Data Compression and Covariance Matrix Inspection: Cosmic Shear

Tassia Ferreira[1,2], Tianqing Zhang[1], Nianyi Chen[1], Scott Dodelson[1]
[1] *Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA*
[2] *PPGCosmo, Universidade Federal do Espirito Santo, Avenida Fernando Ferrari 514, 29075-910 Vitoria, Espirito Santo, Brazil*
(LSST Dark Energy Science Collaboration)
(Dated: April 24, 2020)

In this paper, we start along the path of comparing covariance matrices for cosmic shear statistics generated by two different codes. The main goal is to identify the parts of the covariance matrix that are most significant to parameter estimation, and therefore which ones should be calculated more accurately. We engage different ways of doing this, starting with a simple one-to-one comparison of the elements, then moving on to eigenvalues and finally to the signal-to-noise ratio (SNR). In the spirit of reducing the number of relevant elements, we remove 200 modes associated with the highest eigenvalues, then those with the lowest SNR. We find that it is not possible to locate the most important elements using the first method but, while the analysis with the SNR proved resourceful for a few parameters of interest, like Ω_m, we lost constraining power on the intrinsic alignment (IA) parameters as well as S_8. We also tested ways to shrink the covariance matrix, both at the tomographic level, and for the two-point functions. The former was accomplished using a method based on the Karhunen-Loéve (KL) decomposition to obtain the modes with highest SNR, and we show that, in order to reproduce compatible results, we need at least two KL-modes, but just like in the SNR case, this is not the case for all parameters. Finally, we apply a lossless compression scheme to the covariance matrix, capable of reducing the dimension to the number of free parameters. We were successful in reproducing the constraints obtained previously and show that the elements of the new matrix have an error tolerance of up to 25%.

I. INTRODUCTION

Cosmic shear is a weak lensing effect caused by the large-scale structure of the universe, and is an important tool for constraining cosmology. Here, we will deal with its covariance matrix, which is an essential component in the analysis of the cosmic shear data. For a data vector of length N, the covariance matrix is a symmetric $N \times N$ matrix with $N \times (N+1)/2$ individual elements that capture the auto and cross-correlation of the data vector. As data sets increase, the number of elements in the covariance matrix grows quadratically and becomes harder to analyse. One could potentially speed up computations and provide a simpler method of analysing the covariance matrix by using compression schemes capable of significantly reducing the size of the matrix while still retaining relevant information about the parameters of interest. One way of obtaining this is to use the Massively Optimized Parameter Estimation and Data (MOPED) algorithm, in which, if the noise in the data does not depend on the model parameters, the compression is truly lossless [8, 18]. MOPED has been widely used in literature for a variety of topics, like, for example, analysing CMB data [20], for redshift space galaxy power spectrum and bispectrum [7], for parameter-dependent covariance matrices [9], and, more recently, for compressing the Planck 2015 temperature likelihood [15].

We will focus on cosmic shear measurements from the Dark Energy Survey (DES) [19] Year 1 release; the data vector has 227 elements, varying with angular separation, and different pairs of tomographic redshift bins. Since our parameter space consists of 16 free parameters, we can use MOPED to reduce the $25,878$ independent elements of the covariance matrix, to only 136.

Apart from MOPED, we will also be employing other compression methods, like "discarting" the modes associated with the highest eigenvalues, then those with the lowest signal-to-noise ratio. In an effort to reducing the covariance matrix to about 10% its original size, we remove 200 such modes.

Finally, we apply the map-level compression [2], where linear combinations of the tomographic maps are used to retain as much information as possible. Compression of the tomographic bin pairs then considerably reduces the size of the data vector of the two-point functions. For example, we will see that most of the information in the four tomographic bins used by DESY1 can be compressed into a single linear combination of those bins. Therefore, instead of $(4 \times 5)/2$ two-point functions for each angular bin, we need include only one. For this purpose, the tomographic bins will have the same length, and so the angular cuts to the dataset and covariance matrix will be different from the ones used in the aforementioned DESY1 papers. The resulting covariance matrix has a dimension of 190×190, and so, for one eigenmode, the compressed matrix will have 190 independent elements.

The main goals of this paper are as follows:

- Apply compression schemes to the DESY1 Cosmic Shear.
- Use MOPED for comparing two covariance matrices.
- Introduce noise to the elements in order to test their tolerance by quantifying responses in the likelihood

analysis.

In §II, we start by describing the dataset and the pair of covariance matrices used. We then proceed to review each compression scheme and apply them to DESY1 Cosmic Shear. We validate our findings by comparing them with the constraints obtained with the unmodified covariance matrix. The matrices compressed using MOPED are compared in §III. Our tolerance test is described in §IV, where we compare what happens to the parameter constraints when we introduce noise to elements and eigenvalues separately. Finally, our conclusions are summarised in §V.

II. METHODS

A. DES Cosmic Shear: Data and Analysis

In this section, we introduce the data and covariance matrices that are used in this work. Our tests are carried out using cosmic shear statistics $\xi_{\pm}(\theta)$, focusing on the Year 1 results of the Dark Energy Survey [1, 19] (DESY1). The data is divided into four tomographic redshift bins spanning the interval $0.20 < z < 1.30$, which yields 10 bin-pair combinations, each one containing 20 angular bins between 2.5 and 250 arcmin. We thus begin with 200 data points for each $\xi_+(\theta)$ and $\xi_-(\theta)$, giving 400 in total. We then apply the angular cuts described in [1], which removes the scales most sensitive to baryonic effects; this leaves 167 points for $\xi_+(\theta)$ and 60 for $\xi_-(\theta)$, resulting in 227 data points corresponding to the aforementioned 227×227 covariance matrix.

Table I shows the 16-parameters varied and the priors placed on them. To perform cosmological parameter inference we use the `CosmoSIS` [3, 5, 10–12, 16, 17, 21] pipeline, while employing the `MultiNest` [6] sampler to explore the parameter space, with 1000 `livepoints`, `efficiency` set to 0.05, `tolerance` to 0.1 and `constant efficiency` set to True.

