Covariance Testing

DESC Members
(Dated: July 26, 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

Our analysis is carried out using cosmic shear statistics $\xi_{\pm}(\theta)$, focusing for the most part on the Year 1 results of the Dark Energy Survey [1] (DESY1), and also on predictions for DES Year 3 (DESY3). The data is divided into four tomographic redshift bins spanning the interval $0.20 - 1.30$, which gives us 10 bin combinations, each one containing 20 angular bins between 2.5 and 250 arcmin. We thus have 200 data points for each $\xi_+(\theta)$ and $\xi_-(\theta)$, giving us 400 in total.

We apply the angular cuts described in [2], which removes the scales most sensitive to baryonic effects; this leaves us with 167 points for $\xi_+(\theta)$ and 60 for $\xi_-(\theta)$ (for which baryonic interactions are more relevant), resulting in 227 data, which correspond to a 227×227 covariance matrix.

The two codes employed here for computing the covariance matrices are `Cosmolike` [6] and a simpler version of the code used to analyze the KiDS-450 survey [7], hereafter referred to as CL and BJ, respectively. To perform cosmological parameter inference we use the `CosmoSIS` [9] pipeline, while employing the `MultiNest` [5] sampler to explore the parameter space, with 500 `livepoints`, `efficiency` set to 0.3, `tolerance` to 0.1 and `constant efficiency` set to False.

Parameter	Prior
Cosmological	
Ω_m	$\mathcal{U}(0.1, 0.9)$
$A_s \times 10^9$	$\mathcal{U}(0.5, 5)$
$H_0(\mathrm{kms}^{-1}\mathrm{Mpc}^{-1})$	$\mathcal{U}(55, 91)$
Ω_b	$\mathcal{U}(0.03, 0.07)$
$\Omega_\nu h^2$	$\mathcal{U}(0.0005, 0.01)$
n_s	$\mathcal{U}(0.87, 1.07)$
Astrophysical	
A	$\mathcal{U}(-5, 5)$
η	$\mathcal{U}(-5, 5)$
Systematic	
m^i	$\mathcal{G}(0.012, 0.023)$
Δz^1	$\mathcal{G}(-0.001, 0.016)$
Δz^2	$\mathcal{G}(-0.019, 0.013)$
Δz^3	$\mathcal{G}(0.009, 0.011)$
Δz^4	$\mathcal{G}(-0.018, 0.022)$

TABLE I. List of the priors used in the analysis for parameter constraints. For the cosmological parameters, we fixed $w = -1.0$, $\Omega_k = 0.0$ and $\tau = 0.08$. The astrophysical parameters are associated with the intrinsic alignment, they follow the relation $A(z) = A[(1+z)/1.62]^\eta$, we have also $z_0 = 0.62$. Lastly, for systematics we have m^i corresponding to the shear calibration, and Δz^i for the source photo-z shift, with $i = 1, 4$ in both cases. The lens photo-z shift parameters Δz^i, $i = 1, 5$ were fixed to zero.

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

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 variance. The lowest eigenvalues would correspond to the least variance and therefore hold the most information, whereas the highest eigenvalues would implicate larger error and thus contribute significantly less, which means that the most important elements of the matrix would be those with the lowest eigenvalues. To verify this assumption, we need to look at the parameter constraints: removing the elements with the highest eigenvalues should not alter our results; the opposite, however, should give us a broader constraints. Our procedure consists of first diagonalising the covariance matrix in order to calculate its eigenvalues and then replacing the eigenvalues of interest with numbers of about nine orders of magnitude higher, thus removing their effective contribution;

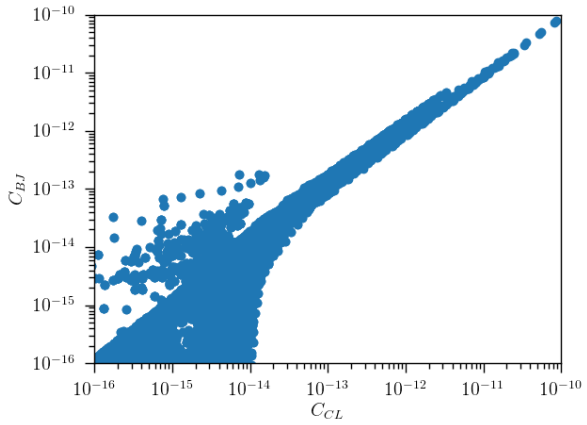

FIG. 1. A simple scatter plot of elements of covariance matrices produced by two separate halo model codes.

we then obtain a new covariance matrix with the modified eigenvalues. Finally, we perform a cosmological analysis with the new covariances matrices, to constrain the parameters of our model.

