

# A fast and reliable method for the comparison of covariance matrices

Tassia Ferreira <sup>1,2★</sup> and Valerio Marra <sup>2,3,4,5</sup>

<sup>1</sup>PPG Cosmo, Universidade Federal do Espírito Santo, 29075-910 Vitória, ES, Brazil

<sup>2</sup>Laboratório Interinstitucional de e-Astronomia – LIneA, 20921-400 Rio de Janeiro, RJ, Brazil

<sup>3</sup>Núcleo de Astrofísica e Cosmologia and Departamento de Física, Universidade Federal do Espírito Santo, 29075-910 Vitória, ES, Brazil

<sup>4</sup>INAF – Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy

<sup>5</sup>IFPU – Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy

Accepted 2022 May 3. Received 2022 April 1; in original form 2021 December 15

## ABSTRACT

Covariance matrices are important tools for obtaining reliable parameter constraints. Advancements in cosmological surveys lead to larger data vectors and, consequently, increasingly complex covariance matrices, whose number of elements grows as the square of the size of the data vector. The most straightforward way of comparing these matrices, in terms of their ability to produce parameter constraints, involves a full cosmological analysis, which can be very computationally expensive. Using the concept and construction of compression schemes, which have become increasingly popular, we propose a fast and reliable way of comparing covariance matrices. The basic idea is to focus only on the portion of the covariance matrix that is relevant for the parameter constraints and quantify, via a fast Monte Carlo simulation, the difference of a second candidate matrix from the baseline one. To test this method, we apply it to two covariance matrices that were used to analyse the cosmic shear measurements for the Dark Energy Survey Year 1. We found that the uncertainties on the parameters change by 2.6 per cent, a figure in agreement with the full cosmological analysis. While our approximate method cannot replace a full analysis, it may be useful during the development and validation of codes that estimate covariance matrices. Our method takes roughly 100 times less CPUh than a full cosmological analysis.

**Key words:** methods: statistical – cosmological parameters – cosmology: observations.

## 1 INTRODUCTION

Cosmology has entered its golden era, with the fast advancement of technology allowing us to build telescopes capable of exploring almost the entire observable universe. As we brace ourselves for the unprecedented amount of data that will be made available in upcoming surveys like the Javalambre Physics of the Accelerating Universe Astrophysical Survey,<sup>1</sup> the Vera C. Rubin Observatory Legacy Survey of Space and Time,<sup>2</sup> *Euclid*,<sup>3</sup> and the Square Kilometre Array,<sup>4</sup> we must tackle the issue of how to process and extract as much information as possible from the data.

This brings us to the issue of code development and validation (see e.g. Chisari et al. 2019; Blanchard et al. 2020). Here, we discuss covariance matrix validation, which is becoming increasingly important in cosmological analyses (Krause et al. 2017; Joachimi et al. 2020; Friedrich et al. 2021). Covariance matrices are vital pieces to the puzzle as they take into consideration not only the statistical and systematic errors of the measurement, but also the correlation between each quantity. The size of a covariance matrix grows as the square of the size  $N$  of the data set, which makes them progressively harder to obtain, whether analytically or through

simulations. Further, analysing and comparing them also becomes exceedingly difficult.

The most certain and forward way of comparing covariance matrices is in terms of their ability to reproduce cosmological constraints, that is, by employing a full Bayesian analysis. Performing this analysis for every new covariance matrix, however, is very time-consuming and computationally expensive. Friedrich et al. (2021), using the Gaussian linear model, studied in detail the impact of different covariance matrices on the uncertainties in parameter estimation, the position of the best-fitting parameters, and the relative  $\chi^2$  value. In this work, we seek a fast way of comparing covariance matrices that eliminates the need for a cosmological analysis to identify the differences between their constraints. In other words, we focus on the impact on the uncertainties in parameter estimation. The motivation is simple: If two covariance matrices produce similar results, then they should, at some level, have comparable features. We find here that even matrices with elements differing by several orders of magnitude (both diagonal and off-diagonal terms) produce parameter estimates that are almost indistinguishable from each other. It is, therefore, conceivable to think that a fair comparison cannot rely on the elements of the full covariance matrix. In Ferreira et al. (2021), one of the highlighted results is the potential in using compressed covariance matrices for comparison. It is shown that two different compressed covariance matrices with consistent parameter constraints also show good agreement in a one-to-one element comparison. Compression methods are powerful tools for reducing the dimensionality of the covariance matrix in order to facilitate and potentially speed up the process of parameter estimation. The most

\* E-mail: [tassia.ferreira@linea.gov.br](mailto:tassia.ferreira@linea.gov.br)

<sup>1</sup> [www.j-pas.org](http://www.j-pas.org)

<sup>2</sup> [www.lsst.org](http://www.lsst.org)

<sup>3</sup> [www.euclid-ec.org](http://www.euclid-ec.org)

<sup>4</sup> [www.skatelescope.org](http://www.skatelescope.org)

successful compression schemes are capable of taking a covariance matrix of size  $N \times N$  and shrinking it down to  $n \times n$ , where  $n$  is the number of free parameters. In this work, we use the Massively Optimised Parameter Estimation and Data compression (MOPED) algorithm, as described by Heavens, Jimenez & Lahav (2000), which works remarkably well in the case of Gaussian data, where the model for the mean only depends linearly on the parameters. When these circumstances are met, the method is said to be lossless in the sense that there is no loss of precision in the parameter constraints.

