Enhancing BOSS bispectrum cosmological constraints with maximal compression

Davide Gualdi1⋆, Héctor Gil-Marín2,3, Robert L. Schuhmann4,5, Marc Manera6,7, Benjamin Joachimi1, Ofer Lahav1

1Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
2Sorbonne Université, Institut Lagrange de Paris (ILP), 98 bis Boulevard Arago, 75014 Paris, France
3Laboratoire de Physique Nucléaire et de Hautes Energies, Université Pierre et Marie Curie, Paris, France
4Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK
5School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK
6Institut de Física d’Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain
7Centre for Mathematical Sciences, DAMTP, Cambridge University, Wilberforce Rd, Cambridge CB3 0WA

June 11, 2018

ABSTRACT

We apply two compression methods to the galaxy power spectrum monopole/quadrupole and bispectrum monopole measurements from the BOSS DR12 CMASS sample. Both methods reduce the dimension of the original data-vector to the number of cosmological parameters considered, using the Karhunen-Loève algorithm with an analytic covariance model. In the first case, we infer the posterior through MCMC sampling from the likelihood of the compressed data-vector (MC-KL). The second, faster option, works by first Gaussianising and then orthogonalising the parameter space before the compression; in this option (G-PCA) we only need to run a low-resolution preliminary MCMC sample for the Gaussianization to compute our posterior. Both compression methods accurately reproduce the posterior distributions obtained by standard MCMC sampling on the CMASS dataset for a k-space range of $0.03 - 0.12 \, h / \text{Mpc}$. The compression enables us to increase the number of bispectrum measurements by a factor of $\sim 23$ over the standard binning (from 116 to 2734 triangles used), which is otherwise limited by the number of mock catalogues available. This reduces the 68% credible intervals for the parameters $(b_1, b_2, f, \sigma_8)$ by $(-24.8\%, -52.8\%, -26.4\%, -21\%)$, respectively. The best-fit values we obtain are $(b_1 = 2.31 \pm 0.17, b_2 = 0.77 \pm 0.19, f(z_{\text{CMASS}}) = 0.67 \pm 0.06, \sigma_8(z_{\text{CMASS}}) = 0.51 \pm 0.03)$. Using these methods for future redshift surveys like DESI, Euclid and PFS will drastically reduce the number of simulations needed to compute accurate covariance matrices and will facilitate tighter constraints on cosmological parameters.

Key words: cosmological parameters, large-scale structure of Universe, methods: analytical, data analysis, statistical

1 INTRODUCTION

Large datasets have recently become available from current cosmological surveys (Planck,1 Ade et al. 2014; Sloan Digital Sky Survey,2 Eisenstein et al. 2011; KiDS de Jong et al. 2013; DES, Dark Energy Survey Collaboration et al. 20163) and even larger ones will be provided in future by DESI4, Levi et al. (2013); Euclid5, Laureijs et al. (2011); PFS6, Takada et al. (2014) and the LSST7, LSST Science Collaboration et al. (2009). In order to exploit their

* Contact e-mail: davide.gualdi.14@ucl.ac.uk
1 http://sci.esa.int/planck/
2 http://www.sdss3.org/surveys/boss.php
3 https://www.darkenergysurvey.org
4 http://desi.lbl.gov
5 http://sci.esa.int/euclid/
6 http://pfs.ipmu.jp
7 https://www.lsst.org/

© 2017 The Authors
full potential, is desirable to go beyond standard two-points statistics (2pt).

Three-points statistics (3pt) are a complementary probe that is possible to investigate both in configuration clustering and Fourier space and have been used extensively in galaxy clustering analyses (Groth & Peebles 1977, Fry 1984, Fry & Gaztanaga 1993, Frieman & Gaztanaga 1994, Mateus et al. 1997, Verde et al. 1998, Heavens et al. 1998, Scoccimarro et al. 1998a, Scoccimarro 2000, Sefusatti et al. 2006). Deviations from General Relativity (Bosch & Jain 2009; Bernardeau & Brax 2011; Gil-Marín et al. 2011) and primordial non-Gaussianities (Fry & Scherrer 1994; Gangui et al. 1994; Verde et al. 2000; Liguori et al. 2010; Tellarini et al. 2016) have been investigated using 3pt statistics. Their potential in lifting degeneracies present at 2pt level has been shown by the most recent measurement on the BOSS dataset, for the bispectrum by Gil-Marín et al. (2017) and for the 3pt correlation function by Slepian et al. (2017a). Baryonic acoustic oscillations (BAO) have also been measured using the 3pt correlation function by Slepian et al. (2017b) and detected using the bispectrum by Pearson & Samushia (2017).

Recently, 3pt statistics have been studied in the case of 21cm emission lines by Hoffmann et al. (2018). For what concerns weak lensing, its effect on 3pt galaxy clustering have been studied by Schmidt et al. (2008). Moreover the weak lensing bispectrum has been object of several studies in recent years (Takada & Jain 2004; Joachimi et al. 2009; Kayo et al. 2013; Kayo & Takada 2013). The skewness of mass aperture statistic was considered by Jarvis et al. (2004) while the 3pt correlation function of cosmic shear was analysed by Schneider et al. (2005); Kilbinger & Schneider (2005). Higher order statistics like the bispectrum via gravitational lensing have been investigated also by Simon et al. (2013); Fu et al. (2014); Simon et al. (2015); Pyne et al. (2017).

Besides being computationally more expensive than 2pt statistics, 3pt statistics present the drawback to be described by very large data-vectors, which in turn require a high number of simulations to accurately estimate their covariance matrix (Hartlap et al. 2007). In Gualdi et al. (2018), Paper I from now on, we presented two methods to compress the redshift-space galaxy bispectrum, namely MC-KL (Markov chain Monte Carlo sampling + Karhunen-Loève compression) and PCA + KL (principal component analysis transformation + Karhunen-Loève compression). MC-KL consists in sampling via MCMC the compressed data-vector’s likelihood. PCA + KL reconstructs the multidimensional physical posterior distribution from the 1D posterior of orthogonalised parameters obtained by diagonalising the Fisher information matrix. Modifications/improvements of the Karhunen-Loève algorithm were introduced also by Heavens et al. (2000) and recently by Heavens et al. (2017); Alting & Wandelt (2018); Alting et al. (2018) also with the target of data compression.

In this work we apply our compression methods to both the power spectrum monopole/quadrupole and to the bispectrum monopole measurements from the CMASS sample of BOSS DR12. While the MC-KL is more flexible than the PCA + KL method since doesn’t require the multidimensional Gaussian posterior assumption, the PCA + KL is much faster in terms of computational time and requires far fewer computational resources (it can be run on standard laptop). We compare both methods and test their convergence in terms of deriving equivalent posterior distributions.

In order to make the PCA + KL method applicable also to parameter spaces with strong degeneracies, for which the posterior Gaussianity approximation is no longer valid, we introduce a pre-Gaussianisation step based on the algorithm developed by Schulmann et al. (2016).

We measure the bispectrum monopole using the same code used for the BOSS DR12 analysis done by Gil-Marín et al. (2017). We vary the size of the triangle vectors by changing the bin size Δk for k, which returns different number of triangular shapes given the minimum and maximum scales. For the same number of triangles the compression returns posterior distributions slightly larger than the MCMC counterparts. However, when compressing a much larger number of triangles (which cannot be done for the MCMC on the full data-vector because of the limited number of mocks available constraint), the posterior distribution becomes more Gaussian and narrow. It eventually returns tighter constraints than the ones obtained by the standard analysis. In Sec. 2 we describe the data set and the galaxy mocks used to estimate the covariance matrix together with the settings of our analysis. In Sec. 3 we present the analytical model used for the data-vector considered and the analytical expression of the covariance matrix used to derive the weights for the compression. In Sec. 4 we recap the compression methods applied including the Gaussianisation extension for the original PCA + KL method. We report the performance of the compression methods compared to the MCMC sampling for the cases in which it is possible to run it on the full data-vector in Sec. 5. We describe the gain in parameter constraints as a function of the number of triangle configurations used in the bispectrum monopole data-vector component in Sec. 6. We test the flexibility and accuracy of the compression methods presented in Sec. 7. Finally we conclude summarising our results in Sec. 8. In Appendix A we report the full derivation of all the analytic expressions used in the analysis. In Appendix B additional validation tests are presented.

2 DATA, MOCKS AND ANALYSIS

2.1 DR12 BOSS data and mocks catalogues

In this paper we use the CMASS galaxy sample (0.43 ≤ z ≤ 0.70) of the Baryon Oscillation Spectroscopic Survey (BOSS Dawson et al. 2013) which is part of the Sloan Digital Sky Survey III (Eisenstein et al. 2011). In the final data release DR12 the CMASS sample contains the spectroscopic redshift of 777202 galaxies (see Gil-Marín et al. 2017 and Alam et al. 2017 for more details).