C. Signal to noise

Another way of analysing covariance matrices is by obtaining the signal-to-noise ratio (SNR). A simple way of achieving this is to note that the total SNR squared is

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

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 SNR 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 SNR. Since C is not diagonal, we have to first diagonalise it and then order the values. So, we write the expected SNR squared as

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

where λ_i are the eigenvalues of the covariance matrix, which is diagonalised with the unitary matrix U, and the eigenvectors are

$$v_i \equiv U_{ij}^t T_j \tag{3}$$

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

After obtaining the SNR for our covariance matrix, we proceed to set the 50 lowest values to seven orders of magnitude lower, which is equivalent to increasing the noise (or decreasing the signal) of these modes. We then obtain a new covariance matrix with the corresponding modified SNR values.

D. Shrinkage

In order to compare covariance matrices and to extract the most important elements, we need to shrink the data vectors and their corresponding covariance matrices. 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 to gain maximum signal to noise ratio. Then we transfer this compression scheme into real-space to compress the two point function. The second method directly operates in the real-space [8], where modes that maximize the Fisher infomation for each of the cosmological parameters are taken to compress the data vector and the covariance matrices.

Shrinking data vector and its covariance matrix is equivalent to projecting the data vector onto a subspace of the original vector space. For example, if our data have a length of n and we want to compress it to a lower dimension $m \leq n$, we need to come up with a basis transformation matrix $W_{m\times n}$ that has orthonormal row vectors, or mathmatically speaking, $< W_{xi} \cdot W_{yi} >= \delta_{xy}$. Once we come up with a basis transformation matrix $W_{m\times n}$, we can compress the data vector y and covariance matrix C into filter data $\tilde{y}$ and $\tilde{C}$

$$\tilde{y} = Wy \tag{4}$$

$$\begin{aligned}\tilde{C} &= < \tilde{y}\tilde{y}^T > \\ &= W < yy^T > W^T = WCW^T \end{aligned} \tag{5}$$

For a data vector with length m, there are $m(m+1)/2$ elements in the covariance matrices to be evaluated. Once the data vector is shrunk to length n, elements in covariance matrices will shrink quadratically. Therefore, shrinking the covariance matrices is a effective way of comparing them.

Here, we will demonstrate the two mentioned method to compress the covariance matrices. In both methods, the data vector is shrunk by 90 percent, so the covariance matrices are shrunk by 99 percent.

1. *Tomographic Compression*

The first method we applied to compress the data vector is based on Karhunen-Loéve decomposition for the shear power spectrum suggested by [3] and later applied to real space two-point function by [**?**] for CFHTLens

survey. This method generally finds an eigenmode with most of the signal-to-noise ratio contribution in the power spectrum, and then it can transform the two point function in real space based on this eigenmode.

With `CosmoSIS`, we can generate the shear power spectrum C_l of convergence a_{lm} for a fiducial cosmology. The fiducial cosmology we choose is the best fit value of the DES Year 1 shear only results. With the shear power spectrum $C_l = S_l + N_l$ and its shape noise N_l, we can calculate the KL modes matrix E_l^p by a general eigenvalue problem:

$$C_l E_l^p = \lambda^p N_l E_l^p \tag{6}$$

And the new observable $b_{lm} = E_p \cdot N^{-1} a_{lm}$. We should notice that C_l is the power spectrum of the convergence of the weak lensing, and E_p is the transformation of basis for the convergence. We should also notice that these eigenmodes are uncorrelated, so the power spectrum of the new observable D_l will be a diagonal matrix, with 1+signal-noise-ratio of the corresponding eigenmodes on the diagonal elements.

$$D_l = < b_{lm} b_{lm}^T > = E_l^p N^{-1} C_l N^{-1} E_l^{pT} \tag{7}$$

Or, if we denote $E_l^p N^{-1}$ as R_l, we can write the compression in one simple linear combination of the C_l

$$D_l = R_l^i C_l^{ij} R_l^j = W_l^{ij} C_l^{ij} \tag{8}$$

The double summation weight W_l^{ij} is the weight we want to perform tomographic compression. Since the KL decomposed modes of shear power spectrum are uncorrelated, we can make a compression here by only taking the first one or two modes with the highest S/N. By doing so, we compressed 10 tomographic combinations to 1 or 2.