On the other hand, compression schemes are non-invertible, which means that, given a compressed matrix, we are unable to return to the original covariance matrix. In this sense, even if we were able to obtain a compressed covariance matrix, either analytically or with simulations, we would have no way of applying it to our real data set of interest.

This work has two main goals: (i) to develop a method to recreate the full matrix given a compressed one, and (ii) to propose a fast and reliable method to compare compressed covariance matrices that discards the need for a full cosmological analysis.

This paper is organized as follows. We start by introducing our data vector and covariance matrices in Section 2. Then, in Section 3, we review the matrix compression method we adopt, while we discuss in Section 4 invertible compression. We present our method to compare covariance matrices in Section 5, where we also analyse two cosmic shear covariance matrices. We conclude in Section 6.

## 2 COSMIC SHEAR

Light from distant galaxies is deflected by the gravitational field of large-scale structures as it travels through the Universe. This creates a correlated distortion of images, known as cosmic shear, which can be used to directly probe the underlying dark matter distribution and provide insight into the growth of structures and the geometry of the Universe. Cosmic shear has thus emerged as a powerful probe for dark energy (Kilbinger 2015; Hikage et al. 2019; Asgari et al. 2021; Secco et al. 2022). We review cosmic shear statistics in Appendix A.

### 2.1 The DESY1 data

The cosmic shear measurements for the Year 1 release of the Dark Energy Survey (DESY1; Troxel et al. 2018) were taken over an area of  $1321 \text{ deg}^2$  of the southern sky and are divided into four tomographic redshift bins from  $0.20 < z < 1.30$  (Zuntz et al. 2018), according to the posterior of the photometric redshift as estimated from *griz* flux measurements (Hoyle et al. 2018).

Each of the 10 bin-pair combinations contains 20 angular bins between 2.5 and 250 arcmin, yielding a data vector of length 200, for each statistic. Not all angular bins are used however, due to cuts that remove angular scales sensitive to baryonic effects, thus reducing the data vector to 167 points for  $\xi_+(\theta)$  and 60 for  $\xi_-(\theta)$ , totalling 227 points.

We assume a flat  $\Lambda$  cold dark matter ( $\Lambda$ CDM) model, with six free parameters,  $\{A_s, \Omega_m, \Omega_b, \Omega_b h^2, H_0, n_s\}$ , and fix  $w = -1$  and  $\tau = 0.08$ . Since we disconsider the baryonic effects, the astrophysical systematics are largely dominated by intrinsic alignment (IA), which describes the coherent orientation of galaxies due to interactions with the underlying gravitational tidal field regions. We vary the amplitude of the non-linear alignment model,  $A_{IA0}$ , and its redshift evolution,  $\eta_{IA}$ , which are related via  $A_{IA0} \equiv A_{IA0}[(1+z)/(1+z_0)]^{\eta_{IA}}$ , where  $z_0 = 0.62$  is the pivot redshift. We also have the shear multiplicative bias,  $m^i$ , which varies with each tomographic bin. Lastly, we vary the photo- $z$  bias,  $\Delta z^i$ , on the distribution of galaxies in each redshift

**Table 1.** List of the priors used in the analysis for parameter constraints using the data set described in Section 2.1.  $\mathcal{U}$  denotes flat in the given range and  $\mathcal{G}$  is Gaussian with mean equal to its first argument and dispersion equal to its second.

Parameter	Prior
Cosmological	
$\Omega_m$	$\mathcal{U}(0.1, 0.9)$
$\log A_s$	$\mathcal{U}(3.0, 3.1)$
$H_0 \text{ (km s}^{-1} \text{ Mpc}^{-1})$	$\mathcal{U}(55, 91)$
$\Omega_b$	$\mathcal{U}(0.03, 0.07)$
$\Omega_b h^2$	$\mathcal{U}(0.0005, 0.01)$
$n_s$	$\mathcal{U}(0.87, 1.07)$
Astrophysical	
$A_{IA0}$	$\mathcal{U}(-5, 5)$
$\eta_{IA}$	$\mathcal{U}(-5, 5)$
Systematic	
$m^i$	$\mathcal{G}(0.012, 0.023)$
$\Delta z^1$	$\mathcal{G}(-0.001, 0.016)$
$\Delta z^2$	$\mathcal{G}(-0.019, 0.013)$
$\Delta z^3$	$\mathcal{G}(0.009, 0.011)$
$\Delta z^4$	$\mathcal{G}(-0.018, 0.022)$

bin. The priors for these 16 parameters are given in Table 1. For brevity, we show the posterior probability density functions only for the matter density parameter  $\Omega_m$  and the amplitude of matter fluctuations  $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ .

Finally, the parameter constraints are obtained with COSMO-SIS (Zuntz et al. 2015), while employing the MULTINEST (Feroz, Hobson & Bridges 2009) sampler and following the same pipeline described in Troxel et al. (2018), with the modified likelihood for the transformed and compressed data sets and covariance matrices used in Ferreira et al. (2021). The MULTINEST run had 1000 livepoints, efficiency set to 0.05, tolerance to 0.1, and constant efficiency set to True.

