \right) + \Delta\Pi_2 \bar{\rho} P_{g\delta}^2 + T_{g\gamma g\gamma} \right) \right]$$
$$+ \left\{ \frac{1}{\mathcal{A}_W(\boldsymbol{r}_{p,i})\mathcal{A}_W(\boldsymbol{r}_{p,j})L_W} \int \frac{dk\,k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) \right.$$
$$\times \left. \tilde{W}(k)^2 \Sigma_c^2 \left(P_{\kappa\kappa}(k) + N_\gamma \right) \right\}. \tag{10}$$

Here, the lens galaxy power spectrum can be written as $P_{\rm gg} = b_g^2 P_{\delta\delta}(k)$ in the linear bias regime, the lens galaxy shot noise power spectrum is $N_g = \frac{1}{\bar{n}_g}$, the shape noise term is $N_\gamma = \frac{\sigma_\gamma^2}{\bar{n}_s}$, the galaxy–shear cross-power spectrum is $P_{g\gamma} = \bar{\rho} b_g r_{\rm cc} P_{\delta\delta}(k)$, and the convergence power spectrum $P_{\kappa\kappa}$ is given in equation (A34).

We compute the $\langle \Sigma_c \rangle$ when performing the measurements, and use $\langle \Sigma_c \rangle \sim 4.7 \times 10^3 \frac{h\mathrm{M_\odot}}{\mathrm{pc}^2}$, $n_s \sim 8\,h^2\,\mathrm{Mpc}^{-2}$ (after accounting for weights) in theoretical covariance calculations. J_2 is the second-order spherical Bessel function, $\sigma_\gamma \sim 0.36/2\mathcal{R}$ is the shape noise, $\Delta\Pi_2 \approx 700\,h^{-1}\,\mathrm{Mpc}$ is the line-of-sight integration length using the lensing window function, $W(k)$ is the projected lens window function in Fourier space (see Appendix A for the expressions for the window function) and L_W is the line-of-sight length of the lens window function. $T_{g\gamma g\gamma}$ is the connected part of the covariance, which we will ignore in numerical calculations. $\mathcal{A}_W$ (defined in equation A36) is the window function-dependent effective area covered by each bin, and accounts for the edge effects due to the survey window. For scales much smaller than the survey window, $\mathcal{A}_W \approx A_W$, where A_W is the survey area. In the $\Delta\Sigma$ measurements in this work, we only divide the lens sample into jackknife regions, but the source sample stays the same and hence the edge effects are small. When calculating numerical predictions for the jackknife errors, we set the window function $\mathcal{A}_W \approx A_W$ (ideally we should set $\mathcal{A}_W$ for jackknife to be same as that of full sample, but in the case of an idealized LOWZ-sized window, $\mathcal{A}_W \approx A_W$ for the scales of interest). In Appendix B, we show the effects of $\mathcal{A}_W$ on the jackknife covariance in the case of clustering measurements.

The covariance for $\Delta\Sigma_{\mathrm{gR}}$ is similar to what is shown in equation (10), except that it does not contain the last term in curly brackets, $\{\}$. This term arises because of the non-zero mean value of the lens density (here the lens sample is assumed to be normalized and hence its mean is 1). This term is independent of the lens overdensity and only depends on the window function of the lens sample. Hence, it get removed when the measurement around random points is subtracted from the measurement around galaxies.

2.4 Covariance matrix estimation methods

To estimate the covariance matrix, we use two different methods. The first is the jackknife method, in which we divide the whole survey into $N_{\mathrm{Jk}} = 100$ approximately equal-area regions ($\sim 90\,\mathrm{deg}^2 \equiv 76^2,\,125^2,\,163^2\,[\mathrm{Mpc}\,h^{-1}]^2$ at $z = 0.16,\,0.27$ and 0.36, respectively). We then make N_{Jk} measurements by dropping one region at a time, so that each measurement contains $N_{\mathrm{Jk}} - 1$ regions. The jackknife variance estimate (diagonals of the covariance matrix) is then

$$\mathrm{Var}_{\mathrm{Jk}}(\widehat{\Delta\Sigma}) = \frac{N_{\mathrm{Jk}} - 1}{N_{\mathrm{Jk}}} \sum_{i=1}^{N_{\mathrm{Jk}}} (\Delta\Sigma_i - \overline{\Delta\Sigma})^2. \tag{11}$$

Our second method is to measure $\Delta\Sigma$ using N_M mock realizations of the lens sample and then compute the standard deviation (Std) of the measurement across all realizations:

$$\mathrm{Var}_{\mathrm{Std}}(\widehat{\Delta\Sigma}) = \frac{1}{N_M - 1} \sum_{i=1}^{N_M} (\Delta\Sigma_i - \overline{\Delta\Sigma})^2. \tag{12}$$

Finally, for comparison, we also show error estimates using subsamples of the survey. We use the same subsampling as in the jackknife method, but in this case we perform the measurements using one subsample at a time. The variance in this case (error on the mean) is

$$\mathrm{Var}_{\mathrm{subsample}}(\widehat{\Delta\Sigma}) = \frac{1}{N_{\mathrm{Jk}}(N_{\mathrm{Jk}} - 1)} \sum_{i=1}^{N_{\mathrm{Jk}}} (\Delta\Sigma_i - \overline{\Delta\Sigma})^2. \tag{13}$$

3 DATA

3.1 SDSS

The SDSS (York et al. 2000) imaged roughly π steradians of the sky, and the SDSS-I and II surveys followed up approximately one million of the detected objects spectroscopically (Eisenstein et al. 2001; Richards et al. 2002; Strauss et al. 2002). The imaging was carried out by drift-scanning the sky in photometric conditions (Hogg et al. 2001; Ivezić et al. 2004), in five bands (*ugriz*; Fukugita et al. 1996; Smith et al. 2002) using a specially designed wide-field camera (Gunn et al. 1998) on the SDSS Telescope (Gunn et al. 2006). These imaging data were used to create the catalogues of shear estimates that we use in this paper. All of the data were processed by completely automated pipelines that detect and measure photometric properties of objects, and astrometrically calibrate the data (Lupton et al. 2001; Pier et al. 2003; Tucker et al. 2006). The SDSS-I/II imaging surveys were completed with a seventh data release (Abazajian et al. 2009), though this work will rely as well on an improved data reduction pipeline that was part of the eighth data release, from SDSS-III (Aihara et al. 2011); and an improved photometric calibration ('ubercalibration', Padmanabhan et al. 2008).

3.2 SDSS-III BOSS

Based on the SDSS photometric catalogue, galaxies were selected for spectroscopic observation (Dawson et al. 2013), and the BOSS spectroscopic survey was performed (Ahn et al. 2012) using the BOSS spectrographs (Smee et al. 2013). Targets were assigned to tiles of diameter $3°$ using an adaptive tiling algorithm (Blanton et al. 2003), and the data were processed by an automated spectral classification, redshift determination and parameter measurement pipeline (Bolton et al. 2012).

We use SDSS-III BOSS data release 12 (DR12; Alam et al. 2015; Reid et al. 2016) LOWZ galaxies in the redshift range $0.16 < z < 0.36$.

The LOWZ sample consists of luminous red galaxies at $z < 0.4$, selected from the SDSS DR8 imaging data and observed spectroscopically in the BOSS survey. The sample is approximately volume-limited in the redshift range $0.16 < z < 0.36$, with a number density of $\bar{n} \sim 3 \times 10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$ (Manera et al. 2015; Reid et al. 2016). We use the same sample as used by Singh, Mandelbaum & Brownstein (2017), who mask out certain regions on the sky which have higher galactic extinction or poor imaging quality (Reyes et al. 2012), which leaves 225 181 galaxies in the sample.

3.3 Re-Gaussianization shapes and photometric redshifts

The shape measurements for the source sample used in this work are described in more detail in Reyes et al. (2012). Briefly, these shapes are measured using the re-Gaussianization technique developed by Hirata & Seljak (2003). The algorithm is a modified version of ones that use 'adaptive moments' (equivalent to fitting the light intensity profile to an elliptical Gaussian), determining shapes of the point spread function (PSF)-convolved galaxy image based on adaptive moments and then correcting the resulting shapes based on adaptive moments of the PSF. The re-Gaussianization method involves additional steps to correct for non-Gaussianity of both the PSF and the galaxy surface brightness profiles (Hirata & Seljak 2003). The components of the distortion are defined as

$$(e_+, e_\times) = \frac{1 - (b/a)^2}{1 + (b/a)^2}(\cos 2\phi, \sin 2\phi), \tag{14}$$

where b/a is the minor-to-major axis ratio and ϕ is the position angle of the major axis on the sky with respect to the RA–Dec. coordinate system. The ensemble average of the distortion is related to the shear as

$$\gamma_+, \gamma_\times = \frac{\langle e_+, e_\times \rangle}{2\mathcal{R}} \tag{15}$$

$$\mathcal{R} = 1 - \frac{1}{2}\left\langle e_{+,i}^2 + e_{\times,i}^2 - 2\sigma_i^2 \right\rangle, \tag{16}$$

where σ_i is the per-component measurement uncertainty of the galaxy distortion, and $\mathcal{R} \approx 0.87$ is the shear responsivity representing the response of an ensemble of galaxies with some intrinsic distribution of distortion values to a small shear (Kaiser, Squires & Broadhurst 1995; Bernstein & Jarvis 2002). A discussion of corrections for shear-related systematic biases and the residual systematic uncertainties can be found in Mandelbaum et al. (2013). These estimates are based on a combination of null tests using the shear catalogue and external image simulations.

The photometric redshifts for the catalogue were estimated using the template-fitting code ZEBRA (Feldmann et al. 2006). Using photometric redshifts for the source sample introduces a bias in galaxy–galaxy lensing through misestimation of the Σ_c factor (with the most severe misestimation arising due to the inclusion of some lens-foreground 'source' pairs due to scatter in photometric redshifts). Nakajima et al. (2012) showed that this bias can be large, but can be estimated to within 2 per cent using a representative calibration sample with spectroscopic redshifts. We compute these correction factors using the method of Nakajima et al. (2012) with the LOWZ lens redshift distributions to be ∼10 per cent and then multiply our measurements with a calibration factor of 1.1.

3.4 Mock source catalogue

We generate 100 mock catalogues of the shape sample by randomly rotating the shapes of galaxies in the real source sample, while keeping their positions (RA, Dec., z) fixed. Random rotations remove any coherent shear (cosmological or due to systematics) in the sample while maintaining the shape noise and measurement noise in each realization. As a result, $\Delta\Sigma$ measurements using rotated (mock) sources will not have any coherent signal and their covariance matrix will only have contributions from shape noise and measurement noise. The comparison of the covariance matrix of mocks with the covariance from real sources will allow us to study the contribution of shape noise and measurement noise to the covariance in the real data.

3.5 QPM mocks

To estimate the galaxy–galaxy lensing covariance matrix using a mock lens sample, we use the QPM mocks (White, Tinker & McBride 2014) that have been used in several BOSS analyses (e.g. Cuesta et al. 2016; Gil-Marín et al. 2016; Grieb et al. 2017). QPM mocks are constructed using the quick particle mesh method (White et al. 2014) to mimic the large-scale clustering properties of BOSS galaxies. In this work, we use 100 QPM mocks with the same sky coverage, mask and jackknife splitting as in the LOWZ sample.

4 RESULTS

In this section, we present our results from measuring galaxy–galaxy lensing using different estimators and different combinations of lens and source galaxies. We perform several tests to study the

Table 1. Table showing the main sources of statistical uncertainty for different combinations of lens, source and estimators in this work. We use the notation from equation (10), with $P_{gg} = b_g P_{\delta\delta}$, N_g is the galaxy shot noise term, N_γ is the shape noise and the terms in curly brackets {} involving W^2 are the window function-dependent covariance contributions to $\Delta\Sigma_g$.

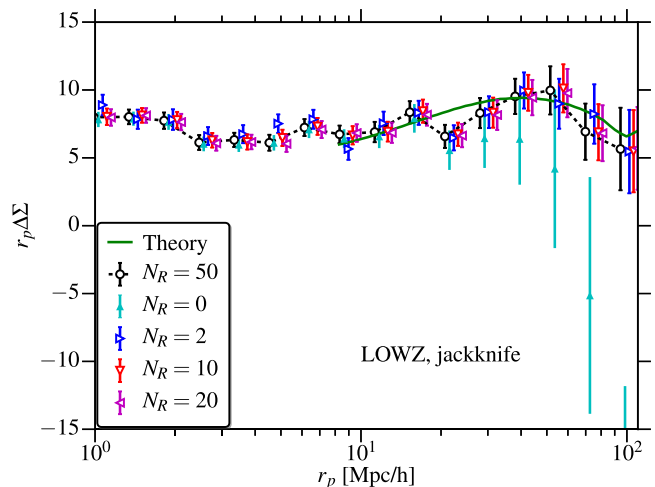| Lens-shape sample	Error term
LOWZ-SDSS	$(P_{gg} + N_g)(P_{\gamma\gamma} + N_\gamma) + P_{g\gamma}^2 + T_{g\gamma g\gamma}$ $+ \{W^2(P_{\gamma\gamma} + N_\gamma)\}$
QPM-SDSS	$(P_{gg} + N_g)(P_{\gamma\gamma} + N_\gamma) + \{W^2(P_{\gamma\gamma} + N_\gamma)\}$
LOWZ-Mock	$(P_{gg} + N_g)N_\gamma + \{W^2 N_\gamma\}$
QPM-Mock	$(P_{gg} + N_g)N_\gamma + \{W^2 N_\gamma\}$
Randoms-Mock	$N_g N_\gamma + \{W^2 N_\gamma\}$

Figure 1. $\Delta\Sigma$ measured for the LOWZ sample, with different numbers of random catalogues used. The errors shown are from the jackknife method. The signal without subtraction of the lensing signal around random points ($N_R = 0$, cyan points) shows the presence of additive systematics in the SDSS source sample. These systematics are removed with the subtraction of the signal measured around random points. Theory predictions use the linear theory+halofit power spectrum with fixed cosmology along with the best-fitting linear bias and $r_{cc} = 1$, and fitting was done for $10\,h^{-1}\,\text{Mpc} < r_p < 65\,h^{-1}\,\text{Mpc}$.

effects on the estimated covariance by using different covariance estimation methods, varying N_R, varying the clustering properties of lens sample and varying the source sample (without and with systematics). A summary of the various terms in the covariance that contribute for different combinations of lens and source samples is presented in Table 1, and a summary of the results is in Fig. 7.