Parameter	Prior
Cosmological	
Ω_m	$\mathcal{U}(0.1, 0.9)$
$\log A_s$	$\mathcal{U}(3.0, 3.1)$
$H_0(\mathrm{kms}^{-1}\mathrm{Mpc}^{-1})$	$\mathcal{U}(55, 91)$
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_{IA}	$\mathcal{U}(-5, 5)$
η_{IA}	$\mathcal{U}(-5, 5)$
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 ($\mathcal{U}$ denotes flat in the given range and $\mathcal{G}$ is gaussian with mean equal to its first argument and dispersion equal to its second). For the cosmological parameters, we fix $w = -1.0$, $\Omega_k = 0.0$ and $\tau = 0.08$. The astrophysical parameters are associated with the intrinsic alignment, they follow the relation $A(z) = A[(1+z)/1.62]^\eta$. 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 covariance matrices used are the following:

- one obtained using `Cosmolike` [14] (CL), which was also used in the initial DESY1 analysis;
- one containing only the Gaussian part, obtained by running the same code used to analyse the KiDS-450 survey [13], using DESY1 parameters and tomographic bins. We refer to it simply as the Gaussian covariance matrix.

Thus, throughout, the covariance labels CL and Gaussian differ for several reasons: first, they are two independent codes and, second, although the code for the KiDS-450 survey does contain all the functionality in CL, we ran with the simplest settings in order to accentuate the differences, as a result, the resulting matrix does not contain the non-Gaussian parts. The ensuing larger differences will help us assess various validation techniques. Where not otherwise stated, the analysis and constraints will be performed on the CL covariance matrix.

Figure 1 shows the results for the projected cosmological constraints for CL and the Gaussian covariance matrix, using the same data vector and cuts. The 2σ constraints are as follows: for CL: $\Omega_m = 0.306^{+0.073}_{-0.060}$, $A = 0.852^{+1.005}_{-1.086}$ and $S_8 = 0.784^{+0.200}_{-0.171}$; and for the Gaussian one: $\Omega_m = 0.309^{+0.073}_{-0.058}$, $A = 0.948^{+0.916}_{-0.985}$ and $S_8 = 0.787^{+0.196}_{-0.166}$. This shows that the differences we have introduced to the calculation of the two matrices are mensurable in the parameter constraints.

B. Eigenvalues

Let us start with the easy task of analysing the eigenvalues of the covariance matrix. Each eigenvalue is associated with a linear combination of the data vector, or a *mode*.

The lowest eigenvalues correspond to modes with the smallest variance but since they are not normalised, it is unclear how this variance compares to the signal in the mode. Let us nonetheless explore the possibility that the modes with the lowest variance provide the most information and therefore dropping the ones with the largest

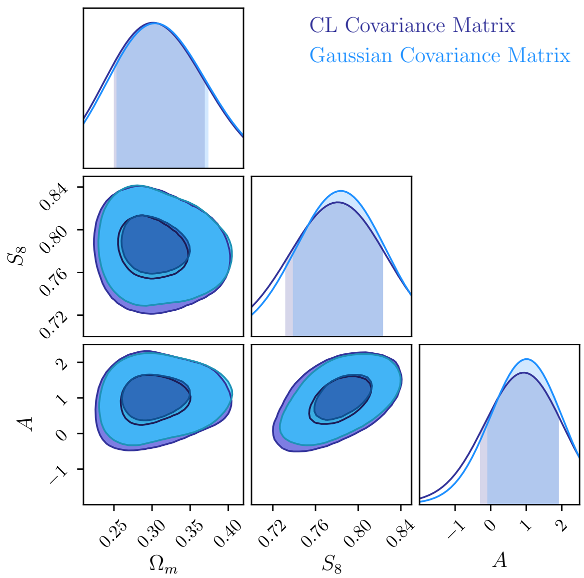

FIG. 1. Constraints on cosmological parameters Ω_m and S_8 and intrinsic alignment parameter $A_{\rm IA}$ for two covariance matrices produced for cosmic shear. The purple curve is for CL while the blue is for the Gaussian. In the 16 parameter space, the volume of the posterior is about 22% larger for the former.

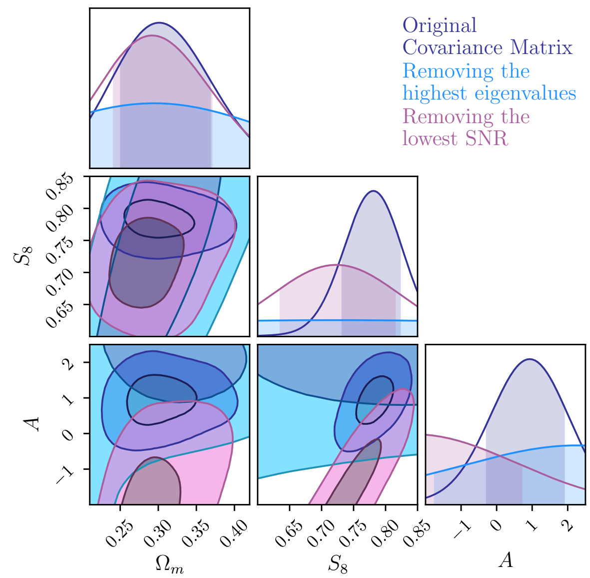

FIG. 2. Constraints on cosmological parameters Ω_m, S_8 and the intrinsic alignment parameter $A_{\rm IA}$ for the original CL covariance matrix (in purple) and for two new covariance matrices obtained by setting the 200 highest eigenvalues of the original matrix to nine orders of magnitude higher (in blue), and by replacing the 200 lowest values of the SNR to seven order of magnitude lower (in magenta).

eigenvalues would not affect the final result.

Our procedure consists of first diagonalising the covariance matrix in order to calculate its eigenvalues and then replacing the large eigenvalues with a larger number (nine orders of magnitude higher), thus removing their effective contribution; we then transform back to the original basis and perform a cosmological analysis with the new covariance matrix, to constrain the parameters of our model.

In the spirit of reducing the covariance matrix to about 10 % its original size, we follow the procedure above to discard the 200 eigenmodes with the largest eigenvalues, Figure 2 shows the results obtained. The constraints are significantly broader for the three parameters shown. This is consistent with the fact that we are throwing away about 90% of the information. However, it is inconsistent with the notion that these modes are irrelevant, in fact, constraints on $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ for the original covariance matrix are $0.784^{+0.200}_{-0.171}$, whereas, for this procedure, we obtain $0.679^{+0.533}_{-0.505}$, showing an increase in almost 200%. It is then clear that a different way of ordering the modes, other than simply looking at the eigenvalues, is called for.