However, we want to evantually compress the two point function data vector of DESY1. A possible way is to calculate the two point function of the KL mode of the shear power spectrum. We can calculate the two point function from the shear power spectrum by

$$\xi_{+/-}^{ij}(\theta) = \int \frac{ldl}{2\pi} J_{0/4}(l\theta) C^{ij}(l) \tag{9}$$

In order to compress the two-point function based on the compression of the C_l, we need to make sure that the compression scheme for C_l is l-independent, that is to say, the two point correlation function of D_l $\tilde{\xi}_{+/-}(\theta)$ can be directly calculated from other two-point functions.

$$\begin{aligned} \tilde{\xi}_{+/-}(\theta) &= \int \frac{ldl}{2\pi} J_{0/4}(l\theta) D(l) \\ &= \int \frac{ldl}{2\pi} J_{0/4}(l\theta) W^{ij} C^{ij}(l) \\ &= W^{ij} \xi_{+/-}^{ij}(\theta) \end{aligned} \tag{10}$$

Where W^{ij}, the l-independent compression weight is calculated by

$$W^{ij} = \frac{\int_{l_{min}}^{l_{max}} (2l+1) W_l^{ij}}{\int_{l_{min}}^{l_{max}} (2l+1)} \tag{11}$$

We make a more conservative angular cut than the angular cut discussed in [ref]. We make sure that the cut for both $\xi_{+/-}$ are uniform regarding tomographic combinations. For ξ_+, we considered an angular scale from 7.195 degree to 250.0 degree. For ξ_-, the angular scale is from 90.579 to 250.0. Therefore, for the purpose of demostrating KL-transform, the raw data vector has a length of 190, and by shrinking 10 tomographic combinations for each angle into 1 KL-mode, the data is shrunk to 19. Therefore the number of elements in the covariance matrices are shrunk by 99 %.

2. 2-Point Correlation Function

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_α that is varied, one captures a single linear mode

$$y_\alpha = U_{\alpha i} D_i \tag{12}$$

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

$$U_{\alpha i} \equiv \frac{\partial T_j}{\partial p_\alpha} C^{-1}{}_{ji} \tag{13}$$

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 χ^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

$$C_{\alpha\beta} = U_{\alpha i} C_{ij} U_{j\beta}^t. \tag{14}$$

To take an example from the full DESY1 for cosmic shear, C_{ij} is a 400×400 matrix, while the number of parameters needed to specify the model is only 16, so $C_{\alpha\beta}$ is a 16×16 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.

In our case, our covariance matrix is 227×227, which we then reduce to 16×16 elements. Figure include figure compares the constraints obtained for the compressed covariance and data with the full one, for Ω_m and σ_8. Include plots and talk about constraints. While this compression does not speed up calculations, it is encouraging to see that we can compress matrices of any size without losing a significant amount of information on the parameters of our models.

One relevant point in this analysis is at which point to take the derivative of each parameter. When we wish to compare the results of our compression scheme with those obtained with the full covariance matrix and data set, it is important to derivate each parameter at their respective mean value (obtained by performing the analysis with the full covariance matrix). The shape and variance of the posterior is not dependent on the derivative, but the best fit value shifts according to the point where the derivative was taken.

We also 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. Figure 3 was obtained for the first method, it shows the correlation matrix for BJ and CL, and that of the difference between them; we find this figure important because we can clearly see the difference between the two matrices by simply looking at only $(16 \times 17)/2$ 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 matrices.