In order to accurately numerically estimate the covariance matrix it is necessary to employ a large suite of mock galaxy catalogues. These are different realizations of the same region of the Universe based on methods such as second-order Lagrangian perturbation theory (2LPT Scoccimarro & Sheth 2002; Manera et al. 2013) or augmented Lagrangian perturbation theory (ALPT) as described in Kitaura & Heß (2013). By measuring the data-vector of interest on each one of these catalogues we can numerically estimate the covariance matrix which will be used in the likelihood evaluation. In this work we use subsets of the 2048 realisations of the MultiDark Patchy BOSS DR12 mocks by Kitaura et al. (2016). This set of mocks has been run using the underlying cosmology: Ω_{Λ} = 0.693,
\[ \Omega_{m}(z = 0) = 0.307, \ \Omega_{b}(z = 0) = 0.048, \ \sigma_{8}(z = 0) = 0.829, \ n_{s} = 0.96, \ h_{0} = 0.678. \]

2.2 Analysis settings

For the power spectrum monopole and quadrupole the bin size was fixed to \( \Delta k = 0.01h/\text{Mpc} \). We measured the bispectrum monopole from both data and mocks using different multiples of the fundamental frequency defined as \( k_{f} = \frac{(2\pi)}{L} \) where \( V_{s} \) is the survey volume which in this case was the cubic box volume \( V_{s} = L_{s}^{3} = (3500 \text{Mpc}/h)^{3} \) used to analyse the galaxy mocks. In particular, the considered bin sizes for the bispectrum are \( \Delta k = (6, 5, 4, 2) \times k_{f} \) respectively, corresponding to 116, 195, 404 and 2734 triangles used between 0.03 < \( k_{i} \) [h/\text{Mpc}] < 0.12. The largest bin size \( \Delta k = 6 \times k_{f} \) corresponds to the one used in the BOSS collaboration analysis done by Gil-Marín et al. (2017). For the \( k \)-range considered in the BOSS analysis the \( \Delta k_{b} (\Delta k = 6 \times k_{f}) \) binning case corresponded to 825 fundamental triangle configurations while \( \Delta k_{2} \) would have corresponded to more than \( \sim 7000 \) triangles.

In all the parameter estimation analyses that we are going to perform, we use the covariance matrix derived from the galaxy catalogues described above (see Sec. 2.1). In particular, we use 1400 mocks to estimate the covariance matrix when running the MCMC sampling on the full data-vector. We use 700 when the analysis is performed using the compressed data-vector.

The largest scales considered in this work are \( k_{\text{min}} = 0.03 h/\text{Mpc} \) for both power spectrum monopole and quadrupole and \( k_{\text{min}} = 0.02 h/\text{Mpc} \) for the bispectrum monopole. The smallest scales considered are \( k_{\text{max}} = 0.09 h/\text{Mpc} \) and \( k_{\text{max}} = 0.12 h/\text{Mpc} \) for power spectrum (monopole and quadrupole) and bispectrum monopole respectively. The lower \( k_{\text{max}} \) used for the power spectrum is due to the fact that we did not include 1-loop corrections for it, hence we consider only scales belonging to the quasi-linear regime. We chose a higher \( k_{\text{max}} \) for the bispectrum since we implemented the effective model developed by Gil-Marín et al. (2014) which works up to non-linear scales.

Thefiducial cosmology chosen for the analysis corresponds to a flat-\( \Lambda \)CDM model close to the one reported in Planck Collaboration et al. (2016). In particular, we set \( \Omega_{m}(z = 0) = 0.31, \ \Omega_{b}(z = 0) = 0.049, \ A_{s} = 2.21 \times 10^{-9}, \ n_{s} = 0.9624, \ h_{0} = 0.6711 \).

In order to compute the covariance terms and the derivatives of the model necessary for the compression, we fix the fiducial value of the galaxy bias model parameters, the growth rate and the amplitude of dark matter fluctuations to the ones obtained by running a preliminary low-resolution MCMC (\( b_{1} = 2.5478, b_{2} = 1.2127, f = 0.7202, \sigma_{8} = 0.4722 \)). The Finger-of-God parameters for both power spectrum and bispectrum \( c_{\text{FoG}}^{P} \) and \( c_{\text{FoG}}^{B} \) have been set to zero after checking that for the range of scales considered (quasi-linear regime) they were compatible with zero. In Section 7 we check that the choice of fiducial parameters used to compute the derivatives of the mean of the data-vector and the analytical covariance matrix does not significantly influence the results of the compression.

### 3 DATA-VECTOR AND COVARIANCE MATRIX

In order to measure the power spectrum and bispectrum from the data and the mocks catalogues we use the estimators described in Gil-Marín et al. (2016a,b). These are based on the weighted field of density fluctuations (Feldman et al. 1994):

\[ F_{\lambda}(r) = \frac{w_{\text{KGP}}(r)}{I_{d}} \left[ w_{c}(r)n(r) - \alpha n_{\text{syn}}(r) \right], \]

where \( w_{c} \) is the weight taking into account all the measurement systematics (redshift failure, fiber collision, target density variations), \( w_{\text{KGP}} \) (Feldman, Kaiser and Peacock) ensures the condition of minimum variance, \( n \) is the observed number density of galaxies, \( n_{\text{syn}} \) is the number density of objects in a synthetic catalogue and \( I_{d} \) is the normalisation of the amplitude of the observed power (\( \lambda = 2, 3 \) for power spectrum and bispectrum, respectively). \( \alpha \) is the ratio between weighted number of observed galaxies over the weighted number of objects in the synthetic catalogues.

3.1 Power spectrum monopole and quadrupole

The redshift-space galaxy power spectrum model adopted in this work is a linear one including redshift-space distortions (RSD) plus a damping function taking into account the Finger-of-God (FoG) effect:

\[ P_{g}^{s}(k, \mu) = D_{\text{FoG}}^{P} \left( k, \mu, \sigma_{\text{FoG}}^{P}(z) \right) Z_{2}^{s}(k)^{2} F_{m}^{\text{lin}}(k), \]

where \( k \) is the module of the wave vector \( k \) and \( \mu \) is the cosine of the angle between the wave vector and the line of sight. The standard redshift-space distortion kernels \( Z_{2}^{s} \) are reported in the Appendix of Gil-Marín et al. (2014) together with the FoG damping function expression, \( \sigma_{\text{FoG}}^{P}(z) \) is the FoG free parameter for the power spectrum. For the range of scales considered in this work the linear RSD model has been proved to be a good approximation (Taruya et al. 2010). The redshift-space galaxy power spectrum can be expanded in terms of Legendre polynomials using its dependence on \( \mu \):

\[ P_{g}^{s}(k, \mu) = \sum_{\ell=0}^{\infty} P_{g}^{(\ell)}(k) L_{\ell}(\mu), \]

where \( L_{\ell}(\mu) \) is the \( \ell \)-order Legendre polynomial. Almost all the signal is contained in the first two even multipoles, the monopole and the quadrupole (\( \ell = 0, 2 \)). These can be found by inverting the above expression:

\[ P_{g}^{(\ell)}(k) = \frac{2\ell + 1}{2} \int_{-1}^{1} d\mu P_{g}^{s}(k, \mu) L_{\ell}(\mu). \]

3.2 Analytical expression for \( P_{g}^{(0, 2)} \) covariance matrices

Defining an estimator as in Appendix A1, it is possible to derive the expression for the Gaussian term of the power spectrum monopole and quadrupole covariance matrices (Appendix A2):

\[ C_{G}^{P_{g}}(k_{1}, k_{2}) = \left( \frac{2\ell + 1}{2} \right)^{2} \frac{2L_{\ell}^{\text{K}}}{N_{p}(k_{1})} P_{g}^{(\ell)}(k_{1}) P_{g}^{(\ell)}(k_{2}). \]
where $\delta^K_{ij}$ is the Kronecker delta between $k_1$ and $k_2$, while $N_p(k_1)$ is the number of pairs of grid points inside the estimator integration volume in Fourier space $V_e = 4\pi k^2 \Delta k$ (Scoccimarro et al. 1998b) and it is proportional to an effective survey volume $V_e$. The $V_e$ normalisation is used to obtain a closer match between the analytic and mocks covariance matrices. Please refer to Eqs. A2 and A11 for more details. We set the cross covariance between power spectrum monopole and quadrupole to zero.

### 3.3 Bispectrum monopole

For the redshift-space galaxy bispectrum we adopt the effective model presented in Gil-Marín et al. (2014), which modifies the redshift-space distortion kernels derived from perturbations theory in order to better fit the data at non-linear scales (see the Appendix of the paper above for the full expressions). The tree level has also been corrected to take into account the Finger-of-God damping effect:

$$B^0_{\ell}(k_1, k_2, k_3) = B^0_{\mathrm{FOG}}(k_1, k_2, k_3, \sigma^B_{\mathrm{FOG}}[z])$$

$$\times \left[ Z_{\ell}^1(k_1) Z_{\ell}^2(k_2) Z_{\ell}^3(k_3) Z_{\ell}^4_{\mathrm{eff}}[k_1, k_2] P^m_{\ell m}(k_1) P^m_{\ell m}(k_2) + \text{cyc.} \right],$$

(6)

where $\sigma^B_{\mathrm{FOG}}[z]$ is the FoG free parameter for the bispectrum. The monopole of the bispectrum corresponds to the average of all the possible orientations of a determinate triangle with respect to the line of sight. It can therefore be obtained by integrating over two angular coordinates:

$$B^0_{\ell}(k_1, k_2, k_3) = \frac{1}{4\pi} \int_{-1}^{1} d\mu_1 \int_{-1}^{1} d\mu_2 B^0_{\ell}(k_1, k_2, k_3)$$

$$= \frac{1}{4\pi} \int_{-1}^{1} d\mu_1 \int_{0}^{2\pi} d\phi B^0_{\ell}(k_1, k_2, k_3),$$

(7)

where $\mu_1$ is the cosine of the angle between the $k_1$ vector and the line of sight. The angle $\phi$ is defined as $\mu_2 = \mu_1 x_{12} - \sqrt{1 - \mu_1^2} \sqrt{1 - x_{12}^2} \cos \phi$ and $x_{12}$ is the cosine of the angle between $k_1$ and $k_2$. More details are given in Appendix A.