C. Signal-to-noise ratio

Instead of looking only at the "noise" – or the eigenvalues of the covariance matrix – a better way to assess the importance of modes is to consider the signal as well. We can define the expected signal-to-noise ratio (SNR) as

$$\left(\frac{S}{N}\right)^2 = \sum_{ij} T_i C^{-1}_{ij} T_j \ , \qquad (1)$$

where T_i and are the predicted theoretical signal for the i^{th} data point and C is 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 just computing T_i^2/C_{ii}. 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

$$\left(\frac{S}{N}\right)^2 = \sum_i \frac{v_i^2}{\lambda_i}, \qquad (2)$$

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

$$v_i \equiv U^T_{ij} T_j \ . \qquad (3)$$

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 the covariance matrix, we proceed to set the 200 lowest values to seven orders of

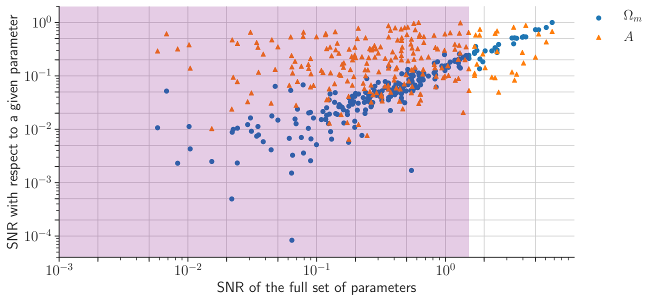

FIG. 3. Scatter plot for the relation between the signal to noise obtained with the covariance matrix for CL for each parameter (x-axis) against that for the full set of parameters (y-axis). The derivatives are shown with respect to Ω_m (blue circle) and for the intrinsic alignment parameters $A_{\rm IA}$ (orange triangle). The purple rectangle spreads until the two hundredth lowest value of SNR, which corresponds to the values that are modified for parameter constraints.

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.

The parameter constraints for this method are shown in Figure 2, where we note that only Ω_m is well constrained (in agreement with those obtained with the original covariance matrix to within a 2σ interval). The constraining power on $A_{\rm IA}$ and S_8, on the other hand, is very much lost, which suggests that the modes removed do indeed carry relevant information for these parameters.

We can investigate this loss by tweaking our understanding of which modes carry information. The "signal" these modes are ordered by is the amplitude of the data points. The parameters , however, are sensitive to the shape as well as the amplitude. To address this, we can identify the SNR for each parameter individually by taking

$$\left(\frac{\partial S/\partial p_\alpha}{N}\right)^2 = \sum_i \frac{(\partial v_i/\partial p_\alpha)^2}{\lambda_i} \tag{4}$$

where $\partial/\partial p_\alpha$ is the derivative with respect to each parameter. This produces the SNR for each parameter of interest. The importance of this procedure is illustrated in Figure 3, which shows the normalised vanilla SNR for a given mode on the x-axis and the SNR for each parameter for the same mode, for brevity, we show only for Ω_m and $A_{\rm IA}$. The shaded region is the one excluded in the previous analysis, but clearly there are some low SNR modes there that contain information about the parameters. This is particularly true for the intrinsic alignment parameter $A_{\rm IA}$, which seems to explain the poor constraints shown in Figure 2. As a result, simply cutting on raw SNR loses constraining power.

D. Tomographic Compression

This compression method is based on Karhunen-Loéve (KL) decomposition for the shear power spectrum suggested by [2] and later applied to real space two-point function in [4] for CFHTLens survey. This method generally finds the eigenmode with most of the signal-to-noise ratio contribution to the power spectrum, and then transforms the two-point function in real space based on this eigenmode.

With `CosmoSIS`, we can generate the shear power spectrum C_ℓ^{ij} of convergence $a_{\ell m}$ for a fiducial cosmology. The cosmology we choose is the best-fit value of the DES Year 1 results for cosmic shear only. With the shear power spectrum $C_\ell = S_\ell + N_\ell$ and its shape noise N_ℓ, we can calculate the Karhunen-Loéve (KL) modes matrix E_ℓ via a general eigenvalue problem

$$C_\ell^{ij} E_\ell^j = \lambda_\ell^i N_\ell^{ij} E_\ell^j . \tag{5}$$

Each row in E_ℓ corresponds to a KL-mode of C_ℓ. Using the Cholesky decomposition, $N_\ell = LL^\dagger$, the new observable can be expressed as $b_{\ell m} = E_\ell \cdot L^{-1} a_{\ell m}$. We should note that C_ℓ is the angular power spectrum of the weak lensing shear, and E_ℓ is similar to a transformation of basis for the shear. So, we can now calculate the power spectrum D_ℓ for the new observable $b_{\ell m}$

$$D_\ell = \langle b_{\ell m} b_{\ell m}^T \rangle = E_\ell L^{-1} C_\ell L^{-1} E_\ell^T , \tag{6}$$

or, if we denote $E_\ell N^{-1}$ as R_ℓ and further denote $U_\ell^{ij} = R_\ell^i R_\ell^j$, we can write the compression in one simple linear combination of the C_ℓ,

$$D_\ell = R_\ell^i C_\ell^{ij} R_\ell^j = U_\ell^{ij} C_\ell^{ij} . \tag{7}$$

The double summation U_ℓ^{ij} is a weight on the tomographic bin-pair, which we can later use to compress the two-point functions. We should point out that these KL-modes are uncorrelated, so the power spectrum of the new observable D_ℓ is a diagonal matrix, with 1+SNR of the corresponding eigenmodes on the diagonal elements. Since the KL-decomposed modes of shear power spectrum are uncorrelated, we can make a compression here by taking only the first one or two modes with the highest SNR. By doing so, we compress ten tomographic bin-pairs to one or two.