### 2.2 The KiDS-1000 data

The measurements for the Kilo-Degree Survey 1000 (KiDS-1000; Asgari et al. 2021) contain  $1006 \text{ deg}^2$  of images, with the primary images taken in the  $r$  band, but with the final set having photometry in *ugriZYJHK<sub>s</sub>* (Wright et al. 2019), after being combined with infrared data from the VISTA Kilo-degree Infrared Galaxy survey (Edge et al. 2013). The data are divided into five tomographic bins,  $z_B$ , based on their best-fitting photometric redshifts and in the range  $0.1 < z < 1.2$ .

There are nine angular bins between 0.5 and 500 arcmin, resulting in a data vector of length 270. The angular cuts are applied to  $\xi_-(\theta)$ , removing scales with  $\theta < 4 \text{ arcmin}$  (Hildebrandt et al. 2017), which leaves 135 data points for  $\xi_+(\theta)$  and 90 for  $\xi_-(\theta)$ . The final data set has length 235, with a  $235 \times 235$  covariance matrix.

Similarly to the DESY1 analysis, we take a flat  $\Lambda$ CDM model, with  $w = -1$ , and five free cosmological parameters,  $\{S_8, \Omega_c h^2, \Omega_b h^2, h, n_s\}$ . There are two astrophysical nuisance parameters: the baryon feedback parameter,  $A_{bary}$  and  $A_{IA}$ , where, for analyses with this data set, the latter does not carry a redshift dependence. The mean of the five redshift distributions is also allowed to vary and are correlated through their covariance matrix. Finally, for the analyses with  $\xi_+(\theta)$ , we have  $\delta_c = \pm \sqrt{c_1^2 + c_2^2}$  to account for the uncertainty of the additive ellipticity bias terms,  $c_1$  and  $c_2$ , assuming that they are constants. Table 2 shows the parameters varied as well as their priors.

**Table 2.** List of the priors used in the analysis for parameter constraints using the data set described in Section 2.2.

Parameter	Prior
Cosmological	
$S_8$	$\mathcal{U}(0.1, 1.3)$
$\Omega_c h^2$	$\mathcal{U}(0.051, 0.255)$
$\Omega_b h^2$	$\mathcal{U}(0.019, 0.026)$
$h$	$\mathcal{U}(0.64, 0.82)$
$n_s$	$\mathcal{U}(0.84, 1.1)$
Astrophysical	
$A_{IA}$	$\mathcal{U}(-6, 6)$
$A_{bary}$	$\mathcal{U}(2.0, 3.13)$
Systematic	
$\delta_1^1$	$\mathcal{G}(0, 1.0)$
$\delta_2^1$	$\mathcal{G}(-0.181, 1.0)$
$\delta_3^1$	$\mathcal{G}(-1.110, 1.0)$
$\delta_4^1$	$\mathcal{G}(-1.395, 1.0)$
$\delta_5^1$	$\mathcal{G}(1.265, 1.0)$
$\delta_c$	$\mathcal{G}(0, 2.3 \times 10^{-4})$

For parameter constraints, we use the `MULTINEST` sampler within `COSMOSIS`, with the same settings described in the previous section, but with the KiDS Cosmology Analysis Pipeline, KCAP (Asgari et al. 2021). We use a modified likelihood to account for the transformed and compressed covariance matrices and data vectors.

### 2.3 The covariance matrices

There are three distinct covariance matrices used in this work, which will be described in this section. What we refer to as the *DES covariance matrix* (DCM) is the DESY1 cosmic shear covariance matrix obtained with `COSMOLIKE` (Krause & Eifler 2017). The covariance is largely dominated by the shape noise and Gaussian components of the covariance, with a halo model framework being used to include the non-Gaussian parts.

We label the second one as the *Gaussian covariance matrix* (GCM), which was generated with the code used for the KiDS-450 survey (Köhlinger et al. 2017), but with the same configuration as DCM, and constituting of only the Gaussian contributions to shape noise, sample variance, and the mixed noise–sample variance term. We do not include the non-Gaussian components and the supersample covariance so that the differences between the covariance matrices and their parameter constraints are accentuated. It is used here along with the DCM to test our algorithm for comparing covariance matrices. It has been shown in Ferreira et al. (2021) that their parameter constraints are accurately reproduced when compressed with MOPED.

The third and last one is the *KiDS covariance matrix* (KCM), which is used for testing our invertible transformation as well as for some of the tests with modifications to one of the blocks of the transformed matrix. It is the same covariance matrix used in the KiDS-1000 survey analysis with cosmic shear.

The second and third covariance matrices are calculated with the code used for the KiDS-450 and KiDS-1000 surveys (Köhlinger et al. 2017; Joachimi et al. 2020). Both were obtained analytically and follow the procedure in Joachimi, Schneider & Eifler (2008) for obtaining second-order cosmic shear measurements under the assumption that density field is Gaussian, the galaxies are uniformly distributed, and the survey has a straightforward geometry. The main differences between them are their size, with GCM being  $227 \times 227$

and KCM  $235 \times 235$ ; GCM was produced for DESY1 cosmic shear statistics and KCM is the covariance matrix used in the KiDS-1000 survey; finally, GCM is a Gaussian covariance matrix, while KCM contains both Gaussian and non-Gaussian terms.

DCM and GCM are used in Sections 4 and 5, whereas KCM is only used in Section 4.