4.1 LOWZ lensing results

We begin by showing the galaxy–galaxy lensing measurements using the LOWZ lens sample. Fig. 1 shows $\Delta\Sigma$ measured using LOWZ lens galaxies and different numbers of random points. When using no randoms (the $N_R = 0$ case), there is evidence for a spurious systematic signal at large scales. This spurious signal arises because the PSF correction method used to measure the galaxy shapes is unable to fully remove all of the PSF anisotropy. The SDSS survey strategy results in large-scale coherent PSF anisotropy which, when improperly removed, causes a large-scale coherent galaxy shape alignment (see Mandelbaum et al. 2005, 2013, for a detailed
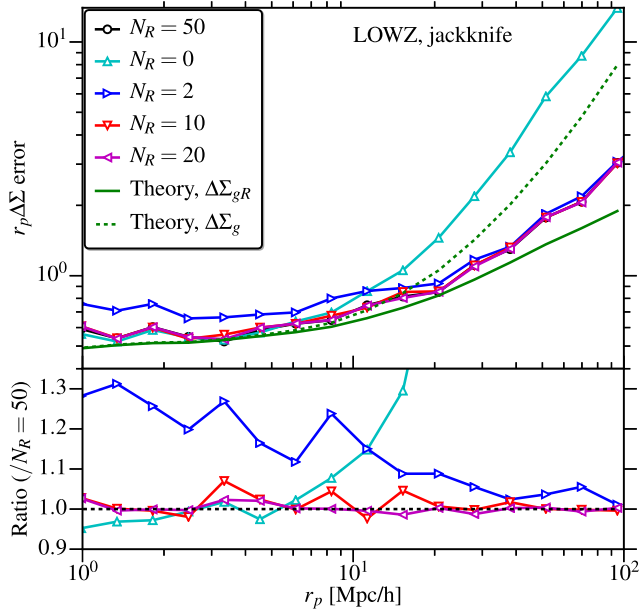
Figure 2. Jackknife errors in the LOWZ $\Delta\Sigma$ measurement (square root of diagonal elements of the covariance matrix), for different numbers of random catalogues. The errors generally follow the $\propto 1/r_p$ scaling expected from shape noise and logarithmic binning in r_p, though there is some saturation at large scales due to the correlated shape noise and systematics. Also shown are the theory predictions, which are consistent with data at small scales, though there are differences at large scales due to systematics that are not included in the theory.

discussion). The fact that this spurious signal gets removed when the measurement around randoms is subtracted has been the primary motivation for the subtraction of the signal around random points in SDSS galaxy–galaxy lensing measurements.

However, the subtraction of the signal around random points also reduces the errors in the measurements, especially at large scales (the noise in the $\Delta\Sigma_g$ term has contributions from systematics as well as shape noise, as we will show in later sections). In Fig. 2, we show the variations in the error estimates (square root of the diagonal covariance matrix elements) with different numbers of random catalogues N_R. At small scales, where the errors follow the expected scaling for shape noise ($\propto 1/r_p$ in logarithmic r_p bins), subtracting the signal around random points increases the error estimates, though with $N_R \gtrsim 10$, the errors converge to $N_R = 0$ case. The errors in this regime should scale with N_R as

$$\left(\frac{\delta\Delta\Sigma(N_R)}{\delta\Delta\Sigma(N_R = 0)} \right)^2 = 1 + \frac{1}{N_R}. \tag{17}$$

Given that the jackknife error estimates using 100 regions have uncertainty of the order of ~ 15 per cent ($\sqrt{2/99}$; Taylor et al. 2013), using $N_R = 10$ is sufficient and henceforth our results will use $N_R = 10$ unless a different value is explicitly given. However, note that when using large numbers of mocks for error estimates, more randoms might be required.

At large scales, contributions to the noise from systematics and the correlated shape noise ($P_{gg}(P_{\gamma\gamma} + N_{\gamma\gamma})$ term) start dominating and hence the error estimates diverge from the $1/r_p$ scaling. The errors are mostly consistent with the theoretical predictions calculated using equation (10). At large scales there is a contribution from the systematics that is not included in the theory predictions, hence the errors diverge from those predictions especially for the $N_R = 0$ case

where systematics are most important. We distinguish between the different terms in the variance in the following sections.

Fig. 3 shows the correlation and cross-correlation matrices for $\Delta\Sigma_g$, $\Delta\Sigma_R$ and $\Delta\Sigma_{gR}$, both from theory and data and their difference. The measurements of $\Delta\Sigma$ around galaxies ($\Delta\Sigma_g$) and randoms ($\Delta\Sigma_R$) are highly correlated for $r_p \gtrsim 10\,h^{-1}$ Mpc. When we subtract the measurement around randoms, this correlated noise gets removed and hence the noise in $\Delta\Sigma_{gR}$ decreases compared to that in $\Delta\Sigma_g$ at large scales. The bin-to-bin correlations also decrease, though there are still some residual correlations due to the clustering of the lens sample and the effects of systematics. Since the theory prediction does not include systematics, the residuals after subtracting the theory correlation matrix from the jackknife are not consistent with zero. A cleaner test of the theoretical expressions will use randomly rotated sources, which do not have any systematic shear correlations, in the next subsection.

4.2 Mock sources

In this section, we quantify the effects of additive shear systematics on the covariance estimation, especially on the differences in the errors with and without $\Delta\Sigma_R$ subtracted. We create 100 mock realizations of the source sample by randomly rotating the source galaxies. The resulting source catalogues should exhibit no coherent signals of cosmological origin or due to systematics. When measuring $\Delta\Sigma$ around the LOWZ galaxies with these randomly rotated source catalogues, we should observe realistic levels of correlated shape noise, but no systematics, cosmic variance or supersample covariance. All terms involving shear correlations – $P_{\gamma\gamma}$, $P_{g\gamma}$ and $T_{g\gamma g\gamma}$ – are zero and hence do not contribute to the covariance.

Fig. 4 shows the jackknife and standard deviation errors obtained with and without subtracting $\Delta\Sigma_R$. Subtracting $\Delta\Sigma_R$ reduces the errors, and the results are consistent with the theory predictions. However, the magnitude of the difference in errorbars for $\Delta\Sigma_{gR}$ versus $\Delta\Sigma_g$ shown here (factor of ~ 2 at the largest scale) is lower compared to what was seen with real SDSS sources (factor of ~ 5), which suggests that a bit more than half the contribution to the errorbars for $\Delta\Sigma_g$ with real sources came from shear correlations, mostly caused by the systematics rather than cosmic shear given the low redshift of this sample. In the case of $\Delta\Sigma_{gR}$, the errors computed using the standard deviation (Std) across the realizations are consistent with the jackknife errors with $\Delta\Sigma_R$ subtraction. In the case of $\Delta\Sigma_g$, the Std errors are lower than the jackknife errors because of the much larger effective window for Std (full survey window) compared to the jackknife ($1/N_{jk}$ of the survey window).

As shown in Figs 3 and 4, in the case of the jackknife window, the predictions from theory are consistent with the data at the ~ 10 per cent level for both $\Delta\Sigma_g$ and $\Delta\Sigma_{gR}$, which is within the noise in the jackknife errors. In the case of $\Delta\Sigma_g$ with the full survey window, the theory predictions for the errors are lower than the jackknife errors. This is likely because when computing the theory predictions, we assume an idealized geometry (see Appendix A), which underestimates the window function effects from the realistically complicated window in the data.

Fig. 4 and the theoretical predictions demonstrate that even in a survey with no known additive systematic errors, measurements of $\Delta\Sigma$ in the r_p range where correlated shape noise is important will have substantially better S/N when using the more optimal $\Delta\Sigma_{gR}$ estimator.
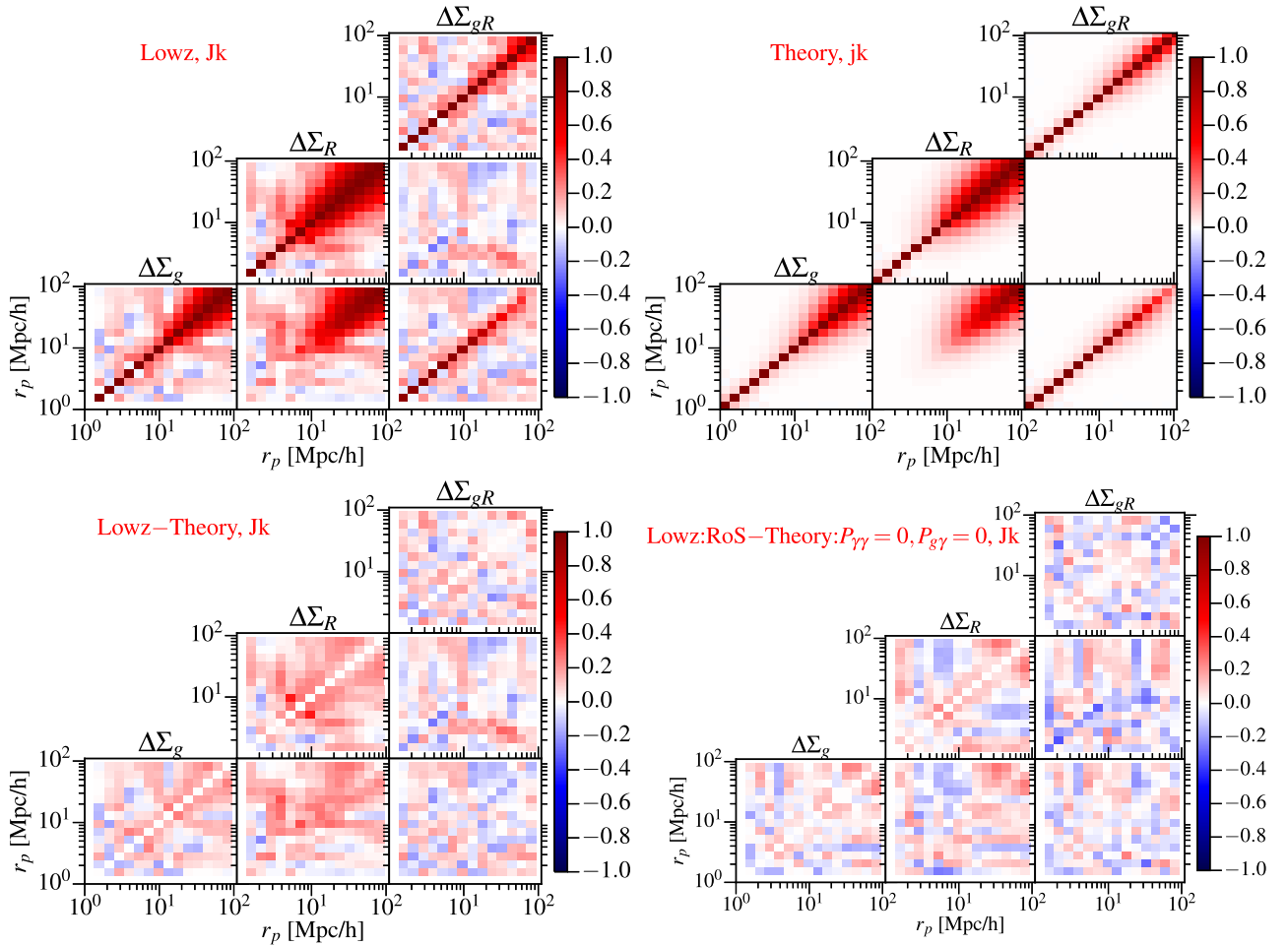
Figure 3. *Upper row*: correlation and cross-correlation matrices for $\Delta\Sigma$ measured around galaxies ($\Delta\Sigma_g$), randoms ($\Delta\Sigma_R$) and the difference of the two ($\Delta\Sigma_{gR}$), both from theory and data (jackknife). Due to the shape noise, there are strong autocorrelations and cross-correlations at large scales in both $\Delta\Sigma_g$ and $\Delta\Sigma_R$. Subtracting the measurement around the random points removes most of the correlated noise (from systematics and correlated shape noise), though there are still some residual bin-to-bin correlations in $\Delta\Sigma_{gR}$, primarily due to the clustering of lens galaxies. *Bottom row*: the difference between the correlation matrices (note we do not take the difference of the covariances here) from the jackknife and the theoretical predictions (*left*) and from the mocks and the theoretical predictions (*right*). In the case of the jackknife matrices, there are systematics that are not included in the theory predictions, so the latter are underpredicated. In the case of the mock sources (or rotated sources, RoS), shear systematics are removed and the theory predictions are consistent with the data within the noise in the jackknife covariances.