We want, however, to eventually compress the two-point function data vector of DESY1 One possible way is to calculate the two-point function of the KL mode of the shear power spectrum,

$$\xi_\pm^{ij}(\theta) = \int \frac{\ell d\ell}{2\pi} J_{0/4}(\ell\theta) C^{ij}(\ell) . \tag{8}$$

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

$$\begin{aligned}\tilde{\xi}_\pm(\theta) &= \int \frac{\ell d\ell}{2\pi} J_{0/4}(\ell\theta) D(\ell) \\ &= \int \frac{\ell d\ell}{2\pi} J_{0/4}(\ell\theta) U^{ij} C^{ij}(\ell) \\ &= U^{ij} \xi^{ij}_\pm(\theta) \ , \end{aligned} \tag{9}$$

where U^{ij}, the ℓ-independent compression weight is given by

$$U^{ij} = \frac{\int_{\ell_{\min}}^{\ell_{\max}} (2\ell+1) U^{ij}_\ell}{\int_{\ell_{\min}}^{\ell_{\max}} (2\ell+1)} \ . \tag{10}$$

We make a more conservative angular cut than the one discussed in [19], making sure that both $\xi_\pm$ are uniform in regard to tomographic combinations. We consider an angular scale for ξ_+ from 7.195° to 250.0°, and for ξ_- from 90.579° to 250.0°. Therefore, for the purpose of demonstrating 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, and so the number of elements in the covariance matrices are reduced by 99%.

With `CosmoSIS`, we calculate the shear power spectrum C_ℓ of DES Year 1 with a fiducial cosmology at the best-fit parameters, and ℓ-range $2-2500$. The left plot in Figure 4, shows the diagonal elements of the signal part and the noise part of C_ℓ, while the right one shows the KL-transformed eigenmode D_ℓ of C_ℓ. We can see that the first KL mode contains most of the SNR contribution to the power spectrum. However, if we want to recover more information, we also should include the second and the cross mode between the first and second KL-mode.

In Figure 5, we plot the normalised KL-eigenmode E^i_ℓ of C_ℓ and its corresponding $U^{ij}_\ell = R^i_\ell R^j_\ell$. Modes with different ℓ are plotted in increasing shades of the colour. We can see that the KL-modes do not depend significantly on the scale factor ℓ, so we also take the weighted average of the eigenmodes E^p_ℓ and its quadratic form U_ℓ over ℓ's and plot them with black lines. We see that for different ℓ, the KL-modes do vary by a slight amount. For the first KL-mode, the tomographic bins with higher redshift gain more weight than those with low redshift. This is also shown by the weight on tomographic combination that the combination of bin 3 and bin 4 gains most of the weight to maximise the signal-to-noise ratio. This agrees with the fact that the diagonal elements C_ℓ for low redshift are much less than those with high redshift.

We ran the likelihood analysis with the first KL-mode and the first two KL-modes, which correspond to a 10-to-1 and 10-to-3 compression, respectively, and show the parameter constraints on the $\Omega_m - S_8 - A$ plane in Figure 8. We can see in the figure that the first KL-mode is generally not enough to recover the information in the data vector. Since the first two modes contain most of the SNR contribution on a map level, we were able to recover the Ω_m constraints. However, the recovery on the S_8 and $A_{\rm IA}$ is not desirable. This could be due to the fact that the SNR-prioritised modes are not the sensitive direction for these parameters, as was also the case in Figure 3. Indeed, the $S_8 - A$ plane shows a strong correlation between these two parameters. This could explain why the S_8 constraints got wider: the KL-modes fail to break the degeneracy on $A_{\rm IA}$, which is mostly contained in the modes that are insensitive to cosmic shear, and are discarded in the compression process.

E. Linear combinations of the data vector

The compression here takes place at the two-point level [20], with the compressed data vector containing linear combinations of the many two-point functions. In principle, this works with only N_p two-point functions where N_p is the number of free parameters, 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

$$y_\alpha = U_{\alpha i} D_i \ , \tag{11}$$

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

$$U_{\alpha i} \equiv \frac{\partial T_j}{\partial p_\alpha} {C^{-1}}_{ji} \ , \tag{12}$$

with T_j being the theoretical prediction for the data point D_j for a fiducial cosmology. An illustration of the matrix $U_{\alpha i}$ is shown in Figure 10, showing the weighting vector for parameters Ω_m, S_8 and $A_{\rm IA}$.

The now much smaller data set $\{y_\alpha\}$, which contains as few as N_p data points, carries its own covariance matrix, with which 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

$$C_{\alpha\beta} = U_{\alpha i} C_{ij} U^T_{j\beta} \ , \tag{13}$$

which also happens to be identical to the Fisher matrix of our likelihood. This compression was first suggested by Tegmark et al. (1997), and is the basis for the MOPED algorithm [8], which extended the original procedure for multiple parameters. Note that our weighing vector given by Eq. 12 does not carry the normalising factor used in the aforementioned paper, nor are they independent of each other. We find that this does not, however, impact the analysis.

In our case, the covariance matrix is 227×227, 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 $(227 \times 228)/2 = 25,878$ independent elements of the covariance matrix a small subset (only 136 in this case) of linear combinations of

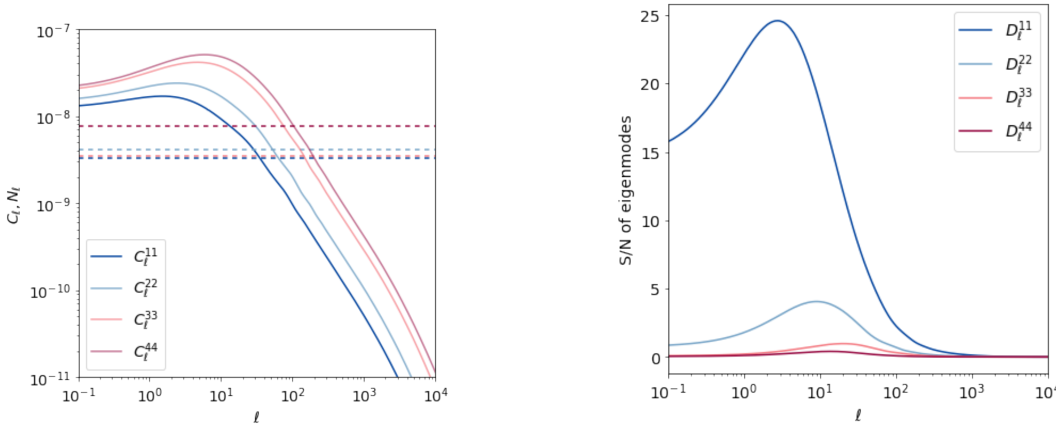

FIG. 4. **Left:** Shear power spectrum of CL. Solid lines are diagonal elements of the signal matrix S_ℓ, and dashed lines are the diagonal elements of noise matrix N_ℓ. **Right:** Signal-to-noise ratio matrix D_ℓ of KL-modes of the power spectrum on the left.

