

# Covariance Testing

DESC Members  
(Dated: May 14, 2019)

There are a number of codes that compute covariance matrices analytically; the plan is to use these to build TJPCov. In this project, we start along the path of comparing these different codes, building up a suite of tools that can be used to compare covariance matrices. We expect these tools to be useful not only for converging on a single accurate code for computing covariance matrices but also more generally for understanding which parts of the covariance matrix carry the most information (and therefore need the most attention to get right) and which are not relevant (so for example matrices that are not positive definite may still be usable if the negative eigenmodes are not relevant).

## I. INTRODUCTION

## II. METHODS

There are several ways to test covariance matrices. In this work, we will illustrate these on cosmic shear statistics  $\xi_{\pm}(\theta)$ , focusing for the most part on the Year 1 results of the Dark Energy Survey [1] (DES Y1), and also on predictions for Year 3 (DES Y3). For each  $\xi_{\pm}(\theta)$  and  $\xi_{\mp}(\theta)$ , we have 200 datapoints, which give 400 in total, divided in 20 angular bins between 2.5 and 250 arcmin, and five tomographic bins, as described in [2]. We apply the same angular cuts to this dataset as in the aforementioned paper, resulting in 227 datapoints, which correspond to a  $227 \times 227$  covariance matrix.

The two codes employed here for computing the covariance matrices are *Cosmolike* [5] and the one used to analyze the KiDS-450 survey [6], hereafter referred to as CL and BJ, respectively. To perform cosmological parameter inference we used the *CosmoSIS* [9] pipeline, choosing the *MultiNest* [4] sampler to explore the parameter space, with 500 livepoints, efficiency set to 0.8, and tolerance to 0.1.

### A. One-to-one Comparison

This is a simple matter of comparing elements of a covariance matrix, usually starting with diagonal elements. Figure 1 shows an example.

### B. Eigenvalues and eigenvectors

An intuitive approach to identifying the most contributing elements of a covariance matrix is to look at its eigenvalues, since we can associate them to the noise in the measurement. The lowest eigenvalues would correspond to the least noise and therefore hold the most information about the parameters, whereas the highest eigenvalues would contribute significantly less. We check this assumption by replacing the 50 highest eigenvalues with a very large number (about nine orders of magnitude higher), thus removing their effective contribution,

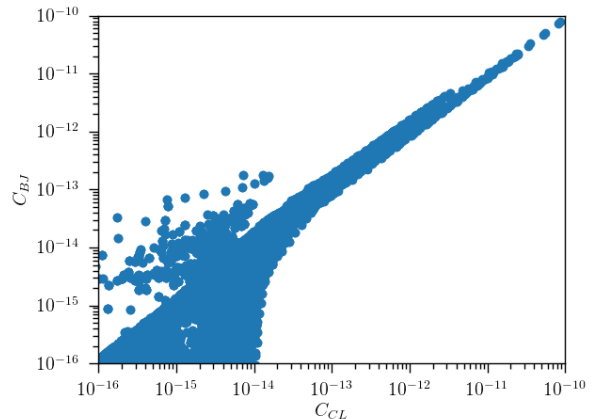


FIG. 1. A simple scatter plot of elements of covariance matrices produced by two separate halo model codes.

and then obtain a new covariance matrix with the modified eigenvalues. To check whether we have indeed kept the most informative modes, we perform an analysis with the new covariance matrix, to constrain cosmological and systematic parameters; which we then compare to the constraints obtained with the original covariance matrix. The results are shown in Fig. 2, for  $\Omega_m$  and  $\sigma_8$ , where we can see that although the constraints are consistent, we have lost a noticeable amount of information of the parameters.

A visual comparison of two matrices produced by CL and BJ is shown in Fig. 3 where we have plotted their eigenvalues in blue and orange, respectively. On average, the eigenvalues differ by an order of  $\approx 15\%$ . Notice that the order of magnitude of the values vary greatly, making it difficult to compare them efficiently; furthermore, it is not clear which elements are most important.