4.3 Lens mocks

In this section, we vary the lens properties to examine how the covariance depends on the lens sample properties.

4.3.1 QPM mocks

In this section, we measure $\Delta\Sigma$ around the galaxies in the QPM mocks using the real and mock source sample. In both cases, shear–galaxy correlations will be absent, i.e. $P_{g\gamma} = T_{g\gamma g\gamma} = 0$, while the former will include $P_{\gamma\gamma}$ terms and the latter will not. Even though the QPM mocks have somewhat different clustering at small scales than the real LOWZ sample (see Fig. B1), the typical separation between galaxies ($\sim$1–2 h^{-1} Mpc) is very similar between the mocks and the LOWZ sample and hence the QPM mocks are adequate to test the effects of lens clustering, P_{gg}, on the galaxy–galaxy lensing covariance. The signal around the QPM mocks should include large-scale systematics and realistic levels of shape noise.

The left-hand panel of Fig. 5 shows the error estimates in the $\Delta\Sigma$ measurements with the $\Delta\Sigma_{gR}$ estimator using $N_R = 10$, and

with the $\Delta\Sigma_g$ estimator (without the measurement around random points subtracted). The jackknife errors with the $\Delta\Sigma_R$ subtraction are consistent with the error estimates using the standard deviation across 200 QPM mocks. Also the jackknife errors for the LOWZ sample are consistent with the jackknife and standard deviation errors computed from the QPM mocks. This consistency confirms that the errors are dominated by the shape noise, $N_\gamma(P_{gg} + N_g)$ with some contributions from systematics, while contributions from cosmic variance and supersample covariance (not included in the signal with the QPM mocks) are subdominant.

As in the case of the LOWZ lens sample, the theoretical predictions for the errors in the left-hand panel of Fig. 5 are lower than the actual errors. This is due to the effect of systematics (there are no connected terms in this case). In the right-hand panel, we show the errors estimated by using the rotated (or mock) sources. In this case, there are no shear correlations (either cosmological or due to systematics) and the theoretical predictions for the errors are consistent with the measured errors in the case of $\Delta\Sigma_{gR}$. The discrepancies in the case of $\Delta\Sigma_g$ are due to the idealized window function used for the extra term in the theory calculations.
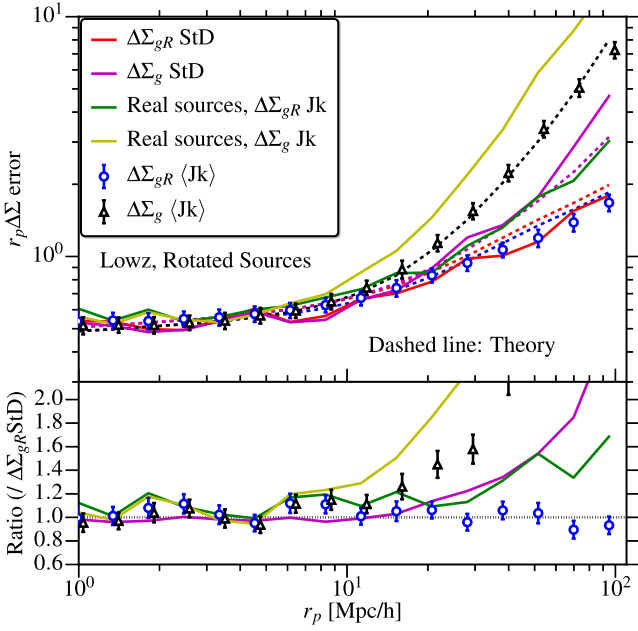
Figure 4. $\Delta\Sigma$ errors using different estimators with LOWZ galaxies as lenses and 100 mock realizations of the source sample obtained by randomly rotating the SDSS source galaxies. Open black and blue markers show the mean and standard deviation of the jackknife errors. Solid red and magenta lines show the errors from the standard deviation (StD) across different realizations of $\Delta\Sigma_g$ and $\Delta\Sigma_{gR}$. Dashed lines are the theory predictions for the curves with corresponding colours. In the bottom panel, we plot the ratio of the different errors with respect to the StD errors of $\Delta\Sigma_{gR}$. Jackknife errors for the real LOWZ sample are also plotted for comparison ($N_R = 10$ for LOWZ). Note that in the $\Delta\Sigma_g$ case, the correlated noise in these mocks is lower than when using real sources due to the removal of the contribution from systematics.
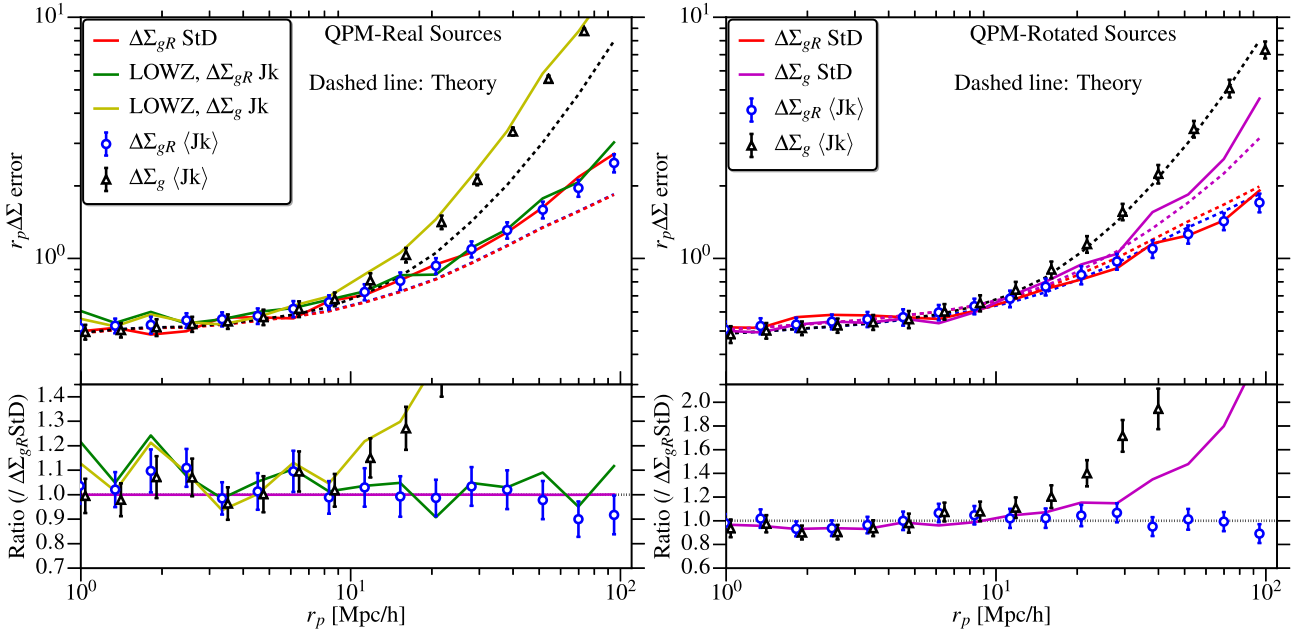
4.3.2 Randoms

In this section, we measure $\Delta\Sigma$ by replacing the LOWZ galaxies with random lens catalogues. In this case, the covariance only has contributions from terms with lens shot noise, $N_g(P_{\gamma\gamma} + N_\gamma)$, as there is no lens clustering, $P_{gg} = 0$. We use 75 random samples that are the same size as the LOWZ sample, along with 10 additional random samples, which are used to compute $\Delta\Sigma_R$. In this section we only show results using the mock source sample, so $P_{\gamma\gamma} = 0$.

Fig. 6 shows the errors in the $\Delta\Sigma$ measurements using randoms lenses, with and without $\Delta\Sigma_R$ subtracted out. Also shown are the error estimates using the standard deviation of the signal measured across all 75 independent realizations.

In the case of $\Delta\Sigma_{gR}$ errors from jackknife, the errors from the standard deviation and theory are consistent. The errors also follow the expected $1/r_p$ scaling (no lens clustering in this case), except at the largest scales where there are some deviations, possibly due to small amounts of large-scale power in the distribution of the random catalogues that enables them to match the selection function of the LOWZ sample. Also, the errors in the case of the random lenses are in general lower than those for the LOWZ sample or the QPM mocks at large scales due to the effects of lens clustering. Finally, for $\Delta\Sigma_g$ the errors do not follow the typical $1/r_p$ scaling because of the additional $W^2 N_\gamma$ term.

4.4 Putting it all together

Using the results of the previous subsections, we can now understand the contributions of various terms in the covariance, using both theoretical predictions and errors estimated using data and mocks.

In Fig. 7, we show the error estimates from various combinations of data and mocks (left-hand panel) and theory calculations using various terms in equation (10) (right-hand panel). In Table 1, we

Figure 5. Same as Fig. 4, using the QPM mocks as the lens sample with real sources (left-hand panel) and rotated (mock) sources (right-hand panel). As in Fig. 4, the errors with real sources are higher due to contributions from systematics. The theoretical predictions are consistent except at large scales due to systematics and the effects of the LOWZ window function, which are not well captured by the idealized window function assumed in the theoretical calculations. Also in the case of the real sources, the LOWZ jackknife errors are consistent with those for the QPM mocks, suggesting that errors from the connected part of the covariance are subdominant.
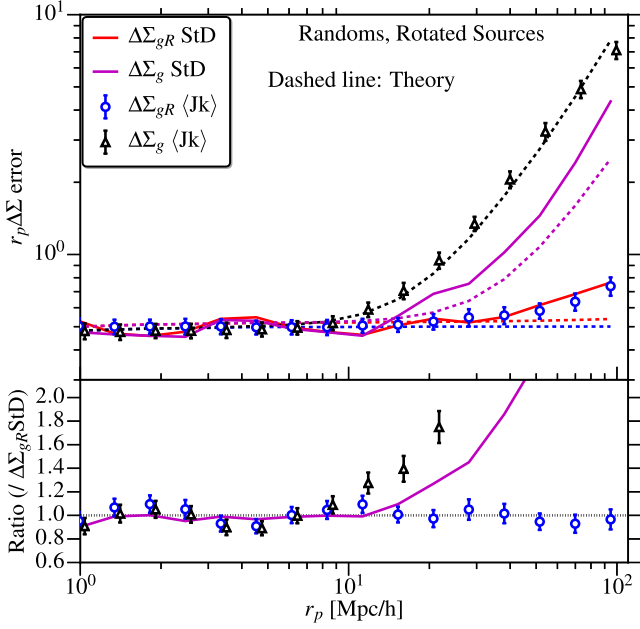
Figure 6. Same as Fig. 4, now using 75 random realizations of the LOWZ lens catalogue (no clustering) with rotated (mock) sources. Since the lens sample has no clustering and sources have no shear correlations, the errors only include shot noise terms and hence scale as $1/r_p$, except at the largest scales where the random points have some clustering as they match the selection function in LOWZ.

also show various terms that contribute to various combinations of data and mocks.

As shown in Table 1, we can use the random lenses with the rotated (mock) sources to compute the contributions of lens shot noise

and source shape noise to the covariance. Including the lens samples with clustering (LOWZ or QPM) then provides the contribution from the clustering of the lenses. Substituting the real sources with mock lenses (QPM) provides the contributions from shear correlations (systematics or cosmological). Thus, we can study all terms except for those arising from lens–source correlations ($P_{g\gamma}$ and $T_{g\gamma g\gamma}$) using the mocks we have used in this paper. Using more realistic simulations as in Shirasaki et al. (2017) will further allow a study of these lens–source correlations terms, though as shown in previous sections, contributions from these terms are subdominant when using the SDSS shape sample.

As demonstrated in Fig. 7, at small scales the errors are dominated by the shot noise terms $N_g N_\gamma$, where N_g is the galaxy shot noise power spectrum and N_γ is the shape noise power spectrum. At larger scales ($r_p \gtrsim 20\,h^{-1}$ Mpc), the term involving the lens clustering, $P_{gg}N_\gamma$, starts dominating (in the literature this term is commonly referred to as 'correlated shape noise'). The contributions from the shear power spectrum terms are in general small, with $P_{\gamma\gamma}$ and $P_{g\gamma}$ terms only contributing $\sim$10 per cent of the error even at $r_p \sim 100\,h^{-1}$ Mpc for this particular survey. In this work, we did not compute the trispectrum terms ($T_{g\gamma g\gamma}$), but based on the comparison between the LOWZ and QPM lens samples (see Fig. 5), we find no evidence that such terms are important at any scale considered in this work for SDSS.