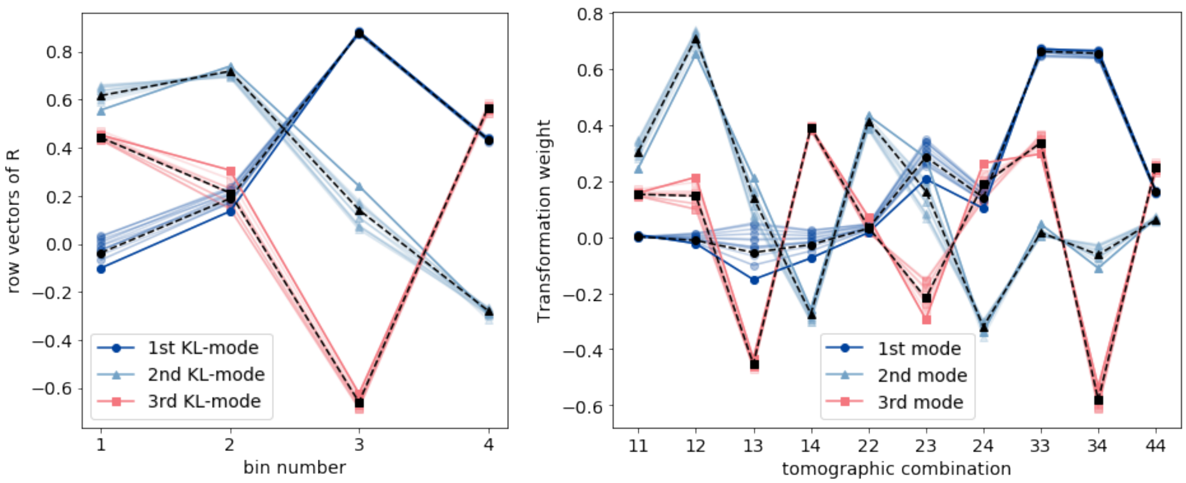

FIG. 5. **Left:** Normalised KL-eigenmodes E_ℓ^p of the shear power spectrum C_ℓ, the changes in shades represent different ℓ. **Right:** Transformation on tomographic bin combination U_{ij} constructed by the KL-eigenmodes. Black lines are the weighted average of each mode.

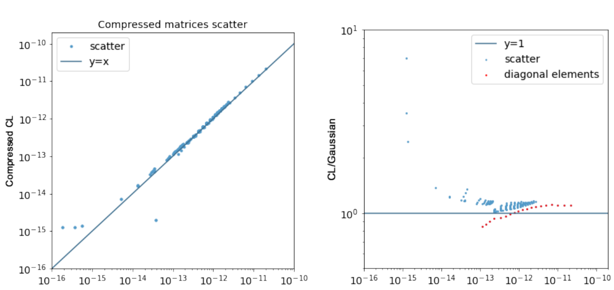

FIG. 6. **Left:** One-to-one scatter of the two compressed matrices following the procedure described in §**??**. **Right:** One-to-one scatter of the ratio of CL over Gaussian and the CL elements.

these 25k elements that really matter. If two covariance matrices give the same set of $C_{\alpha\beta}$, it should not matter whether any of the other thousands of elements differ from one another.

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 χ^2 using two different covariance matrices.

Figure 9 compares the constraints obtained for the compressed covariance matrix and data set with results from the full one. The two curves agree extremely well for the parameters shown: Ω_m, S_8 and $A_{\rm IA}$. This is also true for all the other cosmological and intrinsic alignment parameters, where their mean values agree at the 2σ confidence level. While the volume of the whole constrained parameter space does increase by about 13%, the constraints for most parameters are less than 4% broader, which shows that the information loss is negligible.

One relevant point in this analysis is at which point to

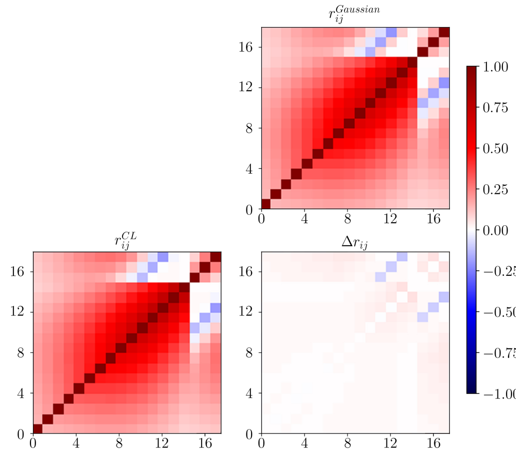

FIG. 7. The correlation matrix of Gaussian (upper right) and CL (bottom left) covariance matrices, and their difference (bottom right) in the elements

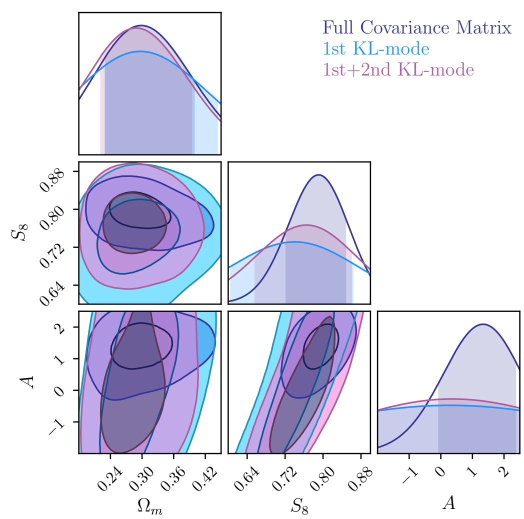

FIG. 8. Cosmological constraints marginalised over all 16 parameters for the 190 × 190 CL covariance matrix and that compressed by the first KL-mode and the first two KL-modes.