3. *Invertible Transformation*

While the compression scheme is good?, it is not devoid of problems, the most concerning one being that we cannot recover the original covariance matrix the compressed one. Our motivation is to have an invertible transformation so that we are able to switch between a compressed and the full covariance. In order to do this, we start off with the compression scheme presented in Eq. 13, which will serve as the basis for a rotation matrix W_α. We then use the Gram-Schmidt decomposition to create $227 - N_p$ vectors orthonormal to U_α, thus obtaining a unitary 227×227 matrix. We then have,

$$C'_{\alpha\beta} = W_{\alpha i} C_{ij} W^t_{j\beta}. \tag{15}$$

When this transformation is applied to the original matrix, C_{ij}, the first $N_p \times N_p$ elements of the new matrix will be the same as we had in $C_{\alpha\beta}$. Since our transformation

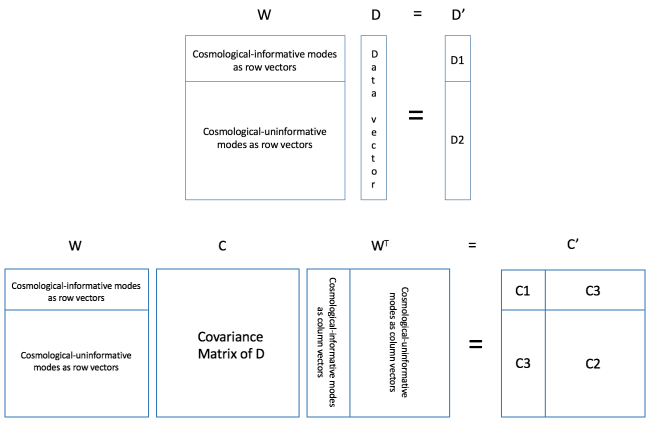

FIG. 2. **Top:** Illustration of transformation of data; **Bottom:** Transformation of the covariance matrix. Detail of each section in the transformed data and covariance are described below.

spreads out the information contained in the compressed matrix, the relevant values for parameter estimation will be contained in the first N_p rows and columns of the new covariance matrix. This means that we will have $N_p\,(2 \times 227 - N_p)$ important elements, which is about an order of magnitude less than in the case of the full covariance matrix.

****

In the last section, we shrunk the data vector and covariance matrices and find the most cosmological-informative modes in the data vector. However, in order to make a tolerence testing of each element in the covariance, we not only need the cosmological-informative modes but also the uninformative modes. Suppose we have found the informative set of modes for cosmology, those modes that are orthogonal to the informative set, or the complementary set of the informative set, form a set of modes that are cosmological-uninformative. We can use Gram-Schmidt decomposition to find these orthogonal modes.

With this inverseable transformation W shown in Figure 2, the data vector and covariance matrix are transformed into sectional blocks. We will describe the meaning of each blocks here. The data vector is split into two blocks. D1 is the block of transformed data points that are sensitive to the changes in cosmological parameter set, that is, the cosmological-sensitive data. D2 is generated by the modes that are orthogonal to the cosmological sensitive mode, they will in priciple stays relatively constant when cosmological parameter changes, and are uninformative for parameter constraints.

The transformed covariance matrix is split into 4 blocks, with the off-diagonal blocks being the transpose of each other, there are 3 different blocks. Block C1 is the variance and covariance of the cosmological informative data,or D1. This block contribute almost everything to the χ^2 calculation. C1 describes how well the cosmological-sensitive data is measured, so the value in

Added error	Standard deviation
$0-20\%$	< 5
$20-40\%$	< 20
$> 50\%$	> 20

TABLE II. The first column is the introduced error to the relevant elements of the rotated covariance matrix for Ω_m. The introduced error corresponds to the an arbitrary increase in each element in the first row and first column of the rotated covariance matrix

this block will affect the parameter constraints. Block C2 is the variance and covariance of the uninformative data,or D2. It plays an minor role in the χ^2, so it will affect the parameter constraints less. C3 is the cross-correlation between D1 and D2, and it will affect the parameter constraints because it describe how D1 is affected by the uncertainty of D2.

By making the transformation described above, we will be able to reorganize the covariance matrices into cosmological infomative blocks and uninformative blocks. Intrinsiclly, we can tolerate more errors in the uninformative blocks and have a stronger requirement on the informative blocks. This will put a lot of simulation time into more valuable work.