### 3.4 Analytical expression for $B^0_{\ell}$ covariance matrix

In order to apply the compression methods presented in Paper I we need an analytical expression for the bispectrum monopole covariance matrix. This allows us to compress a data-vector with an arbitrarily large number of triangle configurations, which on the contrary wouldn’t be possible using a covariance matrix estimated from the galaxy mock catalogs. That is because in order to obtain an accurate numerical estimate of the covariance matrix, the number of simulations used must be much greater than the data-vector’s dimension (Hartlap et al. 2007; Percival et al. 2014). As it has been shown in Paper I, compressing the power spectrum together with the bispectrum, or leaving it uncompressed, does not make any substantial difference in terms of recovered parameter constraints. However, it makes a huge difference in terms of complexity of the covariance matrix that one has to model analytically in order to compress the data-vector. Compressing the power spectrum as well (monopole and quadrupole) also requires modelling their covariance matrices together with the cross-covariance with the bispectrum monopole. Leaving them uncompressed just requires to model the bispectrum monopole covariance matrix. The expression for the Gaussian term of $C^{B^0_{\ell}B^0_{\ell}}$ is derived in Appendix A3 and reads:

$$C^{B^0_{\ell}B^0_{\ell}}(k_1, k_2, k_3; k_4, k_5, k_6) =$$

$$= \frac{D_{123456} V_e}{16\pi^2 N_t(k_1, k_2, k_3)} P^0_{\ell}(k_1) P^0_{\ell}(k_2) P^0_{\ell}(k_3),$$

(8)

where $D_{123456}$ stands for all the possible permutations for which each side of the first triangle is equal to a side of the second one; it has the values (6,2,1) respectively for equilateral, isosceles and scalene triangles, $N_t(k_1, k_2, k_3)$ is the number of independent triplets of grid points in the integration volume in Fourier space $V_e k_{123} = 8\pi^2 k_1 k_2 k_3 \Delta k_1 \Delta k_2 \Delta k_3$. For the values of the effective survey volume and the average galaxy density number used in computing the analytical covariance matrix, we adopt the values $V_e = 2.43 \times 10^9 \text{Mpc}^3$ and $h_0 = 1.14 \times 10^{-3} \text{Mpc}^{-3}$ used by Slepian et al. (2017a) for both power spectrum monopole/quadrupole and bispectrum monopole analytical covariance matrices. In practice we use the analytic expression of the covariance matrix only to determine the weights for the compression. Since all the terms considered scale as $V_e^{-1}$ the effective volume acts only as a scaling factor not affecting the compression performance.

In order to describe the correlation between different triangles in our analytical model of the covariance matrix, we include also a non-Gaussian term of the bispectrum monopole covariance matrix. In the expansion of the bispectrum covariance matrix presented in the Appendix of Paper I, for the bispectrum monopole this corresponds to a term proportional to the product of two bispectra monopoles as shown in Appendix A4:

$$C^{B^0_{\ell}B^0_{\ell}}_{\mathrm{NG}}(k_1, k_2, k_3; k_4, k_5, k_6) =$$

$$= C^{B^0_{\ell}B^0_{\ell}} - \frac{k^3}{16\pi^2 4\pi k^2 \Delta k} B^0_{\ell}(k_1, k_2, k_3) B^0_{\ell}(k_3, k_5, k_6) + 8 \text{ perm.}$$

(9)

It is important to include a term modelling the correlation between different triangles since the number of possible configurations increases very quickly as the bin size decreases. We do not include a corresponding non-Gaussian term into the power spectrum monopole and quadrupole covariances, since the number of data points considered is relatively low, thus the separation between the $k$ modules values is more than sufficient to assume that the correlation between the different modes $k_1$ and $k_2$ is negligible with respect to their variance (approximated by the Gaussian term on the diagonal of the covariance matrix).

### 3.5 Analytical expression for $[B^{(0,2)}_{\ell}, B^{(0)}_{\ell}]$ cross-covariance matrix

Finally we also model the cross-covariance between power spectrum multipoles and bispectrum monopole as described in Appendix A5:
As done in Paper I, we made the assumption that the shot noise is well approximated by a Gaussian distribution (which is reasonable if the galaxy number density is fairly high). Therefore, we just modify the galaxy power spectrum expressions by adding a $\nu^{-1}$ term. We did not take into account the effect of the survey geometry in the theoretical covariance matrix expression, which would affect the large scales inducing an extra correlation among the modes. We leave the inclusion of this correction for future work. Please refer to Howlett & Percival (2017) for a more detailed study on how to include this effect in the covariance matrix.

4 COMPRESSION METHODS

In Paper I we presented two compression methods and applied them to the galaxy bispectrum and power spectrum: MC-KL and PCA + KL. Both methods rely on the Karhunen-Loève algorithm (KL) applied for the first time for multi-parameter inference in cosmology by Tegmark et al. (1997). Using this KL compression it is possible to shrink an arbitrarily large data-vector $x$ to a compressed one $y$ having dimension equal to the number of model parameters considered preserving Fisher information. This is obtained by deriving a set of weights for the full data-vector for each model parameter. Taking the scalar product between the weighting vectors and the original full data-vector $x$ gives the elements $y_i$ of the compressed data-vector. Here we report only the main equations, please refer to Paper I for more details. The weighting vector for each parameter $\theta_i$ is given by:

$$b = \text{Cov}^{-1}(\langle x \rangle)_{i} \, ,$$

where $\text{Cov}^{-1}$ is the inverse of the original full data-vector covariance matrix and $\langle x \rangle_{i}$ is the derivative with respect to the model parameter $\theta_i$ of the mean of the modelled data-vector $x_i$, computed at a fiducial parameter vector $\theta_{\text{fid}}$. In our case the fiducial values are reported in Section 2.2. Therefore, the elements of the compressed data-vector $y$ are given by:

$$y_i = \langle x \rangle_{i}^\top \text{Cov}^{-1} x = b_i^\top x .$$

In the MC-KL method a MCMC sampling algorithm using $y$ as data-vector is run after compression. An estimate of the compressed covariance matrix from the mock catalogues can be obtained as shown in the Appendix of Paper I:

$$\text{Cov}_{y_{i,j}} = \text{Cov} \left[ y_i, y_j \right] = b_i^\top \cdot \text{Cov}_x \cdot b_j ,$$

where $\text{Cov}_x$ is the original covariance matrix.

4.1 PCA + KL

As described in Paper I, instead of orthogonalising the weights as in Zablocki & Dodelson (2016), we perform a principal component analysis (PCA) transformation of our parameter space before applying the KL compression. This is done by diagonalising the Fisher information matrix using the eigenvalue decomposition

$$F_{\theta_{\text{phys}}} = P \, F_{\theta_{\text{PCA}}} \, P^\top \, \text{ where } \, F_{\theta_{\text{PCA}}} = P^\top \, F_{\theta_{\text{phys}}} .$$

and $P$ is the linear transformation matrix. After having diagonalised the Fisher matrix we compress the data-vector with respect to this new set of parameters $\theta_{\text{PCA}}$. The effect of a PCA decomposition is to rotate the parameter space to the axes corresponding to the degeneracies between the original set of parameters. Therefore, taking the outer product of the 1D posteriors of the parameters $\theta_{\text{PCA}}$ in order to get the multidimensional posterior distribution should return a good approximation to the one sampled by the MCMC code.

Since the $\theta_{\text{PCA}}$ are uncorrelated, one can randomly sample the 1D posteriors and then rotate the resulting parameter vector using $P$ back into the physical space. Doing this avoids the use of the MCMC sampling altogether.

As shown in Paper I, this works only for those parameter sets which have a sufficiently low degree of degeneracy such that the approximation of Gaussianity for the multidimensional posterior can be assumed to be valid (no or very weak "banana-shaped" contours). Since this is not always the case, as for our choice of parameters, an additional Gaussianisation pre-step is required.