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 is taken.

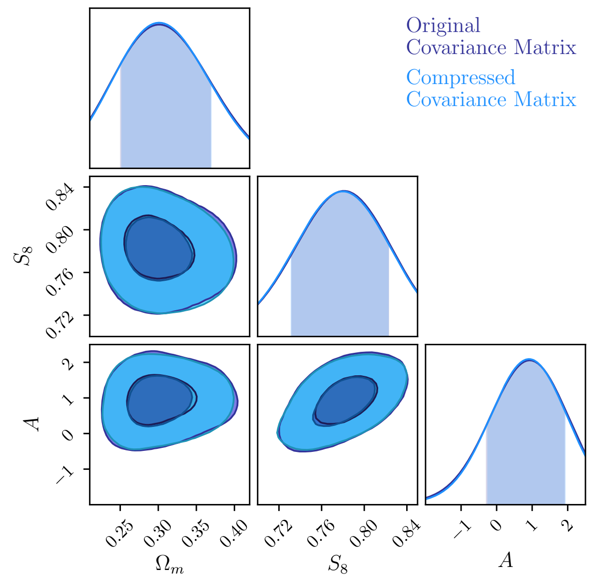

FIG. 9. Constraints on cosmological parameters Ω_m and S_8 and for the intrinsic alignment parameter $A_{\rm IA}$ for the original CL covariance matrix (in purple) and for the compressed one (in blue).

III. COMPARISON OF COVARIANCE MATRICES

A. Element-by-element comparison

Let us start by performing an element-by-element comparison between the two covariance matrices. If there were only a single data point, then the covariance matrix would be one number and comparing two covariance matrices to try to understand why they give different constraints would be as simple as comparing these two numbers. The simplest generalisation is then to do an element-by-element comparison. We make a scatter plot of the elements of the two matrices in the bottom panel of Figure 11, where we can see that the elements of CL are, in general, larger than the Gaussian's, differing by up to 4 orders of magnitude. In some ways, this is useful and reassuring, as it aligns with what we see in the parameter constraints, in Figure 1: larger elements in the covariance matrix translates to less constraining power.

The limitation of this method is that it remains unclear which of the differences are driving the final discrepancies in parameter constraints. This difficulty is an outgrowth of the increasing size of the data sets and hence the growing number of elements of the covariance matrix that any two codes are likely to disagree on. This element-by-element comparison, however, would prove much more useful if we fewer elements to compare. Towards that end, we turn to compressed covariance matrices.

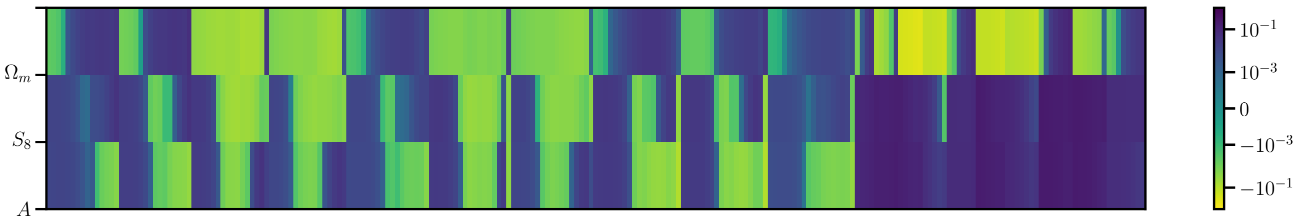

FIG. 10. An illustration of the 227 values of the weights corresponding to Ω_m, S_8 and $A_{\rm IA}$ used for compressing the covariance matrices. Note how similar the weighing vectors for S_8 and $A_{\rm IA}$, and that the largest values for correspond to the last 60 elements, i.e. these will be used to compress the part of the covariance matrix that holds information for ξ_-.

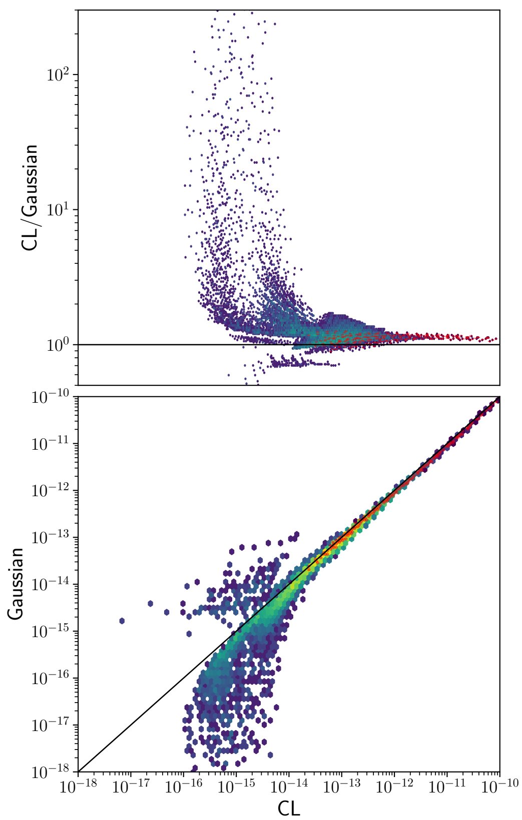

FIG. 11. In both plots, the red points refer to the diagonal elements, and the colour bar varies according to the number of elements in one hexagonal bin, where the darkest blue colour corresponds to only one element, and the brightest yellow shade to 2000. **Top:** Scatter plot of the ratio of the elements of CL and the Gaussian one vs the Gaussian value. For illustrative purposes, we draw a black, horizontal line at CL/Gaussian = 1. **Bottom:** Density of the scatter plot of the positive elements of the covariance matrices Gaussian and CL, with the black line showing $x = y$.