### 3 COMPRESSION SCHEME

Consider a data set represented by the vector  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  with probability distribution  $L(\mathbf{x}; \Theta)$ , where  $\Theta = (\theta_1, \theta_2, \dots, \theta_n)$  is a vector of the model parameters. Take the logarithm of the probability distribution,  $\mathcal{L} = \ln L$ ; the first derivative with respect to the model parameters,  $\mathcal{L}_{,i}$ , known as the score function, tells how sensitive the model is with respect to the parameters. Its second derivative, the Hessian matrix,

$$\mathbf{H}_{ij} = \mathcal{L}_{,ij}, \quad (1)$$

describes the correlation of the estimated values of  $\theta_i$  and  $\theta_j$ . The expectation value of the negative of the Hessian,  $\mathbf{F}_{ij} = -\langle \mathbf{H}_{ij} \rangle$ , gives us the Fisher information matrix, which is an essential quantity in Bayesian statistics and, as we will see, for compressing covariance matrices. To obtain a compression scheme capable of retaining the highest amount of information, we seek to maximize  $\mathbf{F}_{ij}$ ; to do so, we start with the log-likelihood function for a Gaussian probability distribution,<sup>5</sup>

$$2\mathcal{L} = \ln \det \mathbf{C} + (\mathbf{x} - \boldsymbol{\mu})^t \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}), \quad (2)$$

with both the covariance matrix,  $\mathbf{C} = \langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t \rangle$ , and  $\boldsymbol{\mu} = \langle \mathbf{x} \rangle$  dependent on the model parameters  $\Theta$ . We also define the data matrix as

$$\mathbf{D} \equiv (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t. \quad (3)$$

Taking this definition and using  $\ln \det \mathbf{C} = \text{Tr}[\ln \mathbf{C}]$ , we can express equation (2) in a simpler form,

$$2\mathcal{L} = \text{Tr}[\ln \mathbf{C} + \mathbf{C}^{-1} \mathbf{D}]. \quad (4)$$

Derivating for  $\theta_i$  and  $\theta_j$ , we find the quantity we wish to maximize:

$$\langle \mathcal{L}_{,ij} \rangle = \frac{1}{2} \text{Tr}[\mathbf{C}^{-1} \mathbf{C}_{,i} \mathbf{C}^{-1} \mathbf{C}_{,j} + \mathbf{C}^{-1} \mathbf{M}_{ij}], \quad (5)$$

where  $\mathbf{M}_{ij} = \boldsymbol{\mu}_{,i} \boldsymbol{\mu}_{,j}^t + \boldsymbol{\mu}_{,j} \boldsymbol{\mu}_{,i}^t$ .

There are three ways one can proceed from here, with respect to the dependence on the model parameters: (1) There is the more general case, where both the mean and the covariance depend on the model parameters; (2) only the covariance has a dependence, in which case the second term vanishes; and (3) only the mean is dependent, and the first term vanishes. The general case has been tackled by Alsing & Wandelt (2018), where they use the score function to derive  $n$  compressed statistics of the data. The second case often reduces to a Karhunen–Loève eigenvalue problem that results in keeping only the linear combinations of the data with the most informative eigenvalues (Alonso 2018). Finally, the third case is the basis for MOPED (Tegmark, Taylor & Heavens 1997; Heavens et al. 2000), which uses linear compression to radically reduce the data set. Here, we will only be working with the latter, with an extension for other compression methods being planned for future works. In this case, equation (5) reduces to

$$\mathbf{F}_{ij} = \langle \mathcal{L}_{,ij} \rangle = \frac{1}{2} \text{Tr}[\mathbf{C}^{-1} \mathbf{M}_{ij}]. \quad (6)$$

<sup>5</sup>We have dropped the additive constant  $n \ln(2\pi)$ .

For the simplest case of only one parameter, we have

$$\mathbf{F}_{11} = \boldsymbol{\mu}_{,1}' \mathbf{C}^{-1} \boldsymbol{\mu}_{,1}. \quad (7)$$

If we apply a compression of the type  $y = \mathbf{b}'\mathbf{x}$ , we can follow these same steps to obtain the Fisher matrix,

$$\mathbf{F}_{11}^{\text{comp}} = \frac{\mathbf{b}' \mathbf{M}_{11} \mathbf{b}}{\mathbf{b}' \mathbf{C} \mathbf{b}}. \quad (8)$$

To find an extremum such that  $\mathbf{b}' \mathbf{C} \mathbf{b} = 1$ , we apply a Lagrange multiplier,

$$\boldsymbol{\mu}_{,1}(\mathbf{b}' \boldsymbol{\mu}_{,1}) = \lambda \mathbf{C} \mathbf{b}. \quad (9)$$

The solution is the eigenvector that, when normalized, gives

$$\mathbf{b}_1 = \frac{\mathbf{C}^{-1} \boldsymbol{\mu}_{,1}}{\sqrt{\boldsymbol{\mu}_{,1}' \mathbf{C}^{-1} \boldsymbol{\mu}_{,1}}}. \quad (10)$$

Plugging this back in equation (8), we find

$$\mathbf{F}_{11}^{\text{comp}} = \boldsymbol{\mu}_{,1}' \mathbf{C}^{-1} \boldsymbol{\mu}_{,1} = \mathbf{F}_{11}, \quad (11)$$

which shows that the Fisher matrix is unchanged.