4.2 Gaussianisation pre-step

In Paper I the PCA + KL method assumed that it was possible to rotate through a linear transformation the physical parameter space into a new one where the new parameters are orthogonal/uncorrelated between each other. In order to be able to deal with distributions containing non-linear degeneracies (e.g. "banana-shaped" contours), we add a pre-Gaussianisation transformation of the parameter space using the procedure described in Schuhmann et al. (2016). In their work they introduced an extension of the Box-Cox transformations, which are functions of two parameters $(\alpha, \lambda)$:

$$\tilde{\theta}^t = BC((\alpha, \lambda))(\theta^t) = \begin{cases} (\lambda^{-1}[(\theta^t + a)^\lambda - 1] & (\lambda \neq 0) \\ \log(\theta^t + a) & (\lambda = 0) \end{cases}$$

where $\tilde{\theta}^t$ is the transformed $i$-th model parameter while $\theta^t$ is the original $i$-th model parameter. Their method was labelled Arcsinh-Box-Cox transformation (ABC). For each of the model parameters, a set of three ABC transformation parameters $(\alpha, \lambda, t)$ are computed by the algorithm which are then used in the following way:

$$\tilde{\theta}_{\text{Gauss.}}^t = ABC(\theta^t_{\text{phys.}}) = \begin{cases} r^{-1} \sinh[r \, BC((\alpha, \lambda))(\theta^t_{\text{phys.}})] & (t > 0) \\ BC((\alpha, \lambda))(\theta^t_{\text{phys.}}) & (t = 0) \\ r^{-1} \arcsinh[r \, BC((\alpha, \lambda))(\theta^t_{\text{phys.}})] & (t < 0) \end{cases}$$

where $\theta^t_{\text{Gauss.}}$ is the Gaussianised $i$-th model parameter while $\theta^t_{\text{phys.}}$ is the original $i$-th physical model parameter. We then relabel this compression as G-PCA. In order to obtain the transformation parameters of the Gaussianising transformations it is necessary to run a preliminary MCMC sampling using the full data-vector.
we want to prove is that once the transformation parameters have been obtained for the standard number of triangles corresponding to the $\Delta k_6$ binning case, these are valid also for a higher number of triangle configurations included in the bispectrum.

4.3 Analytical covariance matrix: usage

In the following analysis, we are going to use two different options for the analytical covariance matrices. For the MC-KL method we compress only the bispectrum monopole part of the data-vector. To derive the weights in Eq. 11 we use the analytical covariance matrix of the bispectrum monopole given by the sum of the Gaussian term in Eq. 8 and the non-Gaussian one given in Eq. 9. For the G-PCA method the full data-vector needs to be compressed since the computation of the 1D posteriors of the $\theta_{\text{PCA}}$ parameters requires each data vector element to be sensitive to the variation of just one $\theta_{\text{PCA}}$ parameter, as explained in Paper I. Therefore, for the power spectrum monopole/quadrupole we use Eq. 5 as our analytical covariance matrix; similarly for the bispectrum monopole we use Eq. 8 for the covariance matrix (the same as the one we used for the MC-KL case), and finally, we use Eq. 10 for our cross-covariance matrix.

5 RECOVER MCMC-DERIVED POSTERIOR DISTRIBUTION

For MCMC sampling we use emcee (Foreman-Mackey et al. 2013). All the likelihoods have been corrected as suggested by Sellentin & Heavens (2016) in order to take into account the bias induced by estimating the inverse of the real covariance matrix from a limited number of mocks. In order to check whether our analytical estimate of the covariance matrix is good enough to be used for deriving the weights as explained in Sec. 4, we compare to the full MCMC 1D posterior distributions in the left panels of Figures 1 and 2 with results from the MCMC+ MC-KL and G-PCA methods, respectively.

The violin plots include the standard binning case $\Delta k_6$ (116 triangles) and the $\Delta k_5$ case (195 triangles). For these two cases we compare the MCMC (grey and purple) with the compression results (cyan and orange). From each point we subtract the mean...
of the model parameters obtained using the MCMC. This makes it easier to check that the shift in the mean of the compression results with respect to the MCMC ones is small when compared to the size of the inner quartiles of the distribution. This concept is also quantified in the bottom half of Table 1, which shows the shifts in the mean values relative to the 1D 68% credible intervals. In the top half of Table 1 we report the precise values of both the means and the 68% credible intervals for all model parameters. Additionally, Figure B1 in Appendix B shows the comparison between the 2D MCMC posterior distributions and the MC-KL and G-PCA ones for both $\Delta k_6$ and $\Delta k_5$ cases. We conclude that even if a small part of the constraining power is lost (see the $\Delta k_6$ columns in Table 2 for details), both compression methods return posterior distributions which well agree with the MCMC distribution for all model parameters under consideration.

6 INFORMATION CONTENT AND NUMBER OF TRIANGLES

The right panels of Figures 1 and 2 show how using a larger number of triangles tightens the posterior contours of the four model parameters considered. At the same time, the maxima of the 2D posterior distributions converge to the same values for each compression method as the number of triangles is increased.

Note that the shift in the posterior distribution between binning cases is not an artifact of the compression: it is also present when we fit using the standard MCMC method. This can be seen when comparing the location and shape of the 2D contour regions in Figures B1 and B2 in Appendix B for the $\Delta k_6$ and $\Delta k_5$ binning cases. Quantitatively it can be observed by comparing means and standard deviations in Table 1. Thus, both compression algorithms reproduce posterior distributions very similar to the ones derived via MCMC sampling for the relevant binning cases $\Delta k_6$ and $\Delta k_5$. The observed shift between binning cases is due to the strong degeneracy between the model parameters. In particular the shift happens along the degeneration direction of $b_1$, $b_2$ and $f$ with $\sigma_8$. It may have a statistical origin. Further checks on this effect may be performed using the galaxy mocks, for example by fitting several different realizations for both the $\Delta k_6$ and $\Delta k_5$ binning cases using the G-PCA method (which would be much faster than doing parameter estimation via MCMC or MC-KL). We reserve to do these tests in future work. The main result of this paper is that the variance of the parameters is reduced when the number of triangles used increases.

For future surveys the compression can be then used for the main analysis and also to find out the minimum number of triangle configurations for a given $k$-range needed to fully capture the non-Gaussian information contained in 3pt statistics like the bispectrum. The later will indicate how many mock catalogues/simulations are required in order to accurately estimate the covariance matrix. In our analysis the saturation seems to be reached already for the $\Delta k_4$ binning case (404 triangles).

For what concerns $\Delta k_2$, the smallest $k$-bin size considered (2734 triangles), Tables 1 and 2 show that the $\Delta k_2$ posterior distribution is very similar to the $\Delta k_4$ case.

The trend in the information content in terms of the 1D 68% credible intervals as a function of the triangle number used is shown in the left panel of Figure 3, and the improvement quantified in Table 2. From Figure 3 it appears that the parameters constraints improvement as a function of the number of triangles reaches the saturation already for the $\Delta k_4$ case. In terms of percent-

---

**Figure 2.** Joint data-vector $[P_{\ell}^{(0)}, P_{\ell}^{(2)}, B_{\ell}^{(0)}] \text{ posteriors: G-PCA four-parameters case. Same as Figure 1 but for the G-PCA method.}
the improvements achieved via compression for the 1D MCMC on the full data-vector for the parameters considered set of parameters (e.g. left panel of Figure 3). Finally in blue and red are highlighted the improvements obtained via compression for the MCMC columns to the compression ones, it is that the difference between the mean parameter values obtained via MCMC and the ones obtained via compression (MC-KL or G-PCA) are evidently within the 68% credible intervals given by the MCMC on the full data-vector.

Table 1. Four parameter-case, check consistency.

Upper half: ratios is increased is simply due to the strong degeneracy present between... and corresponding 1σ credible intervals obtained via MCMC sampling is shown. Comparing the MCMC columns to the compression ones, it is that the difference between the mean parameter values obtained via MCMC and the ones obtained via compression (MC-KL or G-PCA) are evidently within the 68% credible intervals given by the MCMC on the full data-vector.

| Δk6 | MCMC | MC-KL | G-PCA |
|-----|------|-------|-------|
| b1  | 2.41 ± 0.22 | 2.41 ± 0.23 | 2.49 ± 0.27 |
| b2  | 1.00 ± 0.40 | 1.04 ± 0.42 | 1.08 ± 0.47 |
| f   | 0.69 ± 0.08 | 0.72 ± 0.09 | 0.72 ± 0.09 |
| σ8  | 0.50 ± 0.04 | 0.48 ± 0.05 | 0.48 ± 0.05 |

Table 2. Four-parameter case, constraints improvement. Below are shown the relative variations in percentage of the size of the 68% credible intervals as a function of the k-binning considered (number of triangle configurations used for the bispectrum monopole). In orange and green are highlighted respectively the improvements achieved via compression for the Δk6 and at the saturation level (404 triangles - Δk5) of the bispectrum monopole constraining power case for the considered set of parameters (e.g. left panel of Figure 3). Finally in blue and red are highlighted the improvements obtained via compression for the highest number of triangles considered (2734 triangles - Δk2 binning) for MC-KL and G-PCA respectively.

| Δk5 | MCMC | MC-KL | G-PCA |
|-----|------|-------|-------|
| b1  | 9.2  | -0.3  | 3.3  |
| b2  | 40.3 | 3.5   | 7.5  |
| f   | 12.1 | 4.4   | 4.4  |
| σ8  | 8.5  | -5.1  | -5.5 |

| Δk3 | MCMC | MC-KL | G-PCA |
|-----|------|-------|-------|
| b1  | 0.22 | 4.4   | 18.8 |
| b2  | 0.40 | 2.9   | 16.2 |
| f   | 0.08 | 3.7   | 7.0  |
| σ8  | 0.04 | 6.5   | 10.0 |

ages of the original 1D 68% credible intervals obtained running an MCMC on the full data-vector for the parameters (b1, b2, f, σ8) in the BOSS Δk6 case, the Δk2 MC-KL and G-PCA analyses obtain tighter constraints by (−35%, −45.3%, −22.6%, −22.6%) and (−24.8%, −52.8%, −26.4%, −21%), respectively. These optimal constraints as obtained by the compression methods are also shown in summary Figure 5.