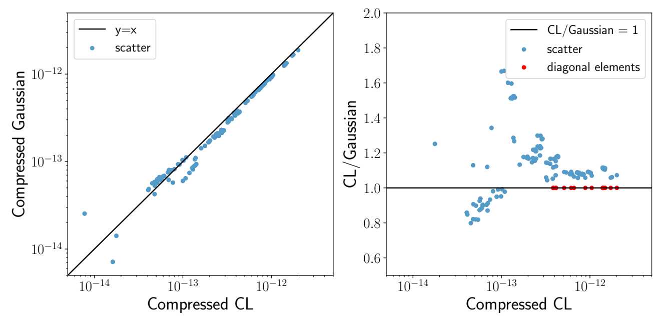

FIG. 12. **Left:** One-to-one scatter of the two compressed matrices following the procedure described in §**??**. **Right:** One-to-one scatter of the ratio of CL over Gaussian and the CL elements.

B. Compressed Matrices Comparison

1. *Using MOPED*

Here, we take two different approaches: first, we assume that $U_{\alpha,i}$ is the same for both covariance matrices and we calculate it with CL. 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. We find that the mean values of the parameter constraints for the two methods agree to 1σ, which shows that they are equivalent to each other. Figure 13 is obtained for the first method, which will be the one adopted from here on, it shows the correlation matrix for Gaussian and CL, and the difference between the diagonal elements. 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.

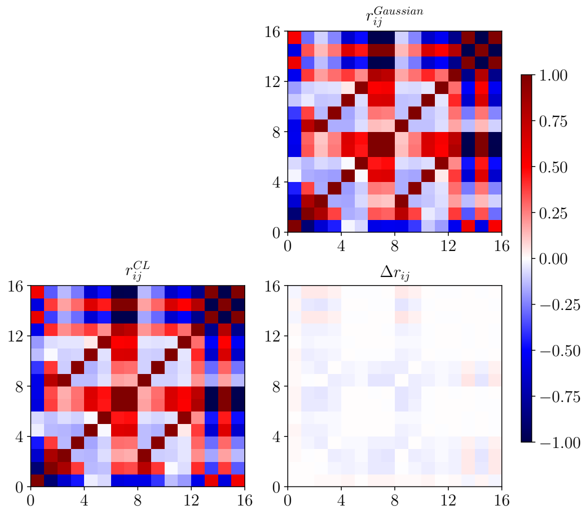

FIG. 13. The upper right and lower left plots display the correlation matrix for Gaussian and CL respectively, and the difference between them, Δr_{ij}, is shown on the lower right.

2. *Using map-level compression*

In Figure 6, we plot the compressed covariance matrices for the CL and the Gaussian covariance with the first mode only, and show a one-to-one comparison of the covariance matrices. By comparing the bottom panel of Figure 11 with the left panel of Figure 6, we notice that the large regions containing the elements with greater difference are now gone. Instead, the two covariance matrices just have a relative constant difference, because of the fact that we did not include non-gaussian effect in one of them. This shows that the divergence between CL and the Gaussian covariance does not considerably affect the overall SNR.

IV. TOLERANCE OF THE COMPRESSED MATRICES

Now that we have shown that we are indeed able to compress the covariance matrix into a much simpler and considerably smaller one, our next step is to analyse the amount of error the elements can tolerate while reproducing compatible parameter constraints. In the next sections we test two different ways of perturbing the covariance matrix: first we consider an error to the elements themselves, and then we follow a similar procedure to study the effects of introducing error to the eigenvalues.

One of the issues that arises when arbitrarily modifying the elements of the covariance matrix is that the new one does not necessarily remain positive definite. In this analysis, we take an extra step to ensure that this characteristic is retained.

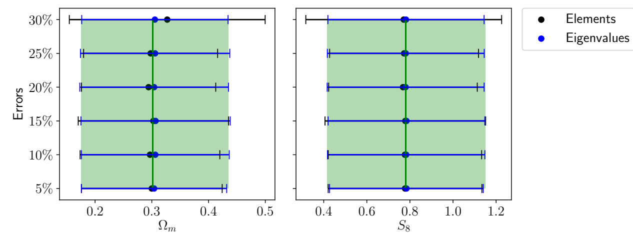

FIG. 14. An error plot showing the changes to the constraints for Ω_m and S_8 for errors added at 5%, 10%, 15%, 25% and 30% of the original elements (in black) and eigenvalues (in blue) of the compressed covariance matrix. The green rectangle covers the 2σ interval obtained for the original CL covariance matrix, and the green line shows the mean value for the respective parameter.

A. Modifying the elements

To quantify the error tolerance of the elements of the covariance matrix, we introduce error in the following manner: consider that we want to test the impact of an error $x\%$; this can either be an increase of a decrease in the original element, such that what we care about most is not whether the parameter constraints will be larger, but rather how different. For this error to be random, but centred at our desired percentage, we draw δ from a Gaussian distribution, $\mathcal{G}(0, \frac{x}{100})$ and calculate the new value to be

$$C^{\rm new}_{\alpha\beta} = (1+\delta)C^{\rm old}_{\alpha\beta} \ . \tag{14}$$

We keep the matrix symmetric by making $C_{\alpha\beta} = C_{\beta\alpha}$, and, finally, we check for positive definiteness. We show the constraints on Ω_m and S_8 in Fig. 14, in black, where the green rectangle spans over the constraints for the original covariance matrix. We see that errors of up to 25% translate to $< 10\%$ difference in the constraints. A 30% error, on the other hand, shows differences of up to 33% and 24%, respectively. It is worth noting here that we also find that S_8 is less sensitive to these noise introduced.

B. Modifying the eigenvalues

Another way of introducing error to the covariance matrix is to perturb its eigenvalues. For a symmetric matrix, we have

$$C_{\alpha\beta} = Q\Lambda Q^{-1} \ , \tag{15}$$

where $\Lambda = \lambda I$, with λ being the eigenvalues and I the identity matrix; and Q is a square matrix whose columns are composed of the eigenvectors of $C_{\alpha\beta}$. The eigenvalues are then perturbed as described in Eq. 14, with an extra step to guarantee $\delta > -1$, then $\lambda^{\rm new} > 0$, thus keeping the matrix positive definite.

The results for this method are also plotted in Fig. 14, in blue, for Ω_m and S_8. In general, we find that we are not able to reproduce significant changes to the parameter constraints, even with 30% errors, as the resulting constraints are all within 5% of the original one. As such, it is not clear that the results obtained using this procedure is equivalent to modifying the actual elements of the covariance matrix.