We also show the contributions of the window function-dependent terms in $\Delta\Sigma_g$. As the size of the window function increases, $W(k)$ approaches a delta function, $\delta_D(k)$. As a result, the contribution of these terms to the covariance decreases with increasing window size, which is the reason why the jackknife errors have higher contributions than the standard deviation from the mocks using full survey window. Once these terms are removed, the errors in $\Delta\Sigma_{gR}$ are consistent from both the jackknife and the full window. Normally we do expect $\Delta\Sigma_{gR}$ to be different between the
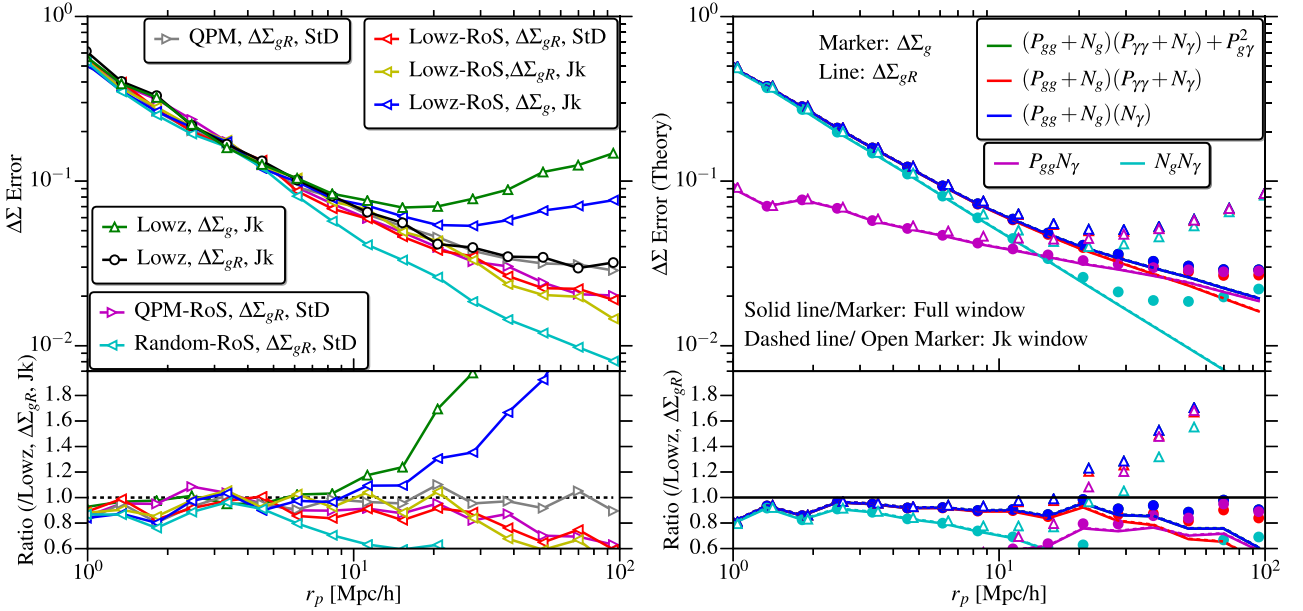

Figure 7. Figure summarizing the main error estimates discussed in this paper. Note that unlike in other plots, the y-axis here is the error on $\Delta\Sigma$ error without a factor of r_p. In the left-hand panel, we show the errors estimated using different combinations of the data and mock catalogues (RoS stands for 'rotated sources' (or mock sources); but unless explicitly mentioned, the curves use the real sources). In the right-hand panel, we show the error estimates from different terms that contribute to the theoretical covariance using the notation of equation (10). In the bottom panels, the curves are divided by the jackknife errors on the real LOWZ $\Delta\Sigma_{gR}$, which contains contributions from systematics that are not included in the theoretical expressions, hence the ratios are systematically below 1. For different curves, the power spectra terms not mentioned in the legend are set to zero, e.g. for the cyan curves ($N_g N_\gamma$) all autopower spectra are zero. There is still an $\{N_\gamma\}$ term in this case in $\Delta\Sigma_{gR}$.
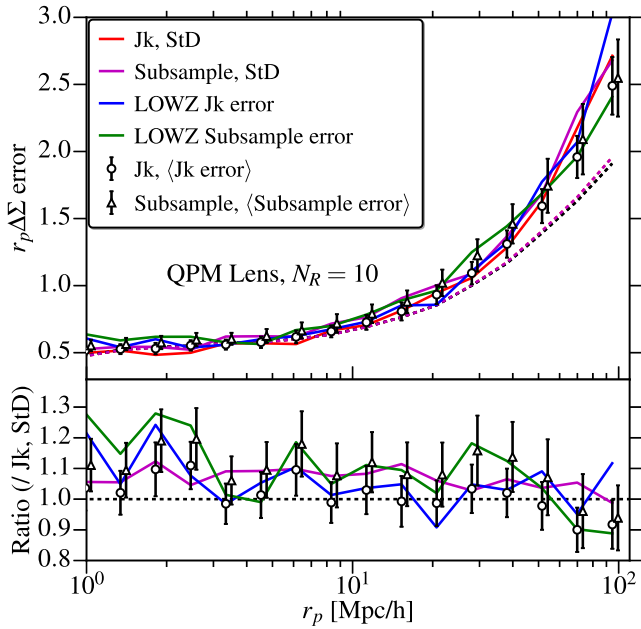
Figure 8. Comparison of the error estimates obtained using the jackknife method and the standard deviation across 100 subsamples of the lens samples (QPM except for the two lines labelled 'LOWZ'). The same subsamples were used for both methods. All calculations use $\Delta\Sigma_g - \Delta\Sigma_R$ with $N_R = 10$. 'Jk, Std' refers to the errors from the jackknife and from the standard deviation of the mock samples as defined in Section 2.4 for $\Delta\Sigma_{gR}$ (signal, not the noise), while 'Subsample, Std' refers to the errors from the standard deviations of $\Delta\Sigma_{gR}$ measured as the mean of subsamples in each realization. $\langle$Jk error$\rangle$ ($\langle$Subsample error$\rangle$) is the mean of the jackknife (subsample) error across the realizations. Dashed lines are the corresponding theory predictions.

jackknife and the full window due to the edge effects (see results for clustering in Appendix B). However, in this work we only apply the jackknife to the lens sample while using the full source sample at all times. Hence, the edge effects in the jackknife and full window cases are the same, though our theory curves underpredict these edge effects since we assumed an idealized window function with circular symmetry and no holes (see Appendix A).

4.5 Comparison of different error estimates

In this section, we compare the error estimates from the jackknife method with 100 regions against those from taking the mean and error on the mean from 100 subsamples. The primary motivation for this comparison is to test for edge effects and to check whether the jackknife method underestimates the errors once the scales are close to the size of the subsamples. The subsamples were defined in the same way for both methods and the division was only done on the lens sample. Each subsample/jackknife region is cross-correlated with the entire shape sample. Subsampling on the lens sample alone is sufficient in the shape noise-dominated regime, since the shape noise for different subsamples will be uncorrelated. In the case of other measurements, e.g. clustering, the measurement across different subsamples will get correlated once the length-scale approaches the size of the subsample, and the errors will be underestimated in both cases.

Fig. 8 shows the comparison of three different error estimates: jackknife, subsampling and standard deviation for the QPM mocks. All error estimates are consistent with each other (within the
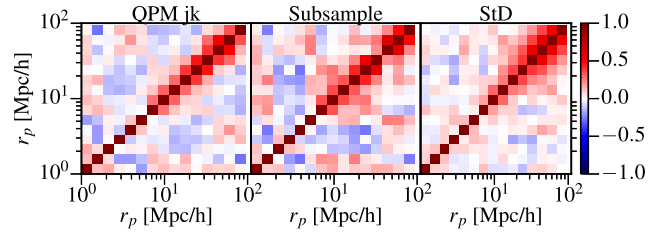

Figure 9. Comparison of the correlation matrices for the QPM mocks from the jackknife, subsampling and standard deviation methods across 200 realizations. All calculations use $\Delta\Sigma_g - \Delta\Sigma_R$ with $N_R = 10$.

uncertainties), though the scatter in the subsampling errors is somewhat higher than the jackknife errors. In Fig. 9, we also show the correlation matrix for the three different error estimates; they are all consistent with each other.

5 CONCLUSIONS

In this work, we have studied the behaviour of covariances in galaxy–galaxy lensing measurements using mock catalogues and theoretical predictions, including a comparison of two different galaxy–galaxy lensing estimators with different covariance properties. The mock catalogues include randomly distributed lenses, QPM mocks that have similar clustering as LOWZ galaxies, and mock source catalogues obtained by randomly rotating the real SDSS source galaxies. Our main results are summarized in Fig. 7 and Table 1.

For the SDSS, at small scales the covariance is dominated by the pure uncorrelated shape noise, which is white, but at larger scales, contributions from lens and shear correlations also matter. Using mock source catalogues obtained by randomly rotating the sources, we show that the errors are dominated by the terms involving the shape noise, $N_\gamma(P_{gg} + N_g)$. When using the real sources from SDSS, we also found evidence of contributions to the covariance from the systematics to the covariance, even when subtracting random signal (i.e. using lens overdensity). While our theory calculations did not include the contributions from systematics or the connected term (including the supersample variance), the consistency of the covariances when using LOWZ and QPM mocks as lenses demonstrates that contributions from the connected terms are subdominant and the differences between the theoretical predictions and the measurements in Fig. 2 arise primarily from the systematics. This conclusion in general depends on the survey configuration, and for different surveys the tradeoffs between shape noise and other covariance contributions must be re-evaluated.

We also demonstrated that the additional variance seen on large scales when not subtracting the shear around random points is only partially contributed by systematics. Even without systematics, on scales where correlated shape noise is important, using the suboptimal estimator (lens density instead of overdensity) for galaxy–galaxy lensing can reduce the per-bin S/N by a substantial factor (up to a factor of 2 on the largest scales considered, which are $\sim$10 times the correlation length of the lens galaxies). Our covariance calculations suggest that this reduction in the covariance primarily arises from the removal of shear correlations (including shape noise), which only depends on the window function of the lens sample. Covariance estimation methods using subsamples (including the jackknife) have a smaller survey window and hence have a higher contribution from this term compared to the full survey window. The tests using mock lens and mock source catalogues are

consistent with this explanation. Our results suggest that the discrepancy between the jackknife error estimates and standard deviation across different realizations observed by (Shirasaki et al. 2017, version 1) can be explained by the fact that they used the lens density instead of the lens overdensity in the galaxy–galaxy lensing measurements, given that the contributions from the additional covariance terms due to use of density depend on the survey window. This assertion has been confirmed in the updated work (version 2) of Shirasaki et al. (2017).

In our calculations of covariances, we also identified the effects of the window function, which can be important when comparing the covariance estimations from empirical methods such as jackknife or subsampling, which divide the survey window into smaller parts. Since in our jackknife estimates we only split the lens sample, these effects are not important in our lensing measurements. In Appendix B, using clustering measurements, we show that the window function effects can alter the covariance by up to 40–50 per cent on scales approaching the size of the subsamples. Finally, for both clustering (Appendix B) and lensing (Section 4.5), we also demonstrated using mocks that the jackknife errors are consistent with the errors from the subsampling methods for the scales that are smaller than the subsample size at the effective redshift of the sample.

Our results emphasize the importance of using the optimal galaxy–galaxy lensing estimator $\Delta\Sigma_g - \Delta\Sigma_R$ even in the absence of systematics, for g–g lensing estimates that extend to scales above a few Mpc, to obtain better covariance properties and to enable use of internal error estimates like the jackknife. Our conclusions are also applicable to galaxy–CMB lensing cross-correlations, as were done by Singh et al. (2017). Finally, we recommend the use of the tests of covariances demonstrated in this work and those in Shirasaki et al. (2017) for ongoing and future surveys, to better understand which terms are dominating the covariances.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Ahn C. P. et al., 2012, ApJS, 203, 21
Aihara H. et al., 2011, ApJS, 193, 29
Alam S. et al., 2015, ApJS, 219, 12
Baldauf T., Smith R. E., Seljak U., Mandelbaum R., 2010, Phys. Rev. D, 81, 063531
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Blake C. et al., 2016, MNRAS, 456, 2806
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, AJ, 125, 2276
Bolton A. S. et al., 2012, AJ, 144, 144
Buddendiek A. et al., 2016, MNRAS, 456, 3886
Clampitt J. et al., 2017, MNRAS, 465, 4204
Cooray A., Hu W., 2001, ApJ, 554, 56
Cuesta A. J. et al., 2016, MNRAS, 457, 1770
Dawson K. S. et al., 2013, AJ, 145, 10
Dodelson S., Schneider M. D., 2013, Phys. Rev. D, 88, 063537
Eisenstein D. J. et al., 2001, AJ, 122, 2267
Feldmann R. et al., 2006, MNRAS, 372, 565
Fischer P. et al., 2000, AJ, 120, 1198
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Gil-Marín H. et al., 2016, MNRAS, 460, 4188
Gillis B. R. et al., 2013, MNRAS, 431, 1439
Grieb J. N. et al., 2017, MNRAS, 467, 2085
Gunn J. E. et al., 1998, AJ, 116, 3040
Gunn J. E. et al., 2006, AJ, 131, 2332
Hamilton A. J. S., Rimes C. D., Scoccimarro R., 2006, MNRAS, 371, 1188
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Heymans C. et al., 2006, MNRAS, 371, L60
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Hirata C., Seljak U., 2003, MNRAS, 343, 459
Hoekstra H., Yee H. K. C., Gladders M. D., 2004, ApJ, 606, 67
Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129
Hu W., Kravtsov A. V., 2003, ApJ, 584, 702
Hu W., White M., 2001, ApJ, 554, 67
Hudson M. J. et al., 2015, MNRAS, 447, 298
Ivezić Ž. et al., 2004, Astron. Nachr., 325, 583
Kaiser N., 1987, MNRAS, 227, 1
Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460
Kilbinger M., 2015, Rep. Prog. Phys., 78, 086901
Kwan J. et al., 2017, MNRAS, 464, 4045
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Leauthaud A. et al., 2012, ApJ, 744, 159
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Li Y., Hu W., Takada M., 2014, Phys. Rev. D, 89, 083519
Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden F. R. Jr, Primini F. A., Payne H. E., eds, ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X. Astron. Soc. Pac., San Francisco, p. 269
Mandelbaum R. et al., 2005, MNRAS, 361, 1287
Mandelbaum R., Hirata C. M., Ishak M., Seljak U., Brinkmann J., 2006, MNRAS, 367, 611
Mandelbaum R., Slosar A., Baldauf T., Seljak U., Hirata C. M., Nakajima R., Reyes R., Smith R. E., 2013, MNRAS, 432, 1544
Manera M. et al., 2013, MNRAS, 428, 1036
Manera M. et al., 2015, MNRAS, 447, 437
Mohammed I., Seljak U., Vlah Z., 2017, MNRAS, 466, 780
More S., Miyatake H., Mandelbaum R., Takada M., Spergel D. N., Brownstein J. R., Schneider D. P., 2015, ApJ, 806, 2