The gold of the invertibility of the transformation is that we can first assign higher tolerence to C2 and assign lower tolerence to C1 and C3, then use the transformation matrix to recover the covariance matrices in the original basis. Then, we can compare the recoverd matrix with the original one, and check out how much tolerence is allowed on each element. Without invertibility, we cannot accomplish this task, which allows us to precisely quantify the tolerence on each element in the covariance matrices.

****

E. Tolerence testing

III. RESULTS

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 χ^2 using two different covariance matrices.

A. One-to-One Comparison

Put previous result of One-to-one comparison here and describe

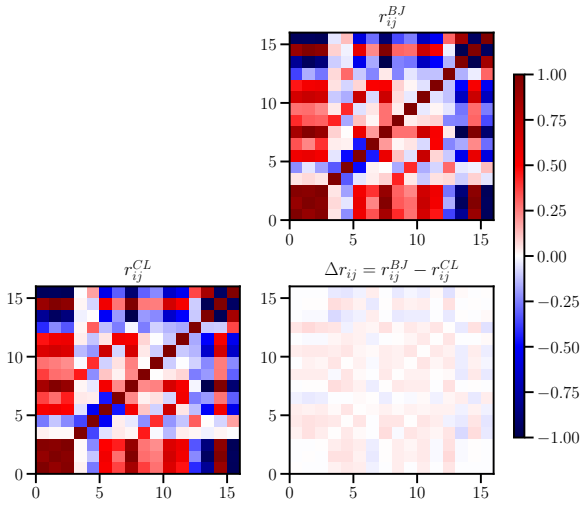

FIG. 3. The upper right and lower left plots display the correlation matrix for BJ and CL respectively, while the lower right is the difference between the two.

B. Eigenvalues

We perform two different analysis: the first consist of modifying the fifty lowest eigenvalues, and for the second, we take the fifty highest. The results of both approaches are compared to those obtained with the original covariance matrix, in Figure Figure 4, where, for brevity, we show only Ω_m and σ_8. We can see in the Posterior Density Functions (pdfs) that the mean values of both parameters are well within the 2σ interval, the same is true for all the fourteen parameters; the size of the constraints is also within less than 5% of each other for almost all of the parameters. Since the constraints are similar for both approaches, it becomes clear that ranking eigenvalues is not the most efficient way of evaluating where the most contributing elements of the covariance matrix lie.

Next, we wish to compare matrices using their eigenvalues. We plot the results obtained for constraints on Ω_m and σ_8 for BJ and CL in Figure 5, where we can see that, while their best fit values agree within a 2σ interval, the constraints are up to 25% broader for the latter. This is also true for the other parameters, where, on average, constraints obtained with CL are about 18% wider. To see if we can identify these differences in the eigenvalues of the covariance matrices, we plot Figure 6. At a first glance, both curves show reasonable agreement, with values differing only by an average of $\approx 15\%$, but we should keep in mind that the order of magnitude of the values vary greatly, which makes it difficult to compare them efficiently with this methodology. Furthermore, from our previous analysis it is not clear which elements are most important.

Figure 7 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

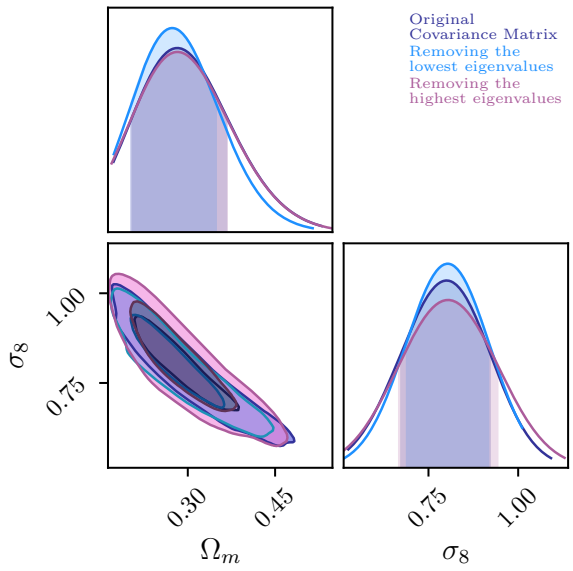

FIG. 4. Constraints on cosmological parameters Ω_m and σ_8 for the original DESY1 covariance matrix (in purple) and for two new covariance matrices obtained by setting the fifty highest eigenvalues of the original matrix to nine orders of magnitude higher (in blue), and by replacing the fifty lowest eigenvalues to nine order of magnitude lower (in magenta).