V. CONCLUSION

In this work we set out to explore efficient ways of comparing, analysing and compressing covariance matrices. We started off looking at the parameter constraints of two 227×227 covariance matrices CL and Gaussian, generated for DESY1 cosmic shear measurements, and saw that, although some of their elements differed by several orders of magnitude, they generated similar constraints. It was clear, then, that not all elements contribute equally to the parameter constraints, and we needed to employ increasingly complicated methods to try and locate the most relevant parts of the covariance matrix.

The first step was then to analyse the eigenvalues. We began with the hypothesis that modes associated with the lowest eigenvalues have the lowest variance and therefore carry most information, as such, those with the highest eigenvalues would contribute less to parameter estimation. This proved to be untrue: "removing" the highest 200 eigenvalues, by setting them to nine orders of magnitude higher resulted in a loss of about 200% on the constraining power. Next, we moved on to the signal-to-noise ratio, and, using a similar procedure adopted for the eigenvalues, we "removed" the modes with the lowest SNR. The results showed us that these modes did not contribute significantly to constraining some cosmological parameters, like Ω_m, but constraints on the intrinsic alignment parameters, and even S_8 were considerably affected. This is consistent with the fact that the IA parameters are more sensitive to low SNR, and it shows us that we need to look at the SNR per parameter before making any cuts, so that we do not lose important information for the parameters that we want to constrain.

Finally, we explored methods of shrinking the covariance matrix. We explored two such methods: a tomographic compression, and another directly on the two-point functions. For the first method, we decompose the shear power spectrum into KL modes, then we look for those with the highest SNR. We thus go from ten tomographic bins to only one or two. The resulting covariance matrix, for one mode, is then reduced from 190×190 to 19×19 or 59×59, showing a reduction of about 99% or 91%, respectively. We show, however, that one mode is not sufficient for constraining the parameters of our model, with the results being similar to our previous tests involving SNR: the constraints for Ω_m, for example, are reproduced with the first and second KL-mode, but this is not the case for the IA parameters. Since essential information of IA parameters is contained in low SNR KL-mode, the high KL-modes failed to break the degeneracy of $A - S_8$ correlation, resulting in wider S_8 constraints. As for the second compression scheme, we use linear combinations of the data vector. By transforming the data vector and covariance matrix with a weighting vector that is parameter dependent, we were able to reduce the 227×227 matrix to a 16×16, and since the Fisher matrix is identical for both the original and compressed ones, the compression scheme is lossless. This is also clear in the parameter constraints, where we show that we are able to reproduce the similar constraints for the two matrices, for all parameters. On the other hand, we compared the elements of the compressed covariance matrix for CL and Gaussian and found that the new elements show reasonable agreement, with their correlation matrices being very similar, and the diagonal elements showing a percentage difference of less than 15%.

One last step was taken to analyse the error tolerance of the compressed covariance matrix. We first introduced random errors of $5\%, 10\%, 15\%, 20\%, 25\%$ and 30% and found that the constraints are consistent with the original ones for an error of up to 25%; errors larger than that cause the constraints to increase. We also applied these errors to the eigenvalues of the covariance matrix, and found the differences in the constraints to be less than 3%, even for those larger than 25%.

Acknowledgments

The author wish to thank Sukhdeep Singh and Hung-jin Huang for useful discussions.

T.F. and T.Z. contributed extensively writing the main paper as well as implementing the covariance comparison and compression. N.C. contributed to the compression code. All authors participated in the discussion and gave valuable suggestions.

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. T.F also acknowledges financial support from CAPES and FAPES.

[1] Abbott, T. M. C., et al. 2018, Physical Review D, 98, arXiv:1708.01530

[2] Alonso, D. 2018, Monthly Notices of the Royal Astronomical Society, 473, arXiv:1707.08950

[3] Antony Lewis, Anthony Challinor, A. L. 2000, The Astrophysical Journal, 538, arXiv:9911177

[4] Bellini, E., Alonso, D., Joudaki, S., & Waerbeke, L. V. 2019, arXiv:1903.0495

[5] Bridle, S., & King, L. 2007, New Journal of Physics, 9, arXiv:0705.0166

[6] Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, arXiv:0809.3437

[7] Gualdi, D., Manera, M., Joachimi, B., & Lahav, O. 2018, Monthly Notices of the Royal Astronomical Society, 476, arXiv:1709.03600

[8] Heavens, A. F., Jimenez, R., & Lahav, O. 2000, Monthly Notices of the Royal Astronomical Society, 317, arXiv:9911102

[9] Heavens, A. F., Sellentin, E., de Mijolla, D., & Vianello, A. 2017, Monthly Notices of the Royal Astronomical Society, 472, arXiv:1707.06529

[10] Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, Journal of Cosmology and Astroparticle Physics, 2012, arXiv:1201.3654

[11] Kilbinger, M., Benabed, K., Guy, J., et al. 2009, Astronomy & Astrophysics, 497, arXiv:0810.5129

[12] Kirk, D., Rassat, A., Host, O., & Bridle, S. 2012, Monthly Notices of the Royal Astronomical Society, 424, arXiv:1112.4752

[13] Köhlinger, F., et al. 2017, Monthly Notices of the Royal Astronomical Society, 471, arXiv:1706.02892

[14] Krause, E., & Eifler, T. 2017, Monthly Notices of the Royal Astronomical Society, 470, arXiv:1601.05779

[15] Prince, H., & Dunkley, J. 2019, Physical Review D, 100, arXiv:1909.05869

[16] Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, Monthly Notices of the Royal Astronomical Society, 341, arXiv:0207664

[17] Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, The Astrophysical Journal, 761, arXiv:1208.2701

[18] Tegmark, M., Taylor, A. N., & Heavens, A. 1997, The Astrophysical Journal, 480, arXiv:9603021

[19] Troxel, M. A., et al. 2018, Physical Review D, D98, arXiv:1708.01538

[20] Zablocki, A., & Dodelson, S. 2016, Physical Review D, 93, arXiv:1512.00072

[21] Zuntz, J., Paterno, M., Jennings, E., et al. 2015, Astronomy and Computing, 12, arXiv:1409.3409