7 CONSISTENCY CHECK

In order to test the validity of our analysis, we compute the reduced χ² and corresponding p-value for each set of parameters obtained using either the MCMC sampling or the compression methods. For all parameter vectors (compressed and uncompressed) this has been done using the data-vector corresponding to the strong degeneracy present between b1, b2, f and σ8. Indeed both the reduced χ² and p-values show that all these models fit the data very well. In Figure 4 we did not show the lines and statistics for the Δk3 cases just for the sake of clarity and because the results are equivalent to those of the other binnings. From the same figure it can also be noticed that the tightest errorbars are those from the power spectrum case.

To demonstrate the flexibility of the compression methods we check their performance when the fiducial parameter set is shifted by ±1σ credible intervals in the Δk6 case. The effect of this is shown in the right panel of Figure 3. For this plot, we centre each 1D distribution by subtracting the mean obtained by running the compression pipelines using the fiducial parameter values. In this way it is possible to observe by how much the posterior distributions derived via MC-KL or G-PCA shift as a function of the chosen fiducial parameter set. In Appendix B the precise numbers are reported in Table B1.

MC-KL appears to be more stable than the G-PCA when the fiducial parameter set is shifted. The explanation of this could be the
fact that G-PCA involves several transformations of the parameter space, including a diagonalisation of the Fisher information matrix which is computed from the analytical model of the covariance matrix.

Nevertheless, it should be noted that we are testing the performances of the compression in a regime of strong degeneracy of the parameter space and therefore shifting the fiducial parameter set by $\pm 1\sigma$ credible intervals actually means increasing/reducing the individual values by $\sim 10-40\%$ (second panel Table 1). Therefore, running a preliminary low-resolution MCMC sampling on the full data-vector (which can be shorter than the one that will be later compressed, as we have done in our analysis) is an efficient solution to determining a reasonable fiducial model for deriving the compression.

7.1 Comparison with BOSS DR12 bias constraints

BOSS galaxy sample results from the bispectrum are reported by Gil-Marín et al. (2017) [in Table 3 at p. 18] from the same CMASS sample data set, at the same redshift, for the following parameter combinations: $b_1\sigma_8 = 1.2479 \pm 0.0072$, $b_2\sigma_8 = 0.641 \pm 0.066$ and $f\sigma_8 = 0.432 \pm 0.018$. If we recast our results obtained using the MCMC for the $\Delta k_8$ case in terms of the same parameter combinations these are: $b_1\sigma_8 = 1.203 \pm 0.008$, $b_2\sigma_8 = 0.557 \pm 0.140$ and $f\sigma_8 = 0.339 \pm 0.019$.

In the BOSS analysis a larger range of scales has been considered. In particular, BOSS analysis goes up to $k \sim 0.2 h/$Mpc for both power spectrum monopole/quadrupole and bispectrum monopole while we stop at $k \sim 0.09 h/$Mpc and $k \sim 0.12 h/$Mpc, respectively. This could explain the larger value we obtained for $b_2\sigma_8$. A more complex model for the power spectrum was used in the BOSS analysis, including loop corrections beyond the tree level approximation. Moreover the BOSS analysis also modelled the effect of the survey window function for both power spectrum and bispectrum.

As we saw from Figure 4, the power spectrum monopole is the most constraining part of the full data-vector, having errorbars of less than 5%. Moreover, in the BOSS analysis the FoG parameters $\sigma^B_{\text{FoG}}$ and $\sigma^P_{\text{FoG}}$ were left free to vary in order to better model the non-linear regime and were detected with high significance ($\sigma^B_{\text{FoG}} = 7.54 \pm 0.70$ and $\sigma^P_{\text{FoG}} = 3.50 \pm 0.14$). The BOSS model also included a noise-amplitude parameter $A_{\text{noise}}$ which modelled divergence from Poissonian shot noise. In our model we had included

---

9 we compare our results with the BOSS analysis standard deviation values obtained considering only the statistical contributions and not the systematics ones.
Compression on BOSS measurements

Figure 4. Reduced $\chi^2$ and $p$-values for the best-fit models obtained using the MCMC, MC-KL and G-PCA compression methods. The $k$-binnings shown are respectively the standard $\Delta k_6$ (navy), an intermediate size $\Delta k_4$ (green) and the smallest one $\Delta k_2$ (pink for MC-KL and red for G-PCA) corresponding to the highest number of triangle used in the bispectrum monopole. The two upper panels are for the power spectrum monopole (left) and quadrupole (right) while the bottom panel refers to the bispectrum monopole. The lower part of each panel shows the relative difference between the data measurements and the different models. Even if for example $b_1$ and $\sigma_8$ values are shifted between the cases of $\Delta k_6$ and $\Delta k_2$, the strong degeneracy has the result of making the two models practically identical.

We used CAMB (Lewis et al. 2000) to compute the linear matter power spectrum. The time difference between MCMC/MC-KL and G-PCA would have been much more significant in the case of a parameter set for which the linear matter power spectrum needs to be recomputed for every model realisation.

8 CONCLUSIONS

In this paper we have shown the results of applying both compression methods for the galaxy redshift-space bispectrum, presented in Paper I, to the measurements from the SDSS-III BOSS DR12 CMASS sample (Gil-Marín et al. 2017). We considered as original data-vector the combination of the power spectrum monopole and quadrupole with the bispectrum monopole, which are obtained by averaging over the angles describing the orientation with respect to the line of sight. The first method called MC-KL consists of running an MCMC sampling on the compressed data-vector obtained by taking the scalar product between the original data-vector and a set of weights derived as first shown by Tegmark et al. (1997). The second method, which we denoted as G-PCA, is the modification of the PCA + KL method presented in Paper I obtained by adding a Gaussianisation transformation of the parameter set (Schuhmann...
et al. 2016) before rotating it using a principal component analysis transformation (PCA) followed by the KL compression. By transforming the physical parameter space into an orthogonal one it is possible to just randomly sample 1D posterior distributions, avoid altogether the need of running a MCMC routine.

In order to derive the posterior distributions for the set of parameter considered, the galaxy bias parameters $b_1$ and $b_2$, the growth rate $f$ and the normalisation of the dark matter perturbations amplitude $\sigma_8$, we numerically estimated the covariance matrix using 1400 and 700 galaxy mocks catalogues for the full data-vector and compressed data-vector cases, respectively.

The following points represent the main conclusions of our analysis:

• In order to obtain the weights for the compression methods we derived an analytic approximation of the leading terms of the covariance matrix relative to the considered data-vector. The final
expressions of these computations are reported in Sec. 3 while the full derivations are shown in Appendix A.

- In Sec. 5 we have shown that both compression methods recover the posterior distributions obtained via MCMC using the full data-vector with little loss of information (≈ 4% and ≈ 13% larger 68% credible intervals than the MCMC ones in average for MC-KL and G-PCA, respectively). More importantly, even if slightly broader, the posterior distributions recovered through compression have the same shape and modes as the MCMC counterparts.

- Adding a pre-Gaussianisation step removes the PCA + KL limitation linked to a strongly degenerate parameter space described in Paper I. It is however necessary to run a preliminary MCMC in order to derive the Gaussianisation transformation parameters. Nevertheless, once these parameters have been derived for a number of triangles case for which it is possible to run an MCMC on the full data-vector, they can be used to compress a data-vector with an arbitrary number of triangles.

- In Sec. 6 we show the main result of this work, namely the substantial improvement in parameter constraints obtained by compressing a much larger number of triangles with respect to standard MCMC data-vector. For the uncompressed data-vector the number of triangles is limited by the number of mock catalogues available to estimate the covariance matrix. For both compression methods and for any number of triangle configuration considered, the dimension of the compressed data-vector is always equal to the number of model parameters constrained.

For the highest number of triangles considered, this leads to an improvement in terms of the 68% 1D credible intervals by (∼35%, ∼45%, ∼23%, ∼23%) and (∼25%, ∼53%, ∼26%, ∼21%) for the MC-KL and G-PCA methods, respectively.

- By way of summary, in Figure 5 we show the results for both MC-KL and G-PCA methods using 2734 triangles and for the MCMC on the uncompressed data-vector containing 116 triangles. The two compression methods agree well and produce substantially tighter and less degenerate constraints. Furthermore the G-PCA approach allowed for a computational speed up, requiring only approximately a third of the time taken by the MCMC and MC-KL methods, including also the low-resolution MCMC necessary for the Gaussianisation transformation. Considering only the PCA part, the speed up factor rises to ∼ 20 – 100 times depending on the parameter set considered.