Nakajima R., Mandelbaum R., Seljak U., Cohn J. D., Reyes R., Cool R., 2012, MNRAS, 420, 3240

Padmanabhan N. et al., 2008, ApJ, 674, 1217

Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, AJ, 125, 1559

Planck Collaboration XIII, 2016, A&A, 594, A13

Reid B. et al., 2016, MNRAS, 455, 1553

Reyes R., Mandelbaum R., Seljak U., Baldauf T., Gunn J. E., Lombriser L., Smith R. E., 2010, Nature, 464, 256

Reyes R., Mandelbaum R., Gunn J. E., Nakajima R., Seljak U., Hirata C. M., 2012, MNRAS, 425, 2610

Richards G. T. et al., 2002, AJ, 123, 2945

Schaan E., Takada M., Spergel D. N., 2014, Phys. Rev. D, 90, 123523

Scoccimarro R., Zaldarriaga M., Hui L., 1999, ApJ, 527, 1

Seljak U. et al., 2005, Phys. Rev. D, 71, 043511

Sheldon E. S. et al., 2004, AJ, 127, 2544

Shirasaki M., Takada M., Miyatake H., Takahashi R., Hamana T., Nishimichi T., Murata R., 2017, MNRAS, 470, 3476

Sifón C. et al., 2015, MNRAS, 454, 3938

Singh S., Mandelbaum R., Brownstein J. R., 2017, MNRAS, 464, 2120

Smee S. A. et al., 2013, AJ, 146, 32

Smith J. A. et al., 2002, AJ, 123, 2121

Smith R. E. et al., 2003, MNRAS, 341, 1311

Strauss M. A. et al., 2002, AJ, 124, 1810

Takada M., Hu W., 2013, Phys. Rev. D, 87, 123504

Takada M., Spergel D. N., 2014, MNRAS, 441, 2456

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152

Taylor A., Joachimi B., Kitching T., 2013, MNRAS, 432, 1928

Tinker J. L., George M. R., Leauthaud A., Bundy K., Finoguenov A., Massey R., Rhodes J., Wechsler R. H., 2012, ApJ, 755, L5

Tucker D. L. et al., 2006, Astron. Nachr., 327, 821

van Uitert E., Hoekstra H., Schrabback T., Gilbank D. G., Gladders M. D., Yee H. K. C., 2012, A&A, 545, A71

van Uitert E. et al., 2016, MNRAS, 459, 3251

Velander M. et al., 2014, MNRAS, 437, 2111

Viola M. et al., 2015, MNRAS, 452, 3529

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87

White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594

York D. G. et al., 2000, AJ, 120, 1579

Zhang P., Liguori M., Bean R., Dodelson S., 2007, Phys. Rev. Lett., 99, 141302

APPENDIX A: COVARIANCE

A1 General case

Here, we derive the expression for the covariance of the cross-correlation function of two fields with non-zero mean. The results depend on the estimator used for that cross-correlation function, as we will show explicitly below (see also Landy & Szalay 1993), and directly motivate the use of estimators that involve subtraction of the mean density for both fields. While we will use the example of clustering in this section, the results are in general true for any tracer of large-scale structure. In Section A3, we will use the results from this section to compute the covariance for the galaxy–shear cross-correlation function.

We are interested in the cross-correlation function of two (biased) tracer fields, g_X, g_Y, of the matter density field ($\rho_{\rm m}$, not just $\delta\rho_{\rm m}$)

$$g_i = (1 + \delta_i + n_i)M_i, \tag{A1}$$

where $i = X$ or Y, M_i is the mean value of the field (mean number density in the case of galaxies) and n_i is the noise in the tracer field (shot noise in the case of galaxies, shape noise in the case of shear). Hereafter, in this section we will assume that the field is normalized so that $M_i = 1$. For notational compactness, we also define

$$\widehat{\delta}_i = \delta_i + n_i \tag{A2}$$

$$g_i = 1 + \widehat{\delta}_i. \tag{A3}$$

In Fourier space

$$\tilde{g}_i(\boldsymbol{k}) = \delta_D(\boldsymbol{k}) + \tilde{\delta}_i(\boldsymbol{k}) + \tilde{n}_i(\boldsymbol{k}) = \delta_D(\boldsymbol{k}) + \tilde{\widehat{\delta}}_i(\boldsymbol{k}), \tag{A4}$$

where $\delta_{\rm D}$ is the Dirac delta function.

We can write the cross-correlation function of two fields as (analogous to normalized $\frac{DD}{RR} - 1$)

$$\widehat{\xi}_{XY}(\boldsymbol{r}) = \widehat{\xi}_{g_1 g_2}(\boldsymbol{r}) = \frac{1}{\mathcal{V}_W(\boldsymbol{r})} \int d^3\boldsymbol{r}' W(\boldsymbol{r}' + \boldsymbol{r})W(\boldsymbol{r}') \left[g_1(\boldsymbol{r}')g_2(\boldsymbol{r}' + \boldsymbol{r}) - 1 \right] \tag{A5}$$

$$= \frac{1}{\mathcal{V}_W(\boldsymbol{r})} \int d^3\boldsymbol{r}' W(\boldsymbol{r}' + \boldsymbol{r})W(\boldsymbol{r}') \left[\widehat{\delta}_1(\boldsymbol{r}')\widehat{\delta}_2(\boldsymbol{r}' + \boldsymbol{r}) + \widehat{\delta}_1(\boldsymbol{r}') + \widehat{\delta}_2(\boldsymbol{r}' + \boldsymbol{r}) + 1 \right] - 1 \tag{A6}$$

$$= \frac{1}{\mathcal{V}_W(\boldsymbol{r})} \int d^3\boldsymbol{r}' W(\boldsymbol{r}' + \boldsymbol{r})W(\boldsymbol{r}')\widehat{\delta}_1(\boldsymbol{r}')\widehat{\delta}_2(\boldsymbol{r}' + \boldsymbol{r}) \tag{A7}$$

$$= \frac{1}{\mathcal{V}_W(\boldsymbol{r})} \int d^3\boldsymbol{r}' W(\boldsymbol{r}' + \boldsymbol{r})W(\boldsymbol{r}') \left[\delta_1(\boldsymbol{r}')\delta_2(\boldsymbol{r}' + \boldsymbol{r}) + n_1(\boldsymbol{r}')n_2(\boldsymbol{r}' + \boldsymbol{r}) \right] \tag{A8}$$

$$\widehat{\xi}_{XY}(\boldsymbol{r}) = \widehat{\xi}_{g_1 g_2}(\boldsymbol{r}) = \xi_{XY}(\boldsymbol{r}) + \xi_{n_X n_Y}(\boldsymbol{r}), \tag{A9}$$

where g_1 belongs to field X and g_2 to field Y. $W(\boldsymbol{r})$ is the survey window function. We have assumed that the noise and δ_i have zero mean and are also uncorrelated with each other on all scales. The normalization factor is the integral over window functions

$$\mathcal{V}_W(\boldsymbol{r}) = \int d^3\boldsymbol{r}' W(\boldsymbol{r}' + \boldsymbol{r})W(\boldsymbol{r}') = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \tilde{W}(\boldsymbol{k})\tilde{W}(-\boldsymbol{k}). \tag{A10}$$

The covariance of the correlation function is given as

$$\mathrm{Cov}(\widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i)\widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j)) = \left\langle \widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i)\widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j) \right\rangle - \left\langle \widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i) \right\rangle \left\langle \widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j) \right\rangle , \tag{A11}$$

where g_1, g_3 belong to field X and g_2, g_4 belong to Y. Using equation (A5), we obtain

$$\mathrm{Cov}(\widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i)\widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j)) = \left\langle \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \mathrm{d}^3\boldsymbol{r} \int \mathrm{d}^3\boldsymbol{r}' W(\boldsymbol{r})W(\boldsymbol{r}')W(\boldsymbol{r}+\boldsymbol{r}_i)W(\boldsymbol{r}'+\boldsymbol{r}_j) \left[g_1(\boldsymbol{r})g_2(\boldsymbol{r}+\boldsymbol{r}_i)g_3(\boldsymbol{r}')g_4(\boldsymbol{r}'+\boldsymbol{r}_j) \right.\right.$$
$$\left.\left. - g_1(\boldsymbol{r})g_2(\boldsymbol{r}+\boldsymbol{r}_i) - g_3(\boldsymbol{r}')g_4(\boldsymbol{r}'+\boldsymbol{r}_j) + 1 \right] \right\rangle - \left\langle \widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i) \right\rangle \left\langle \widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j) \right\rangle \tag{A12}$$

$$\mathrm{Cov}(\widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i)\widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j)) = \left\langle \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \mathrm{d}^3\boldsymbol{r} \int \mathrm{d}^3\boldsymbol{r}' W(\boldsymbol{r})W(\boldsymbol{r}')W(\boldsymbol{r}+\boldsymbol{r}_i)W(\boldsymbol{r}'+\boldsymbol{r}_j)g_1(\boldsymbol{r})g_2(\boldsymbol{r}+\boldsymbol{r}_i)g_3(\boldsymbol{r}')g_4(\boldsymbol{r}'+\boldsymbol{r}_j) \right\rangle$$
$$- \frac{\left\langle \widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i) \right\rangle}{\mathcal{V}_W(\boldsymbol{r}_i)} - \frac{\left\langle \widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j) \right\rangle}{\mathcal{V}_W(\boldsymbol{r}_j)} - 1 - \left\langle \widehat{\xi}_{g_1g_2}(\boldsymbol{r}_i) \right\rangle \left\langle \widehat{\xi}_{g_3g_4}(\boldsymbol{r}_j) \right\rangle . \tag{A13}$$

We use $\langle g_1g_2g_3g_4 \rangle_{ij}$ as short-hand for the first term in equation (A13), which we would like to simplify.

$$\langle g_1g_2g_3g_4 \rangle_{ij} = \left\langle \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \mathrm{d}^3\boldsymbol{r} \int \mathrm{d}^3\boldsymbol{r}' g_1(\boldsymbol{r})g_2(\boldsymbol{r}+\boldsymbol{r}_i)g_3(\boldsymbol{r}')g_4(\boldsymbol{r}'+\boldsymbol{r}_j)W(\boldsymbol{r})W(\boldsymbol{r}')W(\boldsymbol{r}+\boldsymbol{r}_i)W(\boldsymbol{r}'+\boldsymbol{r}_j) \right\rangle . \tag{A14}$$

Writing the g_i in terms of its Fourier space counterpart $\tilde{g}_i$, we get

$$\langle g_1g_2g_3g_4 \rangle_{ij} = \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \mathrm{d}^3\boldsymbol{r} \int \mathrm{d}^3\boldsymbol{r}' \iiiint \prod_{n=1}^{4} \left[\frac{\mathrm{d}^3\boldsymbol{k}_n}{(2\pi)^3} \right] \iiiint \prod_{m=1}^{4} \left[\frac{\mathrm{d}^3\boldsymbol{q}_m}{(2\pi)^3}\tilde{W}(\boldsymbol{q}_m) \right]$$
$$\times \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_1-\boldsymbol{q}_1)\cdot\boldsymbol{r}}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_2-\boldsymbol{q}_2)\cdot(\boldsymbol{r}+\boldsymbol{r}_i)}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_3-\boldsymbol{q}_3)\cdot\boldsymbol{r}'}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_4-\boldsymbol{q}_4)\cdot(\boldsymbol{r}'+\boldsymbol{r}_j)} \left\langle \tilde{g}_1(\boldsymbol{k}_1)\tilde{g}_2(\boldsymbol{k}_2)\tilde{g}_3(\boldsymbol{k}_3)\tilde{g}_4(\boldsymbol{k}_4) \right\rangle \tag{A15}$$

$$\langle g_1g_2g_3g_4 \rangle_{ij} = \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \iint \frac{\mathrm{d}^3\boldsymbol{k}_1}{(2\pi)^3}\frac{\mathrm{d}^3\boldsymbol{k}_3}{(2\pi)^3} \iiiint \prod_{m=1}^{4} \left[\frac{\mathrm{d}^3\boldsymbol{q}_m}{(2\pi)^3}\tilde{W}(\boldsymbol{q}_m) \right] \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_1-\boldsymbol{q}_1)\cdot\boldsymbol{r}_i}\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_3-\boldsymbol{q}_3)\cdot\boldsymbol{r}_j}$$
$$\times \left\langle \tilde{g}_1(\boldsymbol{k}_1)\tilde{g}_2(-\boldsymbol{k}_1+\boldsymbol{q}_1+\boldsymbol{q}_2)\tilde{g}_3(\boldsymbol{k}_3)\tilde{g}_4(-\boldsymbol{k}_3+\boldsymbol{q}_3+\boldsymbol{q}_4) \right\rangle . \tag{A16}$$

We have integrated over $\mathrm{d}^3\boldsymbol{r}$ and $\mathrm{d}^3\boldsymbol{r}'$ and then over $\mathrm{d}^3\boldsymbol{k}_2$ and $\mathrm{d}^3\boldsymbol{k}_4$, to obtain the last expression.