Figure 4 shows an example of one of the eigenvectors, the one associated with the smallest eigenvalue. This low-eigenvalue mode picks up the differences between the correlation function at different angular scales (each vertical line delineates between two-point functions of shears in different tomographic bin pairs).

Another interesting way of analysing covariance matrices is by calculating the signal to noise ratio (S/N). A simple way of achieving this is to note that the total

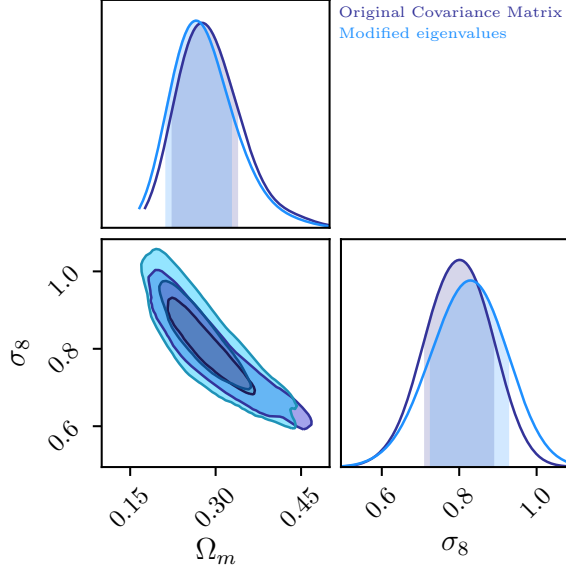


FIG. 2. Constraints on cosmological parameters  $\Omega_m$  and  $\sigma_8$  for the original DESY1 covariance matrix (in purple) and for a new covariance matrix (in blue) obtained by setting the fifty highest eigenvalues of the original matrix to nine orders of magnitude higher. The constraints are consistent, however, we notice a slightly larger  $1\sigma$  error for the latter.

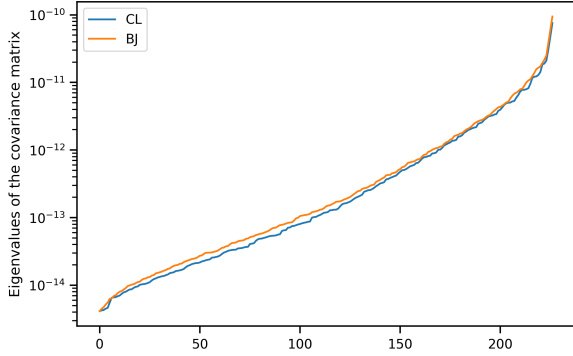


FIG. 3. A plot showing the 227 eigenvalues of two covariance matrices produced for cosmic shear for the Year 3 projection of the Dark Energy Survey. The blue curve is for CL and the orange one is for BJ.

signal to noise squared is

$$\left(\frac{S}{N}\right)^2 = \sum_{ij} D_i C_{ij}^{-1} D_j, \quad (1)$$

where  $D_i$  and  $D_j$  are the data points and  $C$  the covariance matrix. If  $C$  were diagonal, then the eigenvectors would simply be the data points themselves, and we could estimate the signal to noise squared expected in each mode by simply computing  $T_i^2/C_{ii}$  where  $T_i$  is the theoretical prediction. Then we could throw out the modes with the lowest signal to noise. Since  $C$  is not diagonal,

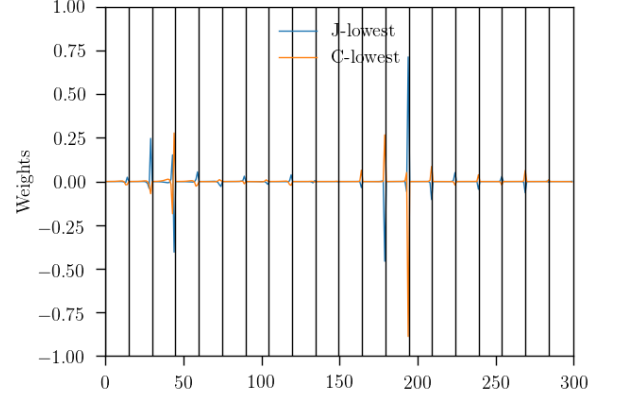


FIG. 4. A simple scatter plot of elements of covariance matrices produced by two separate halo model codes.