- Finally we would like to point out that the compressing methods used in this work represents a straightforward approach to include higher order statistics like the trispectrum or the tetraspectrum in the analysis of current and future data sets. This is due to the fact that the number of elements of the data-vector, after the maximal compression, corresponds exactly to the number of model parameters. Both MC-KL and G-PCA have the potential to fully exploit the constraining power of higher order statistics applied to data-sets from future surveys like DESI, EUCLID and PFS.

ACKNOWLEDGEMENTS

D.G. is grateful to L. Whiteway for the useful discussions and to E. Edmondson for the technical help in using the UCL computer cluster. D.G. is supported by the Perren and the IMPACT studentships. HGM is supported by Labex ILP (reference ANR-10-LABX-63) part of the Idex SUPER, and received financial state aid managed by the Agence Nationale de la Recherche, as part of the programme Investissements d’avenir under the reference ANR-11-IDEX-0004-02 e.M.M. acknowledges funding from STFC Consolidated Grants RG84196 and RG70655 LEAG/506 and has received funding from the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 6655919. O.L. acknowledges support from a European Research Council Advanced Grant FP7/291329. C (Kernighan 1988) and python 2.7 (Rossum 1995) have been used together with many packages like IPython (Perez & Granger 2007), Numpy (van der Walt et al. 2011), Scipy (Jones et al. 2001) and Matplotlib (Hunter 2007). The corner plots have been realised using rvtool developed by Bocquet & Carter (2016) while the violin plots have been created using seaborn by Waskom et al. (2016).

References

Ade P. A. R., et al., 2014, Astron. Astrophys., 571, A16
Alam S., et al., 2017, MNRAS, 470, 2617
Alsing J., Wandelt B., 2018, MNRAS, 476, L60
Alsing J., Wandelt B., Fenney S., 2018, MNRAS, 477, 2514
Bernardeau F., Brax P., 2011, JCAP, 1106, 019
Bocquet S., Carter F. W., 2016, The Journal of Open Source Software, 1
Borisson A., Jain B., 2009, Phys. Rev. D, 79, 103506
Dark Energy Survey Collaboration et al., 2016, MNRAS, 460, 1270
Dawson K. S., et al., 2013, AJ, 145, 10
Eisenstein D. J., et al., 2011, AJ, 142, 72
Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Frieman J. A., Gaztanaga E., 1994, ApJ, 425, 392
Fry J. N., 1984, ApJ, 279, 385
Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447
Fry J. N., Scherrer R. J., 1994, ApJ, 429, 36
Fu L., et al., 2014, MNRAS, 441, 2725
Gangui A., Lucchin F., Matarrese S., Mollerach S., 1994, ApJ, 430, 447
Gil-Marín H., Schmidt F., Hu W., Jimenez R., Verde L., 2011, J. Cosmology Astropart. Phys., 11, 019
Gil-Marín H., Wagner C., Noreña J., Verde L., Percival W., 2014, J. Cosmology Astropart. Phys., 12, 029
Gil-Marín H., et al., 2016a, MNRAS, 460, 4188
Gil-Marín H., et al., 2016b, MNRAS, 460, 4210
Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, MNRAS, 465, 1575
Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385
Gualdi D., Manera M., Joachimi B., Lahav O., 2018, MNRAS, 480, 3246
Harlét J., Simon P., Schneider P., 2007, A&A, 464, 399
Heavens A. F., Matarrese S., Verde L., 1998, Mon. Not. Roy. Astron. Soc., 301, 797
Heavens A. F., Jimenez R., Lahav O., 2000, MNRAS, 317, 965
Heavens A. F., Sellentin E., de Mijolla D., Vianello A., 2017, MNRAS, 472, 4244
Hoffmann K., Mao Y., Mo H., Wandelt B. D., 2018, preprint, (arXiv:1802.02578)
Howlett C., Percival W. J., 2017, MNRAS, 472, 4935
Hunter J. D., 2007, Computing in Science and Engg., 9, 90
Jarvis M., Bernstein G., Jain B., 2004, MNRAS, 352, 338
Joachimi B., Shi X., Schneider P., 2009, A&A, 508, 1193
Jones E., Oliphant T., Peterson P., et al., 2001–, SciPy: Open source scientific tools for Python, http://www.scipy.org/
Kayo I., Takada M., 2013, preprint, (arXiv:1306.4684)
APPENDIX A: ESTIMATORS AND COVARIANCE TERMS

A1 Power spectrum monopole/quadrupole and bispectrum monopole estimators

The analytical model for the redshift-space galaxy power spectrum monopole and quadrupole is given by equation 4. It is therefore natural to define the estimator as:

\[ \hat{P}_g^{(\ell)}(k) = \left( \frac{2\ell + 1}{2} \right) \frac{1}{(2\pi)^3 N_p(k)} \int_{V_p} \int_{V_q} d^3 p d^3 q L_\ell(\mu) \delta_D(q + p) \delta_g(q) \delta_g(p), \]  

(A1)

where \( V_p, q \) are the spherical shell volumes characterised by \( k - \Delta k/2 \leq q, p \leq k + \Delta k/2 \). \( \mu \) is the cosine of the angle with respect to the line of sight of the \( q \) wave vector and \( L_\ell(\mu) \) is the Legendre polynomial of order \( \ell \). \( \delta_D \) is the 3-D Dirac delta. \( N_p \) is the number of grid point pairs in the integration volume in Fourier space and can be computed as:

\[ \frac{V_p}{k_f^3} = \frac{k_f^3}{V_p} \int_{V_p} \int_{V_q} d^3 p d^3 q \delta_D(q + p) \approx \frac{4\pi k^2 \Delta k}{k_f^3}, \]  

(A2)

where \( V_p \approx 4\pi k^2 \Delta k \) is the spherical integration shell defined by \( k - \Delta k/2 \leq q, p \leq k + \Delta k/2 \) as defined in Scoccimarro et al. (1998b). \( k_f \) is the fundamental frequency defined in terms of the survey volume \( V_c \) as \( k_f^3 = \frac{(2\pi)^3}{V_c} \). We check that the estimator defined in Eq. A1 is unbiased:

\[ \langle \hat{P}_g^{(\ell)}(k) \rangle = \left( \frac{2\ell + 1}{2} \right) \frac{1}{(2\pi)^3 N_p(k)} \int_{V_p} \int_{V_q} d^3 p d^3 q L_\ell(\mu) \delta_D(q + p) \langle \delta_g(q) \delta_g(p) \rangle \]  

\[ = \left( \frac{2\ell + 1}{2} \right) \frac{1}{(2\pi)^3 N_p(k)} \int_{V_p} \int_{V_q} d^3 p d^3 q L_\ell(\mu) \delta_D(q + p)^2 (2\pi)^3 P_g^0(p) \]  

\[ = \left( \frac{2\ell + 1}{2} \right) \frac{1}{V_c V_k} \int_{V_p} \int_{V_q} d^3 p d^3 q L_\ell(\mu) \delta_D(q + p) V_c P_g^0(p) \]  

\[ = \left( \frac{2\ell + 1}{2} \right) \frac{1}{V_k} \int_{V_p} \int_{V_q} d^3 p d^3 q L_\ell(\mu) \delta_D(q + p) \frac{P_g^0(p)}{V_c} \]  

\[ \approx \left( \frac{2\ell + 1}{2} \right) \int_{-1}^{+1} d\mu P_g^0(k, \mu) L_\ell(\mu), \]  

(A3)

where we used the approximation made in Joachimi et al. (2009) that \( \delta_D^2 \approx \frac{V_c}{(2\pi)^3} \delta_D = k_f^3 \delta_D \). In the last step it has been made the common approximation that \( p \) and \( q \) are very close to \( k \) in module for thin enough shells (small \( \Delta k \)). The standard definition of the redshift galaxy power spectrum has also been used:

\[ \langle \delta_g(q) \delta_g(p) \rangle = \langle 2\pi \rangle^3 \delta_D(q + p) P_g^0(p) \]  

(A4)

The redshift space galaxy bispectrum is defined as:

\[ \langle \delta_g(q_1) \delta_g(q_2) \delta_g(q_3) \rangle = \langle 2\pi \rangle^3 \delta_D(q_1 + q_2 + q_3) B_g^0(q_1, q_2, q_3) \]  

(A5)

The analytical expression for the bispectrum monopole model was given in Eq. 7.