We now expand the four-point function into two separable parts: the connected or non-Gaussian component $\langle \tilde{\delta}_1\tilde{\delta}_2\tilde{\delta}_3\tilde{\delta}_4 \rangle'$ and the Gaussian component, which using Wick's theorem can be expanded as the sum of the product of two-point functions.

$$\langle g_1g_2g_3g_4 \rangle_{ij} = \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \iint\frac{\mathrm{d}^3\boldsymbol{k}_1}{(2\pi)^3}\frac{\mathrm{d}^3\boldsymbol{k}_3}{(2\pi)^3} \iiiint \prod_{m=1}^{4} \left[\frac{\mathrm{d}^3\boldsymbol{q}_m}{(2\pi)^3}\tilde{W}(\boldsymbol{q}_m) \right] \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_1-\boldsymbol{q}_1)\cdot\boldsymbol{r}_i}\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_3-\boldsymbol{q}_3)\cdot\boldsymbol{r}_j} \left[\langle \tilde{\delta}_1\tilde{\delta}_2\tilde{\delta}_3\tilde{\delta}_4 \rangle' \right.$$
$$+ \langle \tilde{\delta}_1\tilde{\delta}_2 \rangle \langle \tilde{\delta}_3\tilde{\delta}_4 \rangle + \left[\langle \delta_{D,1}\delta_{D,2} \rangle + \langle \tilde{n}_1\tilde{n}_2 \rangle \right] \langle \tilde{\delta}_3\tilde{\delta}_4 \rangle + \langle \tilde{\delta}_1\tilde{\delta}_2 \rangle \left[\langle \delta_{D,3}\delta_{D,4} \rangle + \langle \tilde{n}_3\tilde{n}_4 \rangle \right]$$
$$+ \langle \tilde{\delta}_1\tilde{\delta}_3 \rangle \langle \tilde{\delta}_2\tilde{\delta}_4 \rangle + \left[\langle \delta_{D,1}\delta_{D,3} \rangle + \langle \tilde{n}_1\tilde{n}_3 \rangle \right] \langle \tilde{\delta}_2\tilde{\delta}_4 \rangle + \langle \tilde{\delta}_1\tilde{\delta}_3 \rangle \left[\langle \delta_{D,2}\delta_{D,4} \rangle + \langle \tilde{n}_2\tilde{n}_4 \rangle \right]$$
$$+ \langle \tilde{\delta}_1\tilde{\delta}_4 \rangle \langle \tilde{\delta}_2\tilde{\delta}_3 \rangle + \left[\langle \delta_{D,1}\delta_{D,4} \rangle + \langle \tilde{n}_1\tilde{n}_4 \rangle \right] \langle \tilde{\delta}_2\tilde{\delta}_3 \rangle + \langle \tilde{\delta}_1\tilde{\delta}_4 \rangle \left[\langle \delta_{D,2}\delta_{D,3} \rangle + \langle \tilde{n}_2\tilde{n}_3 \rangle \right]$$
$$+ \langle \delta_{D,1}\delta_{D,2}\tilde{n}_3\tilde{n}_4 \rangle + \text{all permutations} + \langle \tilde{n}_1\tilde{n}_2\tilde{n}_3\tilde{n}_4 \rangle + \langle \delta_{D,1}\delta_{D,2}\delta_{D,3}\delta_{D,4} \rangle \left.\right] . \tag{A17}$$

We have? omitted the positional arguments for $\tilde{\delta}_i$ and $\tilde{n}_i$, which are the same as for g_i in equation (A16). We defined $\delta_{D,i} = \delta_D(\boldsymbol{k}_i)$ and $\langle \tilde{\delta}_i\tilde{\delta}_j \rangle = P_{ij}(\boldsymbol{k}_i)\delta_D(\boldsymbol{k}_i + \boldsymbol{k}_j)$, where $P_{ij}(\boldsymbol{k})$ is the power spectrum.

Simplifying, the terms involving $\langle \xi_{12}(\boldsymbol{r}_i) \rangle \langle \xi_{34}(\boldsymbol{r}_j) \rangle$ cancel out, and using the fact that g_1 and g_3 belonged to field X and g_2 and g_4 belonged to field Y, we can write the covariance as

$$\mathrm{Cov} = \left[\frac{\mathcal{V}_W(\boldsymbol{r}_i-\boldsymbol{r}_j)}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j}\widehat{P}_{XX}(\boldsymbol{k})\widehat{P}_{YY}(\boldsymbol{k}) + \frac{\mathcal{V}_W(\boldsymbol{r}_i+\boldsymbol{r}_j)}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j}\widehat{P}_{XY}(\boldsymbol{k})\widehat{P}_{XY}(\boldsymbol{k}) + T_{XYXY} \right]$$
$$+ \left\{ \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j}\tilde{W}(\boldsymbol{k})\tilde{W}(-\boldsymbol{k})\left(\widehat{P}_{YY}(\boldsymbol{k}) + \widehat{P}_{XX}(\boldsymbol{k}) \right) \right\}$$
$$+ \left\{ \frac{1}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i}\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j}\tilde{W}(\boldsymbol{k})\tilde{W}(\boldsymbol{k})\left(\widehat{P}_{XY}(\boldsymbol{k}) + \widehat{P}_{XY}(\boldsymbol{k}) \right) \right\} . \tag{A18}$$

Here $\widehat{P}_{ij} = P_{ij} + P_{ij,N}$, where $P_{ij,N}$ is the noise power spectrum. T_{XYXY} is the connected term. To simplify expressions, we have assumed that the power spectrum is a slowly varying function of k and that we are working with modes much smaller than the survey size, so that $P(\boldsymbol{k}-\boldsymbol{q}) \approx P(\boldsymbol{k})$ and then $P(\boldsymbol{k})$ can be moved out of the window function integrals. For scales much smaller than the survey size,

$\mathcal{V}_W(\boldsymbol{r}) \to V_W$, where V_W is the physical volume of the survey, the expression simplifies to the more familiar form

$$
\begin{aligned}
\mathrm{Cov} = & \left[\frac{1}{V_W} \left(\int \frac{\mathrm{d}^3 \boldsymbol{k}}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j} \widehat{P}_{XX}(\boldsymbol{k})\widehat{P}_{YY}(\boldsymbol{k}) + \int \frac{\mathrm{d}^3 \boldsymbol{k}}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j} \widehat{P}_{XY}(\boldsymbol{k})\widehat{P}_{XY}(\boldsymbol{k}) \right) + T_{XYXY} \right] \\
& + \left\{ \frac{1}{V_W^2} \int \frac{\mathrm{d}^3 \boldsymbol{k}}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j} \tilde{W}(\boldsymbol{k})\tilde{W}(-\boldsymbol{k}) \left(\widehat{P}_{YY}(\boldsymbol{k}) + \widehat{P}_{XX}(\boldsymbol{k}) \right) \right\} \\
& + \left\{ \frac{1}{V_W^2} \int \frac{\mathrm{d}^3 \boldsymbol{k}}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_j} \tilde{W}(\boldsymbol{k})\tilde{W}(\boldsymbol{k}) \left(\widehat{P}_{XY}(\boldsymbol{k}) + \widehat{P}_{XY}(\boldsymbol{k}) \right) \right\} .
\end{aligned}
\tag{A19}
$$

The terms in square brackets ([ ]) are the usual covariance terms while the terms in braces ({}) arise when the means of the fields are not subtracted. These additional contributions depend on the survey window function and become less important as the survey size increases. In the case of a large uniform survey, $\lim_{V_W \to \infty} \tilde{W}(\boldsymbol{k}) = \delta_{\mathrm{D}}(\boldsymbol{k})$. As a result, the terms in braces ({}) approach zero faster (under the assumption $\widehat{P}(\boldsymbol{k}=0)=0$) and the two estimators (correlating mean zero field or correlating mean non-zero fields) are equivalent. However, in case $\widehat{P}(\boldsymbol{k}=0) \neq 0$, e.g. due to shot noise in case of galaxies, the ({}) terms approach the value of $P(\boldsymbol{k}=0)$. We also emphasize that this additional contribution to the covariance will be present in the analysis in Fourier space as well.

A2 Projected case

The projected correlation function is defined as the integral of the 3D correlation function over the line-of-sight separation, Π.

$$
\widehat{w}(r_p) = \int_{\Pi_{\min}}^{\Pi_{\max}} \mathrm{d}\Pi \, W(\Pi)\widehat{\xi}(r_p, \Pi),
\tag{A20}
$$

where $W(\Pi)$ is the line-of-sight weight function (not necessarily the same as the window function).

To compute the covariance, we start with equation (A17), carry out the line-of-sight integrals assuming the integration length is long (getting delta functions of the form $\delta_{\mathrm{D}}(\boldsymbol{k}_{\parallel,i} - \boldsymbol{q}_{\parallel,i})$). Then, carrying out integrals involving the line-of-sight window functions, we assume that the relevant line-of-sight modes are small, such that power spectrum is only dependent on the projected modes ($k_{\parallel} \ll k_{\perp}$, $P(\boldsymbol{k}) \approx P(\boldsymbol{k}_{\perp})$)

$$
\begin{aligned}
\mathrm{Cov} = & \int d\Pi \, W_Y(\Pi) W_Y(\Pi) \left[\frac{\mathcal{V}_W(\boldsymbol{r}_i - \boldsymbol{r}_j)}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \widehat{P}_{XX}(\boldsymbol{k}_{\perp})\widehat{P}_{YY}(\boldsymbol{k}_{\perp}) \right. \\
& \left. + \frac{\mathcal{V}_W(\boldsymbol{r}_i + \boldsymbol{r}_j)}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \widehat{P}_{XY}(\boldsymbol{k}_{\perp})\widehat{P}_{XY}(\boldsymbol{k}_{\perp}) + T_{XYXY} \right] \\
& + \int \mathrm{d}\Pi \, W_Y(\Pi) W_Y(\Pi) \left\{ \frac{L_W}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \tilde{W}_X(\boldsymbol{k}_{\perp})\tilde{W}_X(-\boldsymbol{k}_{\perp}) \left(\widehat{P}_{YY}(\boldsymbol{k}_{\perp}) + \widehat{P}_{XX}(\boldsymbol{k}_{\perp}) \right) \right. \\
& \left. + \frac{L_W}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \tilde{W}_X(\boldsymbol{k}_{\perp})\tilde{W}_X(\boldsymbol{k}_{\perp}) \left(\widehat{P}_{XY}(\boldsymbol{k}_{\perp}) + \widehat{P}_{XY}(\boldsymbol{k}_{\perp}) \right) \right\} .
\end{aligned}
\tag{A21}
$$

Here, we distinguished between the window functions of tracers X and Y, L_W is the line-of-sight length of the window function (of X) and we ignore the edge effects along the line of sight. Thus, the volume element can be written as

$$
\mathcal{V}_W(\boldsymbol{r}_p) = \mathcal{A}_W(\boldsymbol{r}_p) L_W.
\tag{A22}
$$

$\mathcal{A}_W$ is the physical survey area at the lens redshift.