We can further expand equation (10) to multiple parameters. By employing a Gram–Schmidt decomposition, we obtain  $y_i$ 's that are orthonormal, uncorrelated, and carry as much information as possible about the parameter  $\theta_i$ . We then have

$$\mathbf{b}_n = \frac{\mathbf{C}^{-1} \boldsymbol{\mu}_{,n} - \sum_{q=1}^{n-1} (\boldsymbol{\mu}_{,n}' \mathbf{b}_q) \mathbf{b}_q}{\sqrt{\boldsymbol{\mu}_{,n}' \mathbf{C}^{-1} \boldsymbol{\mu}_{,n} - \sum_{q=1}^{n-1} (\boldsymbol{\mu}_{,n}' \mathbf{b}_q)^2}}, \quad (12)$$

for  $q < n$ . With this in hand, the resulting weighing matrix  $\mathbf{b}$  contains  $n$  rows, and the covariance matrix is shrunk to  $n \times n$ .

Note that the compression vector  $\mathbf{b}$  only depends on the model parameters. If the model is linear in its parameters, then this dependence disappears, and the method becomes lossless.

To use MOPED as described here, we have assumed a Gaussian likelihood and that the fiducial values at which the derivatives are taken are those at the maximum likelihood point. If one does not have prior knowledge of this value, then one could iterate to find it but, as Heavens et al. (2000) have found, this is often unnecessary. We also trust that the cosmological model we have chosen is the correct one; deviations from a baseline cosmological model can be accounted for by introducing additional weighing vectors, as described in Heavens, Sellentin & Jaffe (2020).

In Ferreira et al. (2021), it was established that the MOPED-compressed DESY1 covariance matrix and data set could reproduce the original parameter constraints; here, we find that the same is true for KiDS-1000. While these results are not new, verification for KiDS-1000 is an important step of the analysis since, as the authors point out, not all compression schemes are capable of reproducing the original parameter constraints. In particular, those that relied on eigenvalues or the signal-to-noise ratio showed a loss of constraining power on cosmological parameters since the modes relevant to the IA parameters were discarded by these methods.

#### 4 INVERTIBLE COMPRESSION

The next step in our analysis is to obtain an invertible transformation, based on the MOPED compression scheme, that is able to reproduce the same parameter constraints we previously obtained with both the original and the compressed covariance matrices. This step is vital for using the compressed covariance matrix, produced analytically or through simulations, with the observed data vector.

We note here that this procedure will not reduce the size of the covariance matrix, but rather the number of relevant elements. The transformed matrix is then composed of the compressed covariance matrix and some additional terms. By construction, we do not expect the additional elements to alter the parameter constraints, and we show that this is indeed the case. It is less intuitive, however, that, through the inverse of the invertible transformation, it is possible to generate a new covariance matrix, with elements differing by large orders of magnitude, that retain the same constraining power when using the same data vector. Here, we show how this can be achieved and we highlight the perils of considering a comparison between the elements of the full covariance matrices.

We start with the non-normalized equation (10), and expand it to an invertible,  $N \times N$  transformation matrix,

$$\mathbf{B} = (\mathbf{b} \ U), \quad (13)$$

where  $U$  has dimension  $(N - n) \times N$ . We want to find  $U$  such that

$$\mathbf{C}^{\text{trans}} = \mathbf{B}' \mathbf{C} \mathbf{B} = \begin{pmatrix} \mathbf{b}' \mathbf{C} \mathbf{b} & 0 \\ 0 & U \mathbf{C} U' \end{pmatrix}, \quad (14)$$

which implies

$$\mathbf{b}' \mathbf{C} U = 0. \quad (15)$$

For the above to be true, the rows of  $U$  must be composed of vectors that form the nullspace of  $\mathbf{b}' \mathbf{C}$ . To simplify notation, we represent the transformed covariance matrix partitioned blockwise as

$$\begin{pmatrix} C1 & 0 \\ 0 & C3 \end{pmatrix}. \quad (16)$$

In the following section, we apply this transformation to a toy example, so we can explore, in a more didactic manner, how this transformation alters the covariance matrix.

#### 4.1 Toy example

In order to better understand our invertible transformation and how  $C3$  affects  $\mathbf{C}$ , we take a simple toy example. Consider the Gaussian data described by

$$\{t_i, x_i\} \text{ with } i = 1, 2, 3, \\ C = \sigma^2 I_3, \quad (17)$$

where  $I_3$  is the three-dimensional identity matrix and  $t_i$  is the independent variable associated with  $x_i$ . Next, we propose the following model:

$$\mu(t) = \theta_1 + \theta_2 t, \quad (18)$$

for which we can obtain the likelihood as

$$2\mathcal{L} = (x_i - \mu(t_i)) C_{ij}^{-1} (x_j - \mu(t_j)) \\ = \sum_{i=1,2,3} \frac{(x_i - \mu(t_i))^2}{\sigma^2}. \quad (19)$$