Analogously to the power spectrum multipoles, the estimator for the bispectrum monopole can be defined as:

\[ \hat{B}_g^{(0)}(k_1, k_2, k_3) = \frac{1}{4\pi N_i(2\pi)^6} \int_{V_1} \int_{V_2} \int_{V_3} d^3 q_1 d^3 q_2 d^3 q_3 \delta_D(q_1 + q_2 + q_3) \delta_g(q_1) \delta_g(q_2) \delta_g(q_3) \]  

(A6)

where \( N_i(k_1, k_2, k_3) \) is the number of independent grid points triplets inside the integration volume in Fourier space. As shown in the weak lensing 2D case by Kayo et al. (2013), this is computed as:

\[ N_i(k_1, k_2, k_3) = \frac{V_k^{(3)}}{k_f^6} = k_f^6 \int_{V_1} \int_{V_2} \int_{V_3} d^3 q_1 d^3 q_2 d^3 q_3 \delta_D(q_1 + q_2 + q_3) \approx \frac{8\pi^2 k_1^2 k_2^2 k_3^2 \Delta k_1 \Delta k_2 \Delta k_3}{k_f^6} \]  

(A7)

It is important to notice that the result of the above integral must be symmetric in the \( k \)-vectors arguments. Therefore, the best way to derive the integral results is through geometrical considerations. Starting from \( q_1 \), this can be chosen in a spherical shell with volume \( V_{k_1} \approx 4\pi k_1^2 \Delta k \).
Compression on BOSS measurements

Figure A1. Computation of the integration volume in Fourier space in the case of the bispectrum monopole. Once the side $k_1$ of the triangle is fixed, the other two sides are free to vary in the intersection given by two sphere of radius $k_2 - \Delta k_2/2 \leq r_2 \leq k_2 + \Delta k_2/2$ and $k_3 - \Delta k_3/2 \leq r_3 \leq k_3 + \Delta k_3/2$ respectively. In the Figure above the 2D projection of the annuli of thickness $\Delta k_2$ (blue) and $\Delta k_2$ (red) are shown. The angle $\phi$ correspond to the angle $\phi_{12}$ in the text.

Once $q_1$ is fixed, considering the plane in which both $q_2$ and $q_3$ lie, they must connect to each other inside the 2D intersection formed by the two annuli defined by $k_2 - \Delta k_2/2 \leq q_2 \leq k_2 + \Delta k_2/2$ and $k_3 - \Delta k_3/2 \leq q_3 \leq k_3 + \Delta k_3/2$. This has approximately an area equal to $A_{k_{23}} = k_2 \Delta \phi_{12} \Delta k_2$. From Figure A1 it is possible to see that $\Delta \phi_{12}$ is defined by varying $k_3$ by $\Delta k_3$. $\phi_{12}$ can be obtained from:

$$\cos \phi_{12} = \frac{k_1^2 + k_2^2 - k_3^2}{2 k_1 k_2}, \quad (A8)$$

and therefore $\Delta \phi_{12}$ can be found differentiating with respect to $k_3$:

$$\frac{d \cos \phi_{12}}{dk_3} = -\frac{d \phi_{12}}{dk_3} \sin \phi_{12} = -\frac{k_3}{k_1 k_2} \quad \implies \quad \Delta \phi_{12} = \frac{\Delta k_3 k_3}{k_1 k_2} (\sin \phi_{12})^{-1}. \quad (A9)$$

Finally the volume of the intersection between $k_2$ and $k_3$ is obtained by rotating the area just found around the axis defined by $k_1$:

$$V_{k_{23}} = 2\pi A_{k_{23}} (k_2 \sin \phi_{12}), \quad (A10)$$

which allows to compute $V_{k_{123}} = V_k V_{k_{23}}$ in Eq. A7.

### A2 Power spectrum monopole and quadrupole covariance matrix: Gaussian term

Following the Appendix of Gualdi et al. (2018) we can check that also the bispectrum monopole estimator defined in Eq. A6 is unbiased. Moreover it is possible to compute the Gaussian term of the covariance for the power spectrum monopole and quadrupole as follows:
A3 Bispectrum monopole covariance matrix: Gaussian term

Analogously to the above we now compute the diagonal term of the bispectrum monopole covariance matrix:

$$C_{G}^{p_{G}p_{G}^{\prime}}(k_{1}, k_{2}; k_{3}, k_{4}, k_{5}, k_{6}) =$$

$$= \frac{1}{16\pi^{2}} \frac{(2\pi)^{-6}}{N_{G}(k_{1}, k_{2}, k_{3}) N_{G}(k_{4}, k_{5}, k_{6})} \prod_{i=1}^{6} \int_{V_{q_{i}}} d^{3} q_{i} \delta_D(q_{1} + q_{2} + q_{3}) \delta_D(q_{4} + q_{5} + q_{6}) \times (2\pi)^{9} \delta_D(q_{1} + q_{4} + q_{6}) \delta_D(q_{2} + q_{3} + q_{5}) \delta_D(q_{3} + q_{6} + q_{4}) P_{G}^{p_{G}}(q_{1}) P_{G}^{p_{G}}(q_{2}) P_{G}^{p_{G}}(q_{3}) + \text{perm.}$$

$$= \frac{D_{123456}}{16\pi^{2}} \frac{(2\pi)^{9} k_{f}^{-6}}{N_{G}(k_{1}, k_{2}, k_{3})} \prod_{i=1}^{3} \int_{V_{q_{i}}} d^{3} q_{i} \delta_D(q_{1} + q_{2} + q_{3})^{2} P_{G}^{p_{G}}(q_{1}) P_{G}^{p_{G}}(q_{2}) P_{G}^{p_{G}}(q_{3})$$

$$= \frac{V_{c} k_{f}^{-6}}{16\pi^{2}} \frac{\delta_{k}^{C}}{N_{G}(k_{1}, k_{2}, k_{3})^{2}} \prod_{i=1}^{3} \int_{V_{q_{i}}} d^{3} q_{i} \delta_D(q_{1} + q_{2} + q_{3})^{2} P_{G}^{p_{G}}(q_{1}) P_{G}^{p_{G}}(q_{2}) P_{G}^{p_{G}}(q_{3})$$

$$= \frac{D_{123456}}{16\pi^{2}} \frac{V_{c} k_{f}^{-6}}{N_{G}(k_{1}, k_{2}, k_{3})} \prod_{i=1}^{3} \int_{V_{q_{i}}} d^{3} q_{i} \delta_D(q_{1} + q_{2} + q_{3})^{2} P_{G}^{p_{G}}(q_{1}) P_{G}^{p_{G}}(q_{2}) P_{G}^{p_{G}}(q_{3})$$

$$= \frac{D_{123456}}{16\pi^{2}} \frac{V_{c} k_{f}^{-6}}{N_{G}(k_{1}, k_{2}, k_{3})} \prod_{i=1}^{3} \int_{V_{q_{i}}} d^{3} q_{i} \delta_D(q_{1} + q_{2} + q_{3})^{2} P_{G}^{p_{G}}(q_{1}) P_{G}^{p_{G}}(q_{2}) P_{G}^{p_{G}}(q_{3})$$

where $D_{123456}$ stands for all the possible permutations and has values 6, 2, 1 respectively for equilateral, isosceles and scalene triangles. Again it has been assumed that the power spectrum monopole does not vary significantly inside the integration volume.

A4 Bispectrum monopole covariance matrix: non-Gaussian term

In this work we use only one of the non-Gaussian terms of the bispectrum monopole covariance matrix. This is because we just need to model the covariance matrix analytically in order to derive the weights for the compression. This additional term allows to better capture the correlation between different triangles. We leave to future work the analytic computation of the remaining terms.
Compression on BOSS measurements

\[
C_{NG}^{B_B^2}(k_1, k_2, k_3; k_4, k_5, k_6) =
\]
\[
= \frac{1}{16\pi^2 N_i(k_1, k_2, k_3) N_i(k_4, k_5, k_6)} \int_{V_{q_1}} \int_{V_{q_2}} \int_{V_{q_3}} \int_{V_{q_4}} \int_{V_{q_5}} \int_{V_{q_6}} d^3 q_1 d^3 q_2 d^3 q_3 d^3 q_4 d^3 q_5 d^3 q_6 \delta_D(q_1 + q_2 + q_3) \delta_D(q_4 + q_5 + q_6) \]
\[
\times (2\pi)^6 \delta_D(q_1 + q_2 + q_3) \delta_D(q_3 + q_5 + q_6) B_B^0(q_1, q_2, q_4) B_B^0(q_3, q_5, q_6) + 8 \text{ perm.}
\]
\[
= \frac{1}{16\pi^2 N_i(k_1, k_2, k_3) N_i(k_4, k_5, k_6)} \int_{V_{q_1}} \int_{V_{q_2}} \int_{V_{q_3}} \int_{V_{q_4}} \int_{V_{q_5}} \int_{V_{q_6}} d^3 q_1 d^3 q_2 d^3 q_3 d^3 q_4 d^3 q_5 d^3 q_6 \delta_D(q_1 + q_2 + q_3)
\]
\[
\times \delta_D(q_3 + q_5 + q_6) B_B^0(q_1, q_2, -q_3) B_B^0(q_3, q_5, q_6) + 8 \text{ perm.}
\]
\[
= \frac{1}{16\pi^2 N_i(k_1, k_2, k_3) N_i(k_4, k_5, k_6)} \int_{V_{q_1}} \int_{V_{q_2}} \int_{V_{q_3}} \int_{V_{q_4}} \int_{V_{q_5}} \int_{V_{q_6}} d^3 q_1 d^3 q_2 d^3 q_3 d^3 q_4 d^3 q_5 d^3 q_6 \delta_D(q_1 + q_2 + q_3)
\]
\[
\times \delta_D(q_3 + q_5 + q_6) B_B^0(q_1, q_2, -q_3) B_B^0(q_3, q_5, q_6) + 8 \text{ perm.}
\]
\[
\approx \frac{1}{16\pi^2 N_i(k_1, k_2, k_3) N_i(k_4, k_5, k_6)} B_B^{(0)}(k_1, k_2, k_3) B_B^{(0)}(k_4, k_5, k_6) \int_{V_{q_1}} d^3 q_3 d^3 q_6 \delta_D(q_3 + q_5 + q_6) + 8 \text{ perm.}
\]
\[
= \frac{\delta^K_{34}}{16\pi^2 4\pi k_3^2 k_5^2} B_B^{(0)}(k_1, k_2, k_3) B_B^{(0)}(k_4, k_5, k_6) + 8 \text{ perm.},
\]
\[(A13)\]

where the usual approximations have been used together with Eq. A10 which in the last step has been used to simplify the integration over the volume in Fourier space once one of the \(k\)-vectors is fixed.