Note that $\widehat{P}_{YY}$ can in principle be evaluated at a different epoch as Y_1 and Y_2 are at separation Π_i and Π_j, i.e. $P_{YY}(k_{\perp}) \sim P_{YY}(k_{\perp} \frac{\chi_z}{\chi_z + \Pi})$ where χ_z is the line-of-sight distance to the mean redshift where we are evaluating the covariance. Under the assumption that the power spectrum evolution within the $\Pi_{\max}$ limits is small, we keep $P_{YY}(k_{\perp})$ (ignoring its Π dependence), and simplify the expression as

$$
\begin{aligned}
\mathrm{Cov} = & \left[\frac{\mathcal{A}_W(\boldsymbol{r}_{p,i} - \boldsymbol{r}_{p,j})}{\mathcal{A}_W(\boldsymbol{r}_{p,i})\mathcal{A}_W(\mathbf{r}_{p,j})} \frac{\Delta\Pi_2}{L_W} \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \widehat{P}_{XX}(\boldsymbol{k}_{\perp})\widehat{P}_{YY}(\boldsymbol{k}_{\perp}) \right. \\
& \left. + \frac{\mathcal{A}_W(\boldsymbol{r}_i + \boldsymbol{r}_j)}{\mathcal{A}_W(\boldsymbol{r}_i)\mathcal{A}_W(\boldsymbol{r}_j)} \frac{\Delta\Pi_2}{L_W} \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \widehat{P}_{XY}(\boldsymbol{k}_{\perp})\widehat{P}_{XY}(\boldsymbol{k}_{\perp}) + T_{XYXY} \right] \\
& + \frac{\Delta\Pi_2 L_W^2}{\mathcal{V}_W(\boldsymbol{r}_{p,i})\mathcal{V}_W(\boldsymbol{r}_{p,j})} \left\{ \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \tilde{W}(\boldsymbol{k}_{\perp})\tilde{W}(-\boldsymbol{k}_{\perp}) \left(\widehat{P}_{YY}(\boldsymbol{k}_{\perp}) + \widehat{P}_{XX}(\boldsymbol{k}_{\perp}) \right) \right. \\
& \left. + \int \frac{\mathrm{d}^2 \boldsymbol{k}_{\perp}}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,i}} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}_{\perp}\cdot\boldsymbol{r}_{p,j}} \tilde{W}(\boldsymbol{k}_{\perp})\tilde{W}(\boldsymbol{k}_{\perp}) \left(\widehat{P}_{XY}(\boldsymbol{k}_{\perp}) + \widehat{P}_{XY}(\boldsymbol{k}_{\perp}) \right) \right\} ,
\end{aligned}
\tag{A23}
$$

where we defined

$$
\Delta\Pi_2 = \int \mathrm{d}\Pi \, W(\Pi) W(\Pi)
\tag{A24}
$$

$$
\Delta\Pi_1 = \int \mathrm{d}\Pi \, W(\Pi).
\tag{A25}
$$

For the case of galaxy clustering, we assume $W(\Pi)$ is a top-hat function for $\Pi \in [-100, 100]$, which leads to $\Delta\Pi_2 = \Delta\Pi_1 = 200 \, h^{-1}$ Mpc.

A3 Galaxy lensing case

We now use the formalism of Appendices A1 and A to derive the covariance for the galaxy–galaxy lensing case. We will assume the same sky coverage for the lens and shape samples. Note that the shear is a mean-zero field since lensing is only sensitive to the matter density contrast, and hence some of the terms in equation (A18) will drop out.

We begin by defining the observed shear as the sum of the true shear and noise.

$$\widehat{\gamma} = \gamma + \gamma_N \tag{A26}$$

We also assume that the mean shear around random points is not subtracted. In that case,
the galaxy–shear (projected) cross-correlation can be written as

$$\Delta\Sigma(\boldsymbol{r}_p) = \frac{1}{\mathcal{V}_W(\boldsymbol{r})} \int \mathrm{d}^3\boldsymbol{r}\, \Sigma_c(z_l, z_s) g(\boldsymbol{r})\widehat{\gamma}(\boldsymbol{r} + \boldsymbol{r}_p) W_\gamma(\boldsymbol{r} + \boldsymbol{r}_p) W_g(\boldsymbol{r}) \tag{A27}$$

$$= \frac{1}{\mathcal{V}_W(\boldsymbol{r})} \int \mathrm{d}^3\boldsymbol{r}\, \Sigma_c(z_l, z_s)(1 + \widehat{\delta})(\boldsymbol{r})\widehat{\gamma}(\boldsymbol{r} + \boldsymbol{r}_p) W_\gamma(\boldsymbol{r} + \boldsymbol{r}_p) W_g(\boldsymbol{r}) \tag{A28}$$

$$\Delta\Sigma(\boldsymbol{r}_p) = \frac{1}{\mathcal{V}_W(\boldsymbol{r})} \int \mathrm{d}^3\boldsymbol{r}\, \Sigma_c(z_l, z_s)\widehat{\delta}(\boldsymbol{r})\widehat{\gamma}(\boldsymbol{r} + \boldsymbol{r}_p) W_\gamma(\boldsymbol{r} + \boldsymbol{r}_p) W_g(\boldsymbol{r}). \tag{A29}$$

The covariance of two $\Delta\Sigma$ is

$$\mathrm{Cov}(\Delta\Sigma_{g_1\gamma_2}(\boldsymbol{r}_{p,i})\Delta\Sigma_{g_3\gamma_4}(\boldsymbol{r}_{p,j})) = \left\langle \Delta\Sigma_{g_1\gamma_2}(\boldsymbol{r}_{p,i})\Delta\Sigma_{g_3\gamma_4}(\boldsymbol{r}_{p,j}) \right\rangle - \left\langle \Delta\Sigma_{g_1\gamma_2}(\boldsymbol{r}_{p,i}) \right\rangle \left\langle \Delta\Sigma_{g_3\gamma_4}(\boldsymbol{r}_{p,j}) \right\rangle . \tag{A30}$$

Following the derivation in Appendices A1 and A and noting that the shear has a mean of zero, and $\Delta\Sigma$ is a projected galaxy–matter correlation function, the full covariance analogous to equation (A21) is

$$\begin{aligned}
\mathrm{Cov} = &\left[\int \mathrm{d}\Pi\, W(\Pi) W(\Pi) \frac{\mathcal{V}_W(\boldsymbol{r}_i - \boldsymbol{r}_j)}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}k_\perp\cdot r_i}\mathrm{e}^{\mathrm{i}k_\perp\cdot r_j} \widehat{P}_{gg}(\boldsymbol{k}_\perp)\widehat{P}_{\delta\delta}(\boldsymbol{k}_\perp) \right] \\
&+ \left[\int \mathrm{d}\Pi\, W(\Pi) W(\Pi) \frac{\mathcal{V}_W(\boldsymbol{r}_i + \boldsymbol{r}_j)}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}k_\perp\cdot r_{p,i}}\mathrm{e}^{-\mathrm{i}k_\perp\cdot r_{p,j}} \widehat{P}_{g\delta}(\boldsymbol{k}_\perp)\widehat{P}_{g\delta}(\boldsymbol{k}_\perp) + T_{g\gamma g\gamma} \right] \\
&+ \int \mathrm{d}\Pi\, W(\Pi) W(\Pi) \left\{ \frac{L_W^2}{\mathcal{V}_W(\boldsymbol{r}_i)\mathcal{V}_W(\boldsymbol{r}_j)} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}k_\perp\cdot r_i}\mathrm{e}^{\mathrm{i}k_\perp\cdot r_j} \tilde{W}(\boldsymbol{k}_\perp)\tilde{W}(-\boldsymbol{k}_\perp)\widehat{P}_{\delta\delta}(\boldsymbol{k}_\perp) \right\} ,
\end{aligned} \tag{A31}$$

where the lensing window function is

$$W(\Pi) = \overline{\rho}\, \frac{\Sigma_c(\chi_s, \chi_l)}{\Sigma_c(\chi_s, \chi_l + \Pi)} . \tag{A32}$$

The line-of-sight integral in the terms involving $P_{\delta\delta}$ leads to the shear autocorrelation function, and the final expression is

$$\begin{aligned}
\mathrm{Cov} = &\left[\frac{\mathcal{V}_W(\boldsymbol{r}_{p,i} - \boldsymbol{r}_{p,j})}{\mathcal{V}_W(\boldsymbol{r}_{p,i})\mathcal{V}_W(\boldsymbol{r}_{p,j})} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}k_\perp\cdot r_i}\mathrm{e}^{\mathrm{i}k_\perp\cdot r_{p,j}} \cos 2\phi_{k,i} \cos 2\phi_{k,j} \widehat{P}_{gg}(\boldsymbol{k}_\perp)\Sigma_c^2\left(\frac{\sigma_\gamma^2}{n_s} + P_{\kappa\kappa} \right) \right] \\
&+ \left[\frac{\mathcal{V}_W(\boldsymbol{r}_{p,i} + \boldsymbol{r}_{p,j})\Delta\Pi_2}{\mathcal{V}_W(\boldsymbol{r}_{p,i})\mathcal{V}_W(\boldsymbol{r}_{p,j})} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}k_\perp\cdot r_{p,i}}\mathrm{e}^{-\mathrm{i}k_\perp\cdot r_{p,j}} \cos 2\phi_{k,i} \cos 2\phi_{k,j} \overline{\rho}^2 \widehat{P}_{g\delta}(\boldsymbol{k}_\perp)\widehat{P}_{g\delta}(\boldsymbol{k}_\perp) + T_{g\gamma g\gamma} \right] \\
&+ \left\{ \frac{L_W^2}{\mathcal{V}_W(\boldsymbol{r}_{p,i})\mathcal{V}_W(\boldsymbol{r}_{p,j})} \int \frac{\mathrm{d}^2\boldsymbol{k}_\perp}{(2\pi)^3} \mathrm{e}^{-\mathrm{i}k_\perp\cdot r_{p,i}}\mathrm{e}^{\mathrm{i}k_\perp\cdot r_{p,j}} \cos 2\phi_{k,i} \cos 2\phi_{k,j} \tilde{W}(\boldsymbol{k}_\perp)\tilde{W}(-\boldsymbol{k}_\perp)\Sigma_c^2\left(\frac{\sigma_\gamma^2}{n_s} + P_{\kappa\kappa} \right) \right\} ,
\end{aligned} \tag{A33}$$

where the convergence power spectrum is

$$P_{\kappa\kappa}(k) = \int_0^{\chi_s} \mathrm{d}\chi\, \frac{\overline{\rho}}{\Sigma_c(\chi, \chi_s)} \frac{\overline{\rho}}{\Sigma_c(\chi, \chi_s)} P_{\delta\delta}\left(k\frac{\chi_l}{\chi} \right) . \tag{A34}$$

For lensing, using the full lens and source redshift distribution, we compute $\Delta\Pi_1 \approx 900\,h^{-1}$ Mpc and $\Delta\Pi_2 \approx 700\,h^{-1}$ Mpc.

Again as in Appendix A1, the terms in square brackets ([ ]) are the usual covariance terms for the mean zero fields and the terms in curly brackets ({}) are additional contribution from terms involving $\delta_{D,i}$, arising from the suboptimal estimator without the mean subtracted.

Note that this additional contribution only depends on the window function of the lens sample and is independent of the clustering or number density of the lens sample. Hence, this noise term is consistent across the real lens galaxy sample and uniformly distributed random points, which is why the subtraction of the shear around random points removes this contribution to the covariance. Also the window function-dependence of this term is the reason why the jackknife/subsample methods of estimating errors show increased contribution from this term compared to the standard deviation across independent mock catalogues, since the effective window function for the subsamples is smaller.

A4 Numerical estimates

A4.1 Clustering

To compute numerical estimates, we assume angular symmetry for both the power spectra and the window function. Further, in the case of galaxy clustering, $X \equiv Y$ and $P_{XX} = b_g^2 P_{\delta\delta}$. After carrying out the angular and line-of-sight integrals (for the projected correlation function) in equation (A23), we get

$$\mathrm{Cov}(w_{gg}) = \left[2\frac{\mathcal{A}_W(|\boldsymbol{r}_{p,i} - \boldsymbol{r}_{p,j}|)}{\mathcal{A}_W(r_{p,i})\mathcal{A}_W(r_{p,j})} \frac{\Delta\Pi_2}{L_W} \int \frac{\mathrm{d}k\,k}{2\pi} J_0(kr_{p,i}) J_0(kr_{p,j}) \left(b_g^2 P_{\delta\delta}(k) + \frac{1}{n_g} \right)^2 + T_{gggg} \right]$$
$$+ 4\left\{ \frac{2\Delta\Pi_2}{\mathcal{A}_W(r_{p,i})\mathcal{A}_W(r_{p,j})} \int \frac{\mathrm{d}k\,k}{2\pi} J_0(kr_{p,i}) J_0(kr_{p,j}) \tilde{W}(k)\tilde{W}(k) \left(b_g^2 P_{\delta\delta}(k) + \frac{1}{n_g} \right) \right\}, \tag{A35}$$

where J_n is the Bessel function of order n, Π_{max} is the line-of-sight integration length and

$$\mathcal{A}_W(r) = \int \frac{\mathrm{d}k\,k}{2\pi} J_0(kr)\,(W(k))^2 \tag{A36}$$

$$\mathcal{A}_W(|\boldsymbol{r}_i - \boldsymbol{r}_j|) = \int \frac{\mathrm{d}k\,k}{2\pi} J_0(kr_i) J_0(kr_j)\,(W(k))^2\,. \tag{A37}$$

For the window function, we assume a circular geometry with a survey area of 9000 deg^2, with a mean redshift of $z = 0.27$. $W(k)$ is defined as

$$W(k) = 2\pi R^2 \frac{J_1(kR)}{kR}, \tag{A38}$$

where $R \approx 1275\,h^{-1}$ Mpc is the physical scale corresponding to 95 deg at $z = 0.27$.

Finally, to get the covariance for bins in r_p, $\mathrm{COV}_{\mathrm{bin}}$, we integrate the covariance in equation (A35) as

$$\mathrm{COV}_{\mathrm{bin}} = \frac{\int_{r_{p,i,l}}^{r_{p,i,h}} \mathrm{d}r'_{p,i}\,r'_{p,i} \int_{r_{p,j,l}}^{r_{p,j,h}} \mathrm{d}r'_{p,j}\,r'_{p,j}\,\mathrm{COV}(r'_{p,i}, r'_{p,j})}{\int_{r_{p,i,l}}^{r_{p,i,h}} \mathrm{d}r'_{p,i}\,r'_{p,i} \int_{r_{p,j,l}}^{r_{p,j,h}} \mathrm{d}r'_{p,j}\,r'_{p,j}}, \tag{A39}$$

where $r_{p,i,l}$, $r_{p,i,h}$ are the lower and upper limits of the bins, respectively.