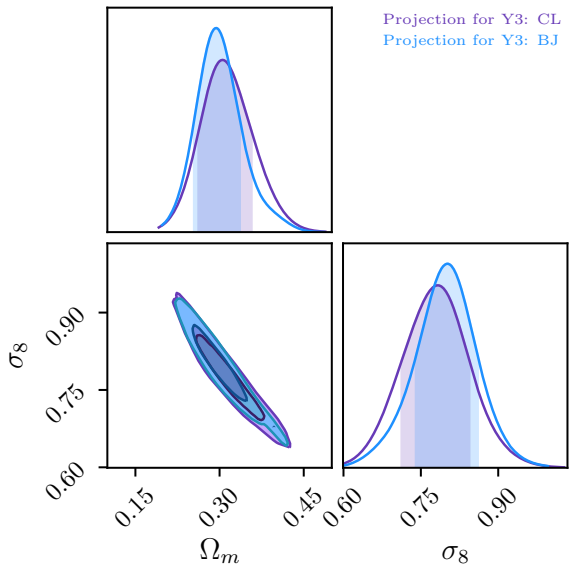

FIG. 5. Constraints on cosmological parameters Ω_m and σ_8 for two covariance matrices produced for cosmic shear for DESY3. The purple curve is for CL while the blue is for BJ. The constraints are about 18% broader for the former, indicating that the two matrices have quantifiable differences between them.

correlation function at different angular scales (each vertical line delineates between two-point functions of shears in different tomographic bin pairs).

C. Signal-to-noise ratio

The parameter constraints for this method are shown in Figure 8, where we note that the mean values are shifted about 12% and 5% for Ω_m and σ_8, respectively, with 17% broader constraints for the former, when compared to the results with the original covariance, which tells us that the modes removed were essential for obtaining better constraints on these parameters.

Another aspect that we must consider when applying this procedure is that the highest modes do not necessarily hold the most information for all of the parameters; this is evident when we analyse the SNR for each parameter individually. To illustrate this we take

$$\left(\frac{\partial S/\partial p_\alpha}{N}\right)^2 = \sum_i \frac{(\partial v_i/\partial p_\alpha)^2}{\lambda_i}, \tag{16}$$

with

$$v_i^\alpha \equiv U_{ij}^t \frac{\partial T_j}{\partial p_\alpha}, \tag{17}$$

where $\partial/\partial p_\alpha$ is the derivative with respect to each parameter θ. This produces the SNR for each parameter of interest. The importance of this procedure is illustrated

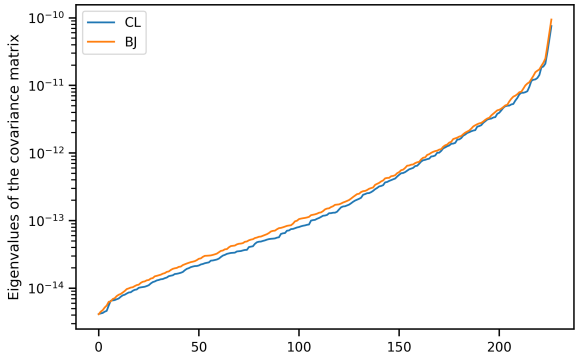

FIG. 6. A log plot showing the 227 eigenvalues of two covariance matrices produced for cosmic shear for the Year 3 projection of the Dark Energy Survey. The blue curve is for CL and the orange one is for BJ.

in Figure 9, where we can see that while the highest SNR for all the parameters does indeed correspond to those that hold most information for parameters Ω_m and A_s, this is not the case, however, for the intrinsic alignment parameters A and η. As a result, we may end up losing constraining power over these parameters when we modify the lowest values of SNR. While this is apparent in the aforementioned figure, it is not very clear when looking at the resulting constraints on these parameters because the constraining power of the data over these parameters is not very strong (errors of about 100%). As such, these results do not encourage us to use this method for identifying the most important elements of the covariance