The next step is then to derive an explicit expression for  $\mathbf{b}$ ,  $U$ ,  $C1$ , and  $C3$  as a function of  $t_i$ ,  $x_i$ ,  $\sigma$ , and  $\theta_i$ . We start with

$$\mathbf{b} = \begin{pmatrix} \frac{1}{\sigma^2} & \frac{1}{\sigma^2} & \frac{1}{\sigma^2} \\ \frac{t_1}{\sigma^2} & \frac{t_2}{\sigma^2} & \frac{t_3}{\sigma^2} \end{pmatrix} \quad (20)$$

and

$$U = \begin{pmatrix} -\frac{t_2 + t_3}{t_1 - t_2} & -\frac{t_1 - t_3}{t_1 - t_2} & 1 \end{pmatrix}, \quad (21)$$



which we can use to find

$$\mathbf{C}^{\text{trans}} = \begin{pmatrix} \frac{3}{\sigma^2} & \frac{t_1+t_2+t_3}{\sigma^2} & 0 \\ \frac{t_1+t_2+t_3}{\sigma^2} & \frac{t_1^2+t_2^2+t_3^2}{\sigma^2} & 0 \\ 0 & 0 & \mathbf{C3} \end{pmatrix}, \quad (22)$$

where

$$\mathbf{C3} = 2\sigma^2 \frac{t_1^2 + t_2^2 + t_3^2 - t_2 t_3 - t_1(t_2 + t_3)}{(t_1 - t_2)^2}. \quad (23)$$

Plugging our values in equation (6), we get

$$F_{ij}^{\text{trans}} = \begin{pmatrix} \frac{3}{\sigma^2} & \frac{t_1+t_2+t_3}{\sigma^2} \\ \frac{t_1+t_2+t_3}{\sigma^2} & \frac{t_1^2+t_2^2+t_3^2}{\sigma^2} \end{pmatrix} = \mathbf{C1}. \quad (24)$$

Making  $\mathbf{C3} = \mathbf{C3}$ , we can revert back the transformation to find  $\mathbf{C}'$ . If we substitute  $\mathbf{C}'$  in equation (6), we find that  $F' = F$ , which shows that the Fisher matrix of the modified covariance matrix does not depend on  $\mathbf{C3}$ .

#### 4.2 Perturbing C3

We now describe the ways in which we perturb the  $\mathbf{C3}$ . The impact these modifications have on the parameter constraints is quantified by performing a full cosmological analysis with the new covariance matrix.

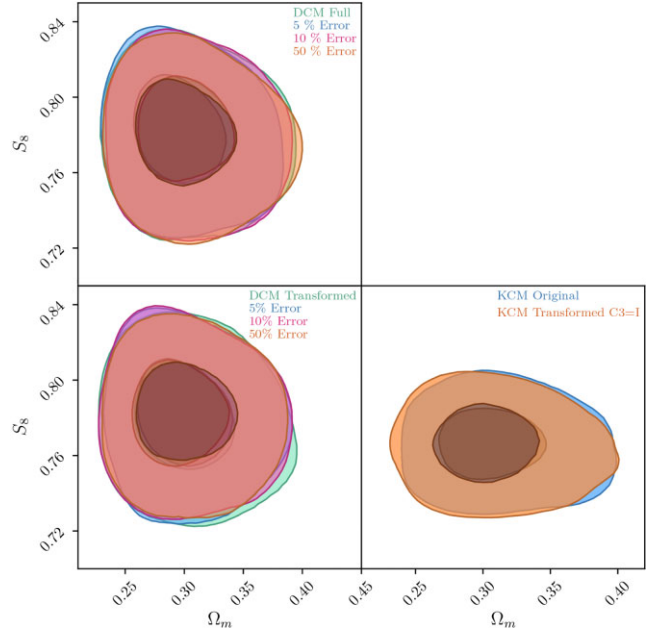
We start with the simple task of making  $\mathbf{C3} = \mathbf{1}$ . This modification increases the diagonal elements by several orders of magnitude and negates all the cross-terms of  $\mathbf{C3}$ . We also modify the elements of the  $\mathbf{C3}$  block of DCM by introducing a Gaussian error of 5 per cent, 10 per cent, and 50 per cent while maintaining its symmetry. In Fig. 1, we see that these configurations are irrelevant to the parameter constraints, as the contour plots show agreement with the original results.

The next question to address is how this property propagates when we apply the inverse of the transformation  $\mathbf{B}$  to the perturbed  $\mathbf{C}^{\text{trans}}$ . We carry out the same procedure for perturbing the elements of  $\mathbf{C3}$  by 5 per cent, 10 per cent, and 50 per cent. We then apply  $\mathbf{B}^{-1}$  to transform it into the space of the original data vector. As expected, the results are similar to those found in the previous analysis, see Fig. 1.

In Fig. 2 (left-hand panel for DCM and right-hand panel for KCM), we show the ratio between the elements of the perturbed covariance matrices and the original one for a 5 per cent (top) and 50 per cent (bottom), and highlight the diagonal elements in red. A trend can be seen where the smallest elements show a greater disagreement, and this decreases as their values increase. Even the largest elements, however, show a difference of about an order of magnitude. By construction, the broad disparities between the elements should not impact the parameter constraints, and we see that this is indeed the case. We see here that covariance matrices obtained by independent algorithms show disagreements between their elements but produce similar constraints. It is therefore essential to note that metrics to compare covariance matrices based solely on these elements themselves may not produce conclusive results. As such, for an unbiased comparison, we propose to concentrate on the covariance matrices compressed with the MOPED algorithm.