we have to first diagonalize it and then order the values. So, we write the expected signal to noise squared as

$$\begin{aligned} \left(\frac{S}{N}\right)^2 &= \sum_{ij} T_i C_{ij}^{-1} T_j \\ &= \sum_i \frac{v_i^2}{\lambda_i}, \end{aligned} \quad (2)$$

where  $\lambda_i$  are the eigenvalues of the covariance matrix, which is diagonalized with the unitary matrix  $U$ , and the eigenvectors are

$$v_i \equiv U_{ij}^t T_j \quad (3)$$

This makes it very clear which modes should be kept and which should be dropped. Modes  $v_i$  for which  $v_i^2/\lambda_i$  is very small can be discarded.

We follow the same procedure described earlier for the eigenvalues, where we remove the lowest values by setting them to even lower values. When doing this, we must be careful, because the highest modes do not necessarily hold the most information for all of the parameters; it is therefore necessary to check this for each parameter before applying the cuts. To test this we take

$$\left(\frac{\partial S/\partial \theta}{N}\right)^2 = \sum_i \frac{(\partial v_i/\partial \theta)^2}{\lambda_i}, \quad (4)$$

with

$$v_i^\theta \equiv U_{ij}^t \frac{\partial T_j}{\partial \theta}. \quad (5)$$

The importance of this is illustrated in Fig. 5, where we can see that while the highest signal to noise for all the parameters does indeed correspond to those that hold most information for parameters  $\Omega_m$  and  $A_s$ , this is not the case for the intrinsic alignment parameters  $A$  and  $\eta$ , so that we end up losing constraining power over these parameters when we modify the lowest values of S/N. This is clearly shown by looking at Fig. 5, which

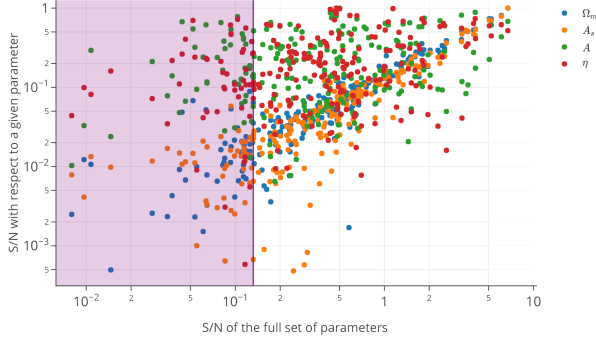


FIG. 5. Scatter plot for the relation between the signal to noise obtained with the covariance matrix for DESY1 for each parameter (x-axis) against that for the full set of parameters (y-axis). The derivatives are shown with respect to cosmological parameters  $\Omega_m$  (blue) and  $A_s$  (orange), and for the intrinsic alignment parameters  $A$  (green) and  $\eta$  (red). The purple rectangle spreads until the fiftieth lowest value of S/N, which corresponds to the values that were modified for parameter constraints.

shows constraints for the aforementioned parameters for the modified S/N and compares them against the ones obtained for the original covariance matrix.

### C. Parameter Estimation

Ultimately, what matters is how well the likelihood does at extracting parameter constraints. Since most analyses assume a Gaussian likelihood, this boils down to how well the contours in parameter space agree when computing the  $\chi^2$  using two different covariance matrices.