A4.2 Galaxy lensing

We carry out the angular integrals in equation (A33), to get

$$\mathrm{Cov}(\Delta\Sigma) = \left[\frac{\mathcal{A}_W(\boldsymbol{r}_{p,i} - \boldsymbol{r}_{p,j})}{\mathcal{A}_W(\boldsymbol{r}_{p,i})\mathcal{A}_W(\boldsymbol{r}_{p,j})} \frac{1}{L_W} \int \frac{\mathrm{d}k\,k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) \Sigma_c^2 \left(b_g^2 P_{\delta\delta}(k) + \frac{1}{n_g} \right) \left(P_{\kappa\kappa}(k) + \frac{\sigma_\gamma^2}{n_s} \right) \right]$$
$$+ \left[\frac{\mathcal{A}_W(\boldsymbol{r}_{p,i} - \boldsymbol{r}_{p,J})}{\mathcal{A}_W(\boldsymbol{r}_{p,i})\mathcal{A}_W(\boldsymbol{r}_{p,J})} \frac{\Delta\Pi_2}{L_W} \int \frac{\mathrm{d}k\,k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) \left(b_g r_{cc} \overline{\rho} P_{\delta\delta}(k) \right)^2 + T_{g\gamma g\gamma} \right]$$
$$+ \left\{ \frac{1}{\mathcal{A}_W(\boldsymbol{r}_{p,i})\mathcal{A}_W(\boldsymbol{r}_{p,j})} \int \frac{\mathrm{d}k\,k}{2\pi} J_2(kr_{p,i}) J_2(kr_{p,j}) \tilde{W}(k)\tilde{W}(k) \Sigma_c^2 \left(P_{\kappa\kappa}(k) + \frac{\sigma_\gamma^2}{n_s} \right) \right\}, \tag{A40}$$

$\mathrm{Cov}(\Delta\Sigma)$ is then integrated to get the covariance in bins as described in equation (A39).

APPENDIX B: CLUSTERING RESULTS

In this appendix, we present the comparison of different estimators and error estimation methods for the galaxy clustering measurements.

We begin by defining the standard LS estimator for clustering (Landy & Szalay 1993)

$$\widehat{\xi}_{\mathrm{LS}}(r_p, \Pi) = \frac{(D-R)^2}{RR} = \frac{DD - 2DR + RR}{RR}, \tag{B1}$$

where ξ is the 3D correlation function, r_p is the projected separation on the sky and Π is the line-of-sight separation between the pair of galaxies. The use of $D - R$ indicates that we are correlating overdensity fields. The estimator can then be expanded into its standard pair counting form. DD denotes summation over all galaxy–galaxy pairs within the bin, DR are the cross pairs between galaxies and randoms and RR are the random–random pairs.

In addition, we define the basic estimator from pair counting

$$\widehat{\xi}_{s2}(r_p, \Pi) = \frac{DD}{RR} - 1 \tag{B2}$$

Motivated by the galaxy–galaxy lensing estimator without subtraction of the mean galaxy density, we also define the estimator correlating an overdensity field $(D - R)$ with a density field (D)

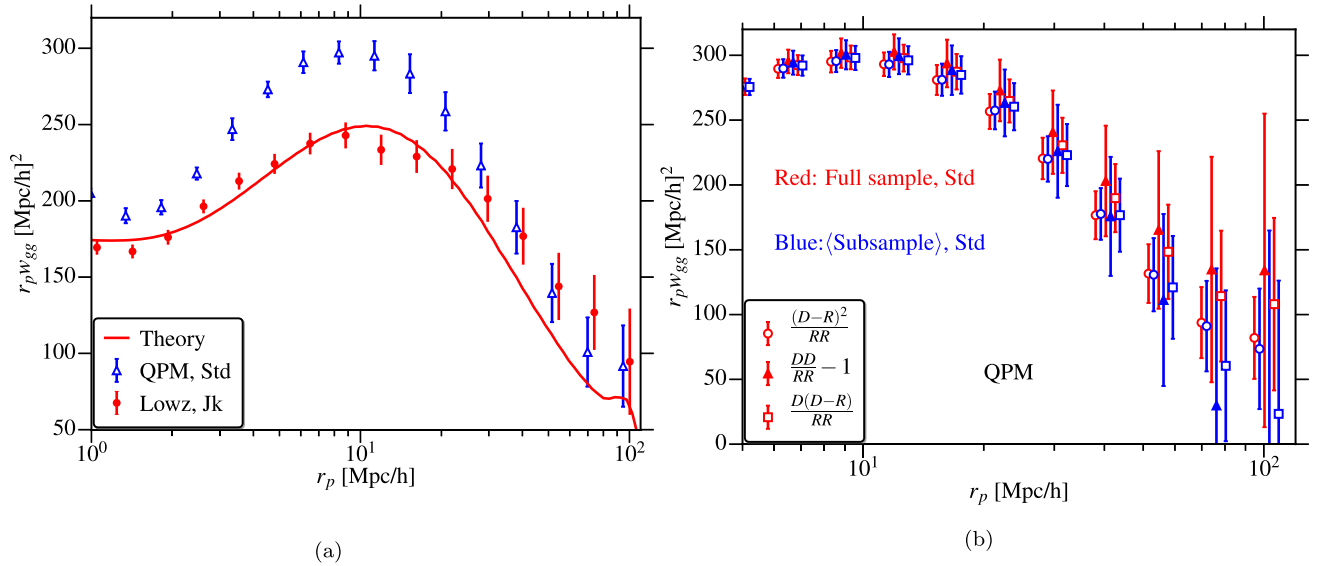$$\widehat{\xi}_{s1}(r_p, \Pi) = \frac{D(D-R)}{RR} = \frac{DD - DR}{RR}. \tag{B3}$$

Figure B1. (a) Comparison of the projected clustering for the LOWZ sample and the QPM mocks. The LOWZ errors are from the jackknife method, while the QPM values are the mean and standard deviation across 150 realizations. The red line is the Planck 2015 ΛCDM prediction along with the best-fitting bias from Singh et al. (2017); note that the points on large scales have correlated errors. (b) Comparison of clustering measurements using different estimators (note different r_p range). ⟨Subsample⟩ refers to the mean signal across the subsamples, in each realization. We then take the mean and standard deviation of ⟨Subsample⟩ across realizations.

We want to work with projected correlation functions, analogous to galaxy–galaxy lensing. Thus, we integrate ξ_{gg} over the line of sight to obtain the projected correlation function w_{gg}.

$$\widehat{w}_{gg}(r_p) = \int_{-\Pi_{max}}^{\Pi_{max}} \widehat{\xi}(r_p, \Pi) d\Pi. \tag{B4}$$

The choice of Π_{max} depends on two considerations: we want to choose large Π_{max} to capture the full correlation function and to mitigate the effect of redshift space distortions (Kaiser 1987). However, in a survey of finite redshift window, the bins at large Π are also noisier which increases the noise in the projected correlation function as well. In this work, we use $\Pi_{max} = 100 \, h^{-1}$ Mpc with linear bins of size $d\Pi = 10 \, h^{-1}$ Mpc.

In Fig. B1, we show the clustering measurement for the LOWZ sample as well as for one realization of the QPM mocks, with jackknife errors for both. At small scales, the clustering between the mocks and data does not agree very well with a maximum difference of order ~ 30 per cent. This is expected, since the QPM mocks are generated using low-resolution simulations, which only resolve the large-scale density field (White et al. 2014). Here, we compare the estimators and error estimations self-consistently from the QPM mocks, and thus the failure to exactly match the LOWZ sample clustering is not important.

In Fig. B2, we compare the clustering error estimates using different estimators. The LS estimator yields the lowest errors followed by the estimator in equation (B3). The relative trends between the estimators are consistent with the theory estimates from expressions in equations (A18) and (A35) and, more generally, with the idea that each time you substitute a zero-mean field with a field that has a non-zero mean, the variance increases. We caution that for the clustering measurements, we have not completely explored the consistency between theory and empirical error estimates. Our theory estimates do not capture the full contributions from non-linear bias, redshift space distortions and the connected part of the covariance. Also in equation (A18) $\widehat{P}_{ij}(\mathbf{k}) = \widehat{P}(\mathbf{k})$ and some terms from equation (A18) will be removed in the case of the estimator in equation (B3); see equation (A23).

In Fig. B2, we also compare the error estimates from jackknife and standard deviation across different mock realizations, using different estimators. Our results suggest that the jackknife overestimates the errors at all scales, even when the scales are larger than the jackknife region size. This is contrary to the expectations, since at scales larger than the jackknife region size, the assumption that the regions are independent is violated and thus the errors are expected to be underestimated. However, since the jackknife regions are much smaller than the survey size, the contribution from supersample variance and other window function-dependent terms is expected to be larger in the jackknife. In the case of the LS estimator, our theory estimates (the difference between solid and dashed green lines) suggest that the increase in error from edge effects is also important at large scale and can lead to the jackknife errors being overestimated by 10–20 per cent. For the non-optimal estimators, the increased contribution from the additional terms identified in Appendix A dominates and hence the increase in jackknife errors compared to the standard deviation is substantially more, when compared to the LS estimator.
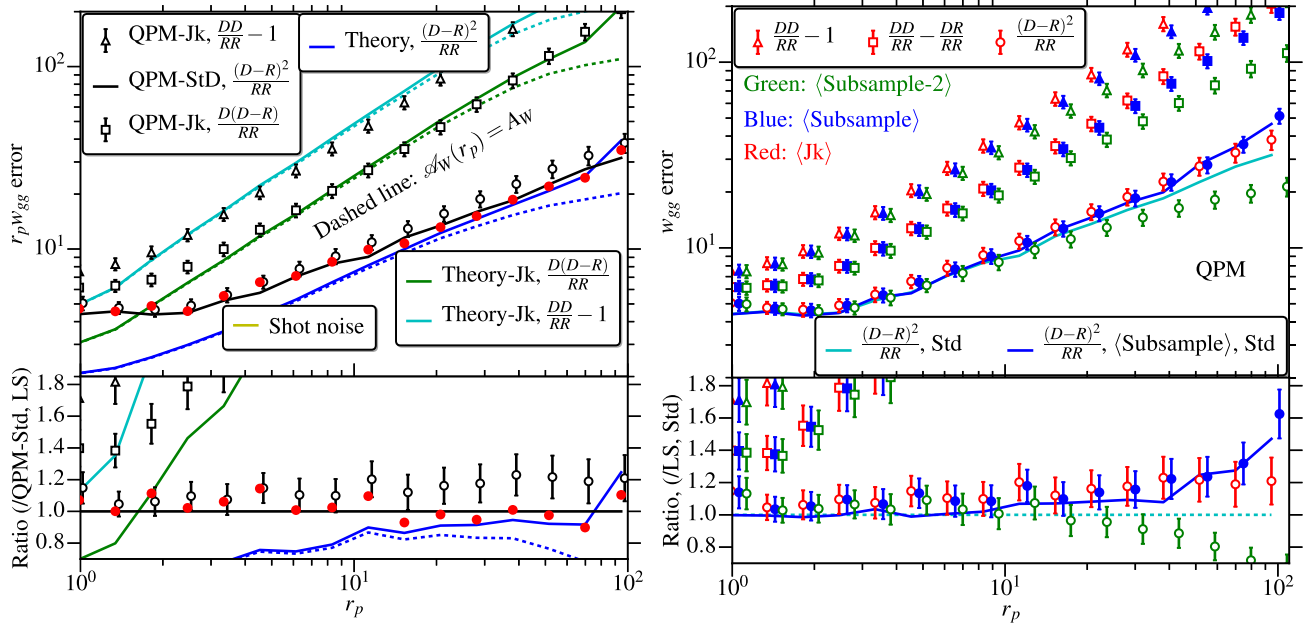
Figure B2. *Left-hand panel*: comparison of errors for different clustering estimators defined in equations (B1)–(B3). The LS ($\frac{(D-R)^2}{RR}$) estimator, equation (B1), gives the lowest error followed by the estimator in equation (B3) and then the estimator in equation (B2). Also shown are the jackknife errors for the LS estimator, with the jackknife overestimating the errors by ∼10–20 per cent at all scales. We additionally show the estimated errors from the theory predictions. Note that the theory estimates use the linear theory+halofit matter power spectrum, and do not include contributions from non-linear galaxy bias and connected parts of the covariance, hence the theory errors are underestimated at small scales. The difference between the solid and dashed theory lines are due to the edge effects as estimated by $\mathcal{A}_W$. *Right-hand panel*: comparison between the different error estimation methods in mocks. ⟨Subsample⟩ refers to the mean and standard deviation across subsamples. ⟨Subsample-2⟩ is similar to ⟨Subsample⟩, except for each subsample we also count the cross terms with other subsamples. This reduces edge effects but also leads to correlations between different subsamples at large scales.

This paper has been typeset from a TeX/LaTeX file prepared by the author.