\section{A5 Cross-covariance term}

For what concerns the cross-covariance term between power spectrum (monopole/quadrupole) and bispectrum monopole we use only the first leading term in our model:

\[
C_{P_1}^{\ell_1 P_2} B_2(q_1, q_2, k_3; k_4, k_5, k_6) =
\]
\[
= \frac{1}{4\pi} \left(\frac{2\ell_1 + 1}{2}\right) \frac{(2\pi)^6 k_3^3}{N_i(k_1) N_i(k_2, k_3, k_4)} \int_{V_{q_1}} \int_{V_{q_2} \cup V_{q_3} \cup V_{q_4}} d^3 q_1 d^3 q_2 d^3 q_4 \int_{V_{q_5} \cup V_{q_6}} d^3 q_5 d^3 q_6 \delta_D(q_1 + q_2 + q_4) \delta_D(q_2 + q_3 + q_4) L_{\ell_1}(\mu_1)
\]
\[
\times 2(2\pi)^6 \delta_D(q_1 + q_2) \delta_D(p_1 + q_3 + q_4) P_B(q_2) B_B^0(q_2, q_3, q_4) + 2 \text{ perm.}
\]
\[
= \frac{1}{2\pi} \left(\frac{2\ell_1 + 1}{2}\right) \frac{k_3^3}{N_i(k_1) N_i(k_2, k_3, k_4)} \int_{V_{q_1}} d^3 q_1 d^3 q_2 d^3 q_3 \delta_D(q_1 + q_2 + q_3) \delta_D(q_2 + q_3 + q_4) P_B^0(q_2) B_B^0(q_2, q_3, q_4) + 2 \text{ perm.}
\]
\[
= \frac{1}{2\pi} \left(\frac{2\ell_1 + 1}{2}\right) \frac{\delta^K_{12}}{N_i(k_2) N_i(k_3, k_4)} \int_{V_{q_1}} d^3 q_1 d^3 q_2 d^3 q_4 \delta_D(q_2 + q_3 + q_4) P_B^0(q_2) B_B^0(q_2, q_3, q_4) + 2 \text{ perm.}
\]
\[
\approx \frac{1}{2\pi} \left(\frac{2\ell_1 + 1}{2}\right) \frac{\delta^K_{12}}{N_i(k_2) N_i(k_3, k_4)} B_2^{(0)}(k_2, k_3, k_4) + 2 \text{ perm.},
\]
\[(A14)\]

where once more we have used the same approximation of the power spectrum multipoles and bispectrum monopole not varying significantly inside the integration volume.
APPENDIX B: VALIDATION TESTS

In Table B1 we report the results obtained compressing the bispectrum with respect to the shifted fiducial parameter sets. This is to test whether the performance of the compression is affected by the choice of fiducial set of parameter values. In particular, we consider two cases by varying the fiducial cosmology by adding/subtracting 1σ 1D credible intervals (derived from the MCMC) to all the parameters. The table quantifies that the shifts in the means of the 1D posterior distributions produced by considering a non-optimal fiducial cosmology are small compared to the 1σ 1D credible intervals of the MCMC results.

In Figures B1 and B2 the 1 and 2-D posterior distributions obtained via MCMC/MC-KL/G-PCA for the test cases relative to the Δk6 and Δk5 binning cases are shown. MC-KL recovers with very good approximation the 1 and 2-D posterior distributions derived by the MCMC. G-PCA shows a slightly greater loss of information for the Δk6 case. However this is noticeably closer to the MCMC/MC-KL result when the number of triangles used is increased (Δk5 case).

In Figure B3 we compare the best-fit model obtained by varying four parameters (b1, b2, f, σ8) with the best-fit model corresponding to a fit done via standard MCMC sampling with only three parameters varied, (b1, b2, f), with σ8 = σ8fidc. For the three parameter case we find running the MCMC: b1 = 1.98 ± 0.01, b2 = 0.39 ± 0.06, f(CMASS) = 0.52 ± 0.03 with σ8fidc(CMASS) = 0.61.

Thereby we show that the discrepancy between the results of this paper and the ones presented in the BOSS collaboration analysis Gil-Marín et al. (2017) is only due to the different range of scales considered. Indeed, by limiting our analysis to a smaller range of scales in k-space, the degeneracy between the amplitude-like parameters b1 and σ8 is much stronger. That is visible in Figure B3, where the models given by sets of parameters with very different b1, b2 and σ8 parameters produce very similar predictions of the signals all with good χ2red and p-values.
Compression on BOSS measurements

(MCMC $\Delta k_5$

MC-KL $\Delta k_5$

G-PCA $\Delta k_5$

(a) MC-KL

(b) G-PCA

Figure B2. Joint data-vector $[P(0) f B(0)]$ posteriors: MC-KL and G-PCA four-parameter $\Delta k_5$ case.

Both a) and b) are the same as for Figure B1 for the $\Delta k_5$ case.

Table B1. Four parameter-case, checking consistency for shifted fiducial cosmology.

Upper half: Mean values of the posterior distributions and 68% credible intervals for the MCMC and the MC-KL / G-PCA compression methods. We report the values for the $\Delta k_6$ binning case for both compression methods in three cases consisting in using for the compression: the fiducial cosmology, the fiducial cosmology shifted by $+1\sigma$ and the fiducial cosmology shifted by $-1\sigma$.

Lower half: In the compression columns we report the relative difference between the posterior modes obtained via MCMC and the ones obtained via compression (MC-KL or G-PCA). In the MCMC columns the relative size of the 68% credible intervals obtained via MCMC sampling is shown. By comparing the MCMC columns to the compression ones, it is clear that the difference between the mean parameter values obtained via MCMC and the ones obtained via compression (MC-KL or G-PCA) are evidently within the 68% credible intervals given by the MCMC on the full data-vector.

| $\Delta k_6$ | $\Delta k_6 + 1\sigma$ | $\Delta k_6 - 1\sigma$ |
|--------------|------------------------|------------------------|
| MCMC MC-KL G-PCA | MCMC MC-KL G-PCA | MCMC MC-KL G-PCA |
| $b_1$ | 2.41 ± 0.22 | 2.41 ± 0.23 | 2.49 ± 0.27 |
| $b_2$ | 1.00 ± 0.40 | 1.04 ± 0.42 | 1.08 ± 0.47 |
| $f$ | 0.69 ± 0.08 | 0.72 ± 0.09 | 0.72 ± 0.09 |
| $\sigma_8$ | 0.50 ± 0.04 | 0.48 ± 0.05 | 0.48 ± 0.05 |
| $\Delta \theta_{\text{mc}}^{\text{comp}} - \theta_{\text{mc}}^{\text{inc}}$ | [\%] | $\theta_{\text{mc}}^{\text{comp}} - \theta_{\text{mc}}^{\text{inc}}$ | [\%] |
| $b_1$ | 2.15 -0.26 | 8.57 0.31 |
| $b_2$ | 3.47 28.68 | 25.29 13.26 |
| $f$ | 0.84 0.51 | 6.96 0.26 |
| $\sigma_8$ | -3.25 -2.91 | -8.94 -1.39 |
Figure B3. Reduced $\chi^2$ and $p$-values for the best-fit parameters obtained using the MCMC/MC-KL methods with varying $\sigma_8$ and for the MCMC leaving $\sigma_8 = \sigma_{8\text{ fid}}$ fixed. The $k$-binnings shown for the four parameter case ($b_1$, $b_2$, $f$, $\sigma_8$) are respectively the standard $\Delta k_6$ (navy) for the MCMC and the $\Delta k_2$ (pink) for the MC+KL. The line corresponding to the fit obtained by letting free to vary only the parameters ($b_1$, $b_2$, $f$) is shown in green. The two upper panels are for the power spectrum monopole (left) and quadrupole (right) while the bottom panel refers to the bispectrum monopole. The lower part of each panel shows the relative difference between the data measurements and the different models. Even if for example $b_1$ and $\sigma_8$ values are shifted in the cases of $\Delta k_6$ and $\Delta k_2$, this is due to the strong degeneracy between them and both models are practically identical to the one given by the three parameters fit ($b_2$, $b_2$, $f$) with $\sigma_8 = \sigma_{8\text{ fid}}$. The only way to converge to the results obtained by the BOSS collaboration is to consider a larger range of scales (as they have done) for both power spectrum and bispectrum which however involves a more complex modelling of the data-vector.