### D. Shrinkage

There have been several methods proposed in the literature to compress the data vectors, extracting as much information as possible. Here we consider two: first compression at the map level [3], where linear combination of the tomographic maps are used. If there are 4 tomographic bins, an uncompressed analysis would require ten separate 2-point functions (or 20 for cosmic shear), whereas a compression scheme leads to just a few uncorrelated maps. If there were 3 such maps, then only three 2-point functions would need to be used for the likelihood analysis.

We characterize a given element in the data vector by its angular and tomographic indices:  $a_i = a_{lm,\alpha}$  where  $l, m$  denote the indices corresponding to given spherical harmonics and  $\alpha$  is a given tomographic bin, or equivalently  $a_i = a_\alpha(\vec{\theta})$  in real space. The compression occurs in terms of tomographic bins, so that

$$b_\mu(\vec{\theta}) = \sum_\alpha F_{\mu\alpha} a_\alpha(\theta) \quad (6)$$

where the  $F$ 's are chosen [7] so that the correlation functions of the  $b$ 's are diagonal in tomographic space:

$$\begin{aligned} w_{\mu\nu}(\theta) &= \langle b_\mu(\vec{\theta}_1) b_\nu(\vec{\theta}_2) \rangle |_{\vec{\theta}_1 - \vec{\theta}_2 \in \theta} \\ &= \delta_{\mu,\nu} w_\mu(\theta). \end{aligned} \quad (7)$$

Usually with  $N_t$  tomographic bins, one must consider  $N_t(N_t+1)/2$  correlation functions, but – due to the transformation that diagonalizes the elements – there are only  $N_t$  correlation functions to consider (for a given  $\theta$ ). Even better, these can be ordered by the information they contain, so fewer than  $N_t$  can be used. In the example used in [3], 16 tomographic bins were assumed, so that the standard treatment would require 136 correlation functions, but only 3 were needed in order to extract accurate constraints.

This method suggests a way of extracting the most important pieces of the full covariance matrix  $C$ . We simply compute the covariance matrix of the  $w_\mu$  in terms of  $C$  and keep only the most important terms. That is,

$$\begin{aligned} C_{\mu\nu}^b &\equiv \langle (w_\mu - \bar{w}_\mu)(w_\nu - \bar{w}_\nu) \rangle \\ &= \sum_{\alpha\alpha'\beta\beta'} F_{\mu\alpha} F_{\mu\alpha'} F_{\nu\beta} F_{\nu\beta'} C_{\alpha\alpha'\beta\beta'}. \end{aligned} \quad (8)$$

Here the angular indices have been suppressed (there are two of them, one for each  $w_\mu$ ), but the full tomographic complexity has been retained. The covariance matrix on the right includes a total of  $N_t^2 \times N_t^2$  terms (some of which are equal because of symmetry) corresponding to all possible pairs of two-point functions. But  $C^b$  on the right contains only  $N_t^2$  elements, and again these are ordered, so we can use only a subset of them. In the simplest case, where only one linear combination is needed so  $\mu = \nu = 1$ , a single number (for each pair of angular bins) captures all the relevant information from the full covariance matrix.

The second compression takes place at the 2-point level [8], with the compressed data vector containing linear combinations of the many 2-point functions. In principle, this might work with only  $N_p$  2-point functions where  $N_p$  is the number of parameters varied, and each mode, or linear combination, contains all the information necessary about the parameter of interest.

For each parameter  $p_\alpha$  that is varied, one captures a single linear mode

$$y_\alpha = U_{\alpha,i} D_i \quad (9)$$

where  $D_i$  are the data points and the coefficients are defined as

$$U_{\alpha,i} \equiv \frac{\partial T_j}{\partial p_\alpha} C^{-1}_{ji} \quad (10)$$

where  $T_j$  is the theoretical prediction for the data point  $D_j$ . The now much smaller data set  $\{y_\alpha\}$ , which contains as few as  $N_p$  data points carries with it its own covariance matrix, with which the  $\chi^2$  can be computed for each point

in parameter space. Propagating through shows that this covariance matrix is related to the original  $C_{ij}$  via