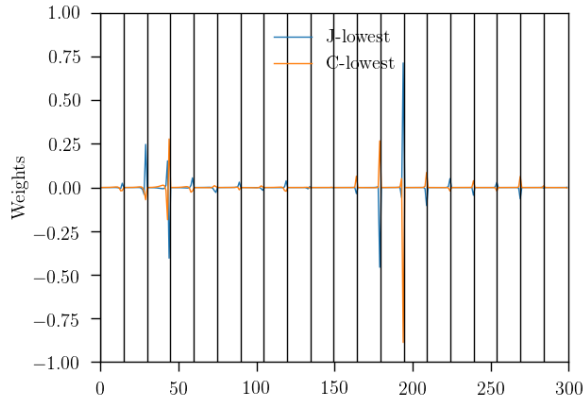

FIG. 7. A simple scatter plot of elements of covariance matrices produced by two separate halo model codes.

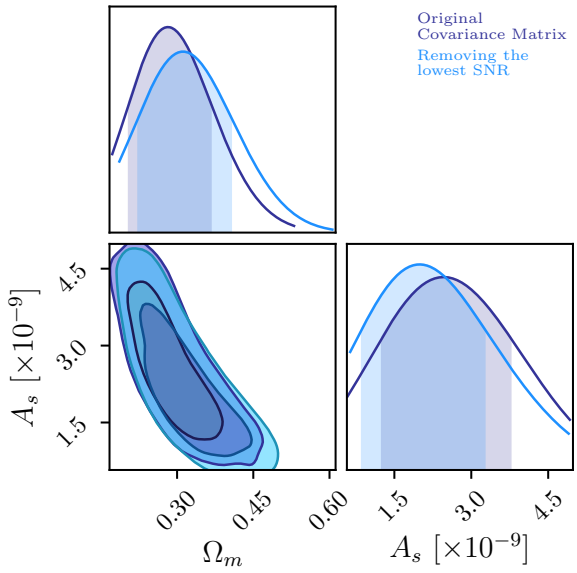

FIG. 8. Constraints on cosmological parameters Ω_m and A_s for the original DESY1 covariance matrix (in purple) and for a covariance matrix (in blue) obtained by setting fifty elements corresponding to the lowest SNR to a value seven orders of magnitude lower, in order to evaluate their contribution to parameter constraints.

matrix.

D. Shrinkage

In this part, we will show the result of our covariance shrinkage, including the weighting function and the compressed covariance matrices.

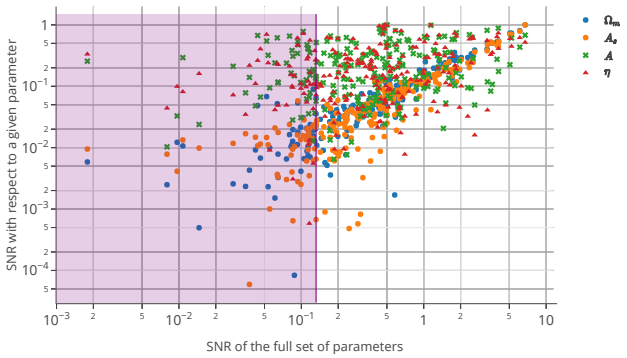

FIG. 9. Scatter plot for the relation between the signal to noise obtained with the covariance matrix for DESY1 for each parameter (x-axis) against that for the full set of parameters (y-axis). The derivatives are shown with respect to cosmological parameters Ω_m (blue) and A_s (orange), and for the intrinsic alignment parameters A (green) and η (red). The purple rectangle spreads until the fiftieth lowest value of SNR, which corresponds to the values that were modified for parameter constraints.

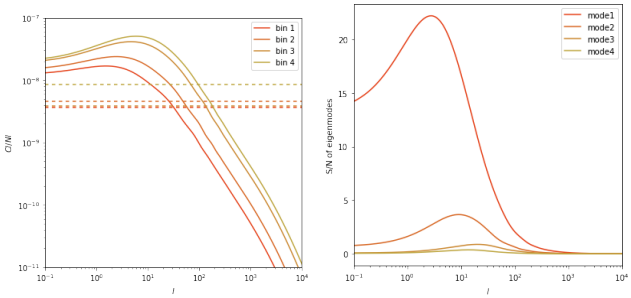

FIG. 10. **Left:** Sheer power spectrum of DESY1. Solid lines are diagonal elements of the signal matrix S_l, and dashed lines are the diagonal elements of noise matrix N_l; **Right:** Signal to noise ratio matrix D_l of KL-modes of the power spectrum on the left.