### 5 COMPARING COVARIANCE MATRICES

We have shown in the previous section that it is imperative that a comparison between covariance matrices be done on its compressed set. This is fortunate because the reduced size results in a speed gain for comparison. In this section, we explain the framework we have



**Figure 1.** Top: DES Y1 constraints on the cosmological parameters  $\Omega_m$  and  $S_8$  for the original covariance matrix (green) and for three covariance matrices produced by applying the inverse transformation to those with 5 per cent (blue), 10 per cent (pink), and 50 per cent (orange) perturbation applied to the  $\mathbf{C3}$  block of the transformed DCM. Bottom left: DES Y1 constraints on  $\Omega_m$  and  $S_8$  for the transformed covariance matrix (green) and for a 5 per cent (blue), 10 per cent (pink), and 50 per cent (orange) perturbation applied to the  $\mathbf{C3}$  block of DCM. Bottom right: KiDS-1000 parameter constraints for the original and transformed covariance matrix on  $\Omega_m$  and  $S_8$ . The darker curve, in blue, is for the original covariance matrix and the lighter curve, in orange, is for when the block  $\mathbf{C3}$  of KCM is replaced by the identity matrix.

developed for obtaining a reliable metric of comparison that discards the need for a full cosmological analysis.

Our method separates the comparison into two parts: the analysis of the diagonal elements, the  $n$ -dimensional vector  $\mathcal{D}$ , and of the independent elements of the correlation matrix, the  $n(n-1)/2$ -dimensional vector  $\mathcal{C}$ . The corresponding calculated differences will then be related to the parameter constraints in terms of their contour levels. To find the differences, we use a Monte Carlo approach by employing Powell's approach for minimizing a function (Powell 1964).

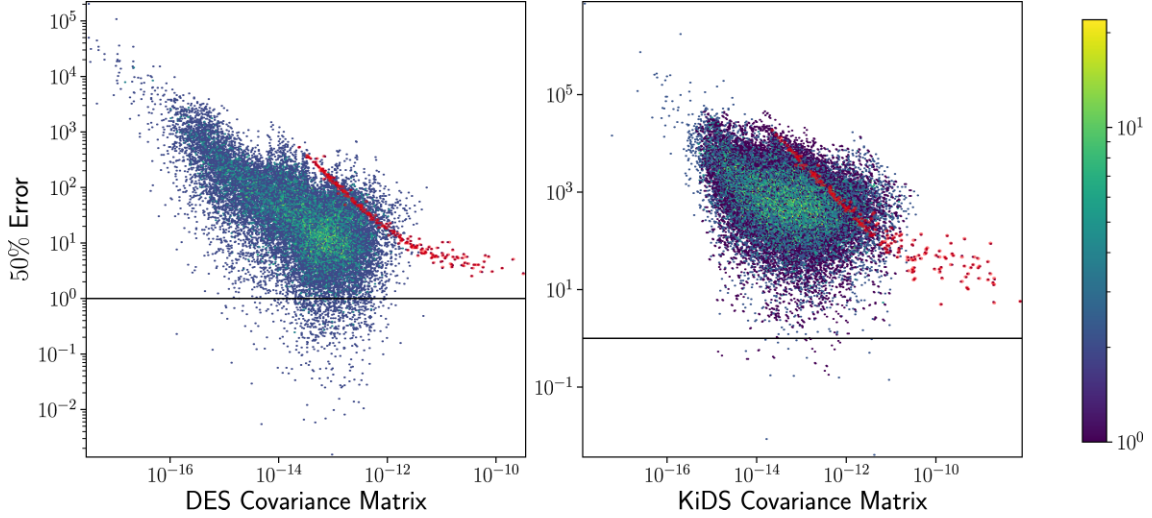
Take two compressed covariance matrices:  $\mathbf{C}_{\text{base}}$  and  $\mathbf{C}_{\text{test}}$ .<sup>6</sup> For each step  $i$ , we create a mock sample  $\{\mathcal{D}_{\delta,i}\}$  (or  $\{\mathcal{C}_{\delta,i}\}$ ) by perturbing  $\mathcal{D}_{\text{base}}$  (or  $\mathcal{C}_{\text{base}}$ ) with a given error percentage  $\delta$ . For the diagonal part, the mocks are generated by drawing  $\mathcal{E}_{\mathcal{D}}$  from a multivariate Gaussian distribution  $\mathcal{G}[0_n, \delta^2 \mathbf{I}_n]$ , such that

$$\mathcal{D}_{\delta} = (1 + \mathcal{E}_{\mathcal{D}})^2 \mathcal{D}_{\text{base}}. \quad (25)$$

In the case of the correlation matrix, we encounter the restriction that the values must be in the range  $[-1, 1]$ . Applying this by force could result in a perturbation not cohesive with our chosen  $\delta$ . This is resolved by switching to the hyperbolic tangent function and correcting for the Jacobian,

$$z = \tanh^{-1} \mathcal{C}_{\text{base}}, \quad \delta z = \mathcal{E}_{\mathcal{C}} \cosh^2 \left( z + \frac{\mathcal{E}_{\mathcal{C}}}{2} \right), \quad (26)$$

<sup>6</sup>Note that these are the Fisher matrices in the parameters.



**Figure 2.** Scatter plots of the ratio between the elements of the perturbed covariance matrix with a 50 per cent Gaussian error and those of the original one. The left-hand panel shows the ratio between the DESY1 covariances, while the right-hand panel shows the ratio relative to KiDS-1000, and the x-axis representing the elements of the respective covariance matrix. The red dots represent the ratio between the diagonal elements of the respective matrices.

where  $\mathcal{E}_C$  is drawn similarly to  $\mathcal{E}_D$ . Our perturbed vector then becomes

$$C_\delta = \tanh(z + \delta z). \quad (27)$$

Since the hyperbolic function increases asymptotically towards one, we are able to ensure a smooth perturbation.