$$C_{\alpha\beta} = U_{\alpha i} C_{ij} U_{j\beta}^t. \quad (11)$$

To take an example from DES-Y1,  $C_{ij}$  is a  $400 \times 400$  matrix, while the number of parameters needed to specify the model is only 16, so  $C_{\alpha\beta}$  is a  $16 \times 16$  matrix. We have apparently captured from the initial set of  $400 \times 401/2 = 80,200$  independent elements of the covariance matrix a small subset (only 136 in this case) of linear combinations of these 80k elements that really matter. If two covariance matrices give the same set of  $C_{\alpha\beta}$ , we do not care whether any of the other eighty thousand elements differ from one another.

When comparing the constraints obtained via the compressed dataset and corresponding covariance matrix to that obtained with the full, original ones, we must be careful when computing  $U_{\alpha,i}$ . This is because the curve shifts depending on where the derivative is taken. For better results, we must take the derivative at the best-fit value, otherwise the constraints will shift, as discussed in more detail in Appendix A. The power of this compression scheme can be appreciated in Fig.

We apply this methodology to comparing the covariance matrices of interest, i. e. CL and BJ. In order to do this, we take two different approaches: first, we assume that  $U_{\alpha,i}$  is the same for both covariance matrices and we calculate it with BJ. The second approach is that each compression scheme must use the original covariance matrix that will be compressed, so that  $U_{\alpha,i}$  will be different for each covariance matrix. Fig. 6 was obtained for the first method, it shows the correlation matrix for BJ and CL, and that of the difference between them; this figure is important because we can clearly see the difference between the two matrices by simply looking at only  $16 \times 16$  elements, as opposed to having to analyse the larger correlation matrix for the full covariance matrices. It is also crucial that the matrices used for comparison here are those obtained via the same compression scheme, so that we can be sure that their differences are indeed only related to the differences in the original  $227 \times 227$  matrices.

### III. RESULTS

### IV. DISCUSSION

### V. CONCLUSION

#### Acknowledgments

The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des

Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and

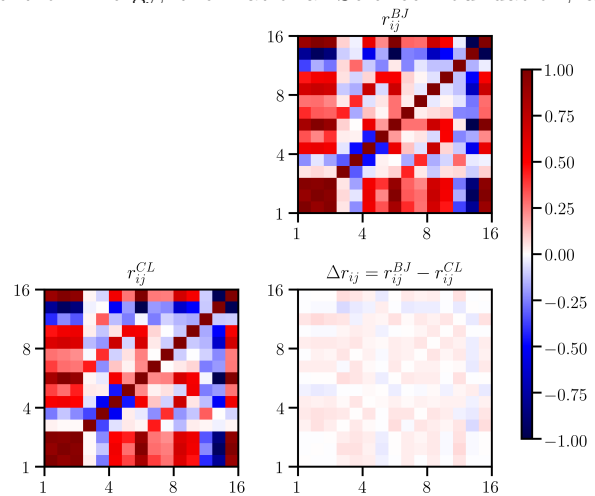


FIG. 6. The upper right and lower left plots display the correlation matrix for BJ and CL respectively, while the lower right is the difference between the two.

the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne - France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515.

- 
- [1] Abbott, T. M. C., et al. 2017, arXiv:1708.01530
  - [2] —. 2018, Phys. Rev. D, 98, 043526
  - [3] Alonso, D. 2018, Mon. Not. Roy. Astron. Soc., 473, 4306
  - [4] Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601
  - [5] Krause, E., & Eifer, T. 2017, Mon. Not. Roy. Astron. Soc., 470, 2100
  - [6] Khlinger, F., et al. 2017, Mon. Not. Roy. Astron. Soc., 471, 4412
  - [7] Note1. ????, note that this definition of  $F$  differs from that in [3] in that it includes the inverse of the noise matrix.
  - [8] Zablocki, A., & Dodelson, S. 2016, Phys. Rev., D93, 083525
  - [9] Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, 45