1. *Tomographic Compression*

With `CosmoSIS`, we calculated the shear power spectrum C_l of DES Year 1 with a fiducial cosmology at the best fit parameters. The l-range we consider is $2-5000$. On the left of Figure 10, we show the diagonal elements of the signal part and the noise part of C_l.

On the right of Figure 10 shows the KL-transformed eigenmode D_l of C_l. We can see that the first KL mode contains most of the S/N contrinbution to the power spectrum. If we want to recover most infomation, we can also include the second mode.

In Figure 11, we plot the normalized KL-eigenmode E_l^i of C_l and its corresponding $W_l^{ij} = E_l^i E_l^j$. Modes with different l are plotted with different depth of the

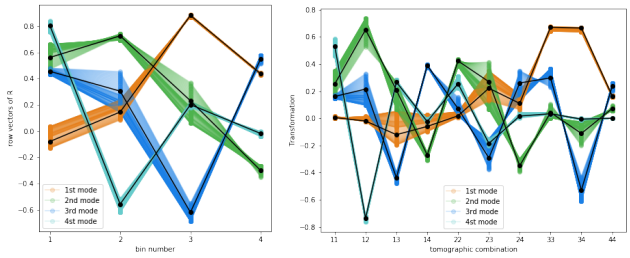

FIG. 11. **Left:** Normalized KL-eigenmodes E_L^P of the shear power spectrum C_l, darkness of the color is representing different l; **Right:** Transformation on tomographic bin combination W_{ij} constructed by the KL-eigenmodes. Black lines are the weighted average of each mode.

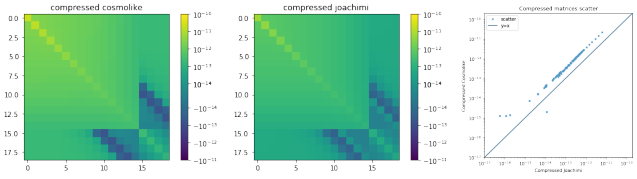

FIG. 12. **Left:** CL covariance matrix compressed by the first KL-mode; **Mid:** BJ covariance matrix compressed by the first KL-mode;**Right:** One-to-One scatter of the two compressed matrices

color. We can see that the KL-modes do not depend a lot on scale factor l by a large portion, so we take the weighted average of the eigenmodes E_l^p and its quadratic form W_l over ls and plot them with black lines.

For different l, the KL-modes do varies by a slight amount. We also observed that tomographic bins with higher redshift gains more weight than those with low redshift. This is also shown by the weight on tomographci combination that the combination of bin 3 and bin 4 gains most of the weight to maximize signal to noise ratio. This agrees with the fact that the diagonal elementsCl for low redshift is much less than those with high redshift. *Is my N_l actually right?*

We take the first KL mode and compressed the tomographic combination for each angle θ. As a result, 190 data points for 19 angles are compressed into 19 numbers. We then shrink the CL and BJ covariance matrices with the same compression scheme generated. The compressed CL and BJ covariance matrices are plotted in Figure 12

If we now again make a one-to-one comparison of these two compressed covariance matrices. We can easily notice that the huge clump at the bottom right for scatter of original matrices, which represent elements with great difference, are gone in the compressed covariance. Instead, the two covariance matrices just have a relative constant difference because of the different model they use. This shows that the divergence between CL and BJ covariance does not affect the overall signal to noise ratio a lot.

E. Constraints of Tomographic Compression

F. Tolerence Testing

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 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] Bellini, E. 2019
[5] Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601
[6] Krause, E., & Eifler, T. 2017, Mon. Not. Roy. Astron. Soc., 470, 2100
[7] Khlinger, F., et al. 2017, Mon. Not. Roy. Astron. Soc., 471, 4412
[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