Next, we produce the sample covariance matrix  $S_\delta$  from the generated mocks

$$S_\delta = \frac{1}{m-1} \sum_{i=1}^m (\mathcal{D}_{\delta,i} - \overline{\mathcal{D}_\delta})(\mathcal{D}_{\delta,i} - \overline{\mathcal{D}_\delta})^T, \quad (28)$$

where  $m$  is the number of mock samples. The overbar denotes the average mock sample. We then calculate the fiducial  $\chi^2$  distribution,

$$\chi_\delta^2 = (\mathcal{D}_\delta - \mathcal{D}_{\text{base}}) S_\delta^{-1} (\mathcal{D}_\delta - \mathcal{D}_{\text{base}})^T, \quad (29)$$

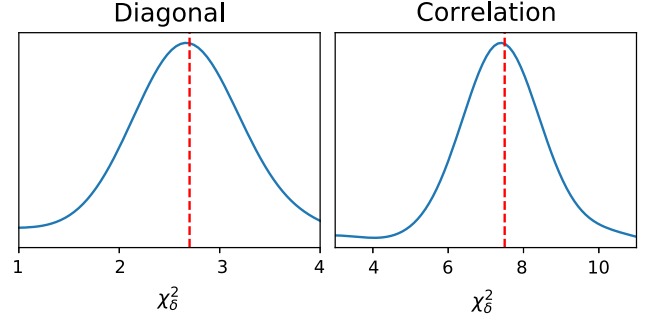
and, for each  $\delta$ , we find

$$\chi_{\text{test},\delta}^2 = (\mathcal{D}_{\text{test}} - \mathcal{D}_{\text{base}}) S_\delta^{-1} (\mathcal{D}_{\text{test}} - \mathcal{D}_{\text{base}})^T. \quad (30)$$

We iterate  $\delta^D$  to find a value for which  $\chi_{\text{test},\delta}^2$  is the maximum of  $\chi_\delta^2$ , for  $\delta \leq 20$  per cent. Finally, we find  $\sigma_\delta = (\delta_+ - \delta_-)/2$ , where  $\delta_+$  is the value that makes  $\chi_{\text{test}}^2$  fall at the right-hand border of the 68 per cent probability interval of the  $\chi^2$  distribution, and similarly for  $\delta_-$ . The steps for finding  $\delta^C$  are analogous to what we have described.

We developed a comprehensible code in PYTHON<sup>7</sup> that uses this procedure to test  $\mathcal{D}$  and  $\mathcal{C}$ . We found convergence for  $m > 5000$ . The value of  $\delta_{\text{test}}^D \pm \sigma_\delta$  estimates the distance between the two compressed matrices,  $C_{\text{base}}$  and  $C_{\text{test}}$ , as far as the uncertainties or correlations in the parameters are concerned.

As equations (25) and (27) show, we adopted a very agnostic point of view: We look for the strength  $\delta$  of the Gaussian noise  $\mathcal{E}$  that makes one matrix similar to the other. This zero-mean noise is supposed to arise because of the different choices that are made in the modelling of the covariance matrices. It is worth stressing that we are not quantifying the bias that may separate the two matrices,



**Figure 3.**  $\chi^2$  distributions that are used to determine  $\delta_{\text{test}}$  for the diagonal elements of the covariance matrix (left-hand panel) and for the elements of the correlation matrix (right-hand panel). The method finds the  $\delta_{\text{test}}$  value such that  $\chi_{\text{test}}^2$  (red dashed line) falls at the maximum of the distribution.

which, given the many elements, would be difficult to tackle in a systematic way.

### 5.1 Cosmic shear covariance matrices

We test this method by comparing the DCM and GCM covariance matrices. We find that the diagonal elements of the compressed GCM differ from the ones of the compressed DCM by

$$\delta_{\text{test}}^D = 2.6 \pm 0.5 \text{ per cent}, \quad (31)$$

and the correlations by

$$\delta_{\text{test}}^C = 7.5 \pm 0.6 \text{ per cent}. \quad (32)$$

We illustrate these findings in Fig. 3, which shows the  $\chi^2$  distributions used by our method: The values of  $\delta_{\text{test}}$  are such that the  $\chi_{\text{test}}^2$  values fall at the maximum of the distributions.

In order to validate these results, we carry out a full Bayesian analysis as shown in Fig. 4, where one can see that the two matrices give similar constraints. We then estimate from the chains the two  $16 \times 16$  covariance matrices on the parameters, that is, the second moments of the posterior distribution. These moments give the Gaussian errors that are relative to the full (possibly non-Gaussian)

<sup>7</sup> [github.com/t-ferreira/Covariance\\_comparison](https://github.com/t-ferreira/Covariance_comparison)

posterior. In order to compare with the results of equations (31) and (32), which are based on the Fisher matrices on the parameters, we then extract the diagonal elements  $\mathcal{D}_{\text{DCM}}$  and  $\mathcal{D}_{\text{GCM}}$  and the correlation vectors  $\mathcal{C}_{\text{DCM}}$  and  $\mathcal{C}_{\text{GCM}}$  from the inverse of the covariance matrices of the DCM and GCM chains. Note that the latter matrices are different from the Fisher matrix, which is only sensitive to the maximum of the likelihood. Next, we define

$$\mathcal{E}_{\mathcal{D}} = \sqrt{\frac{\mathcal{D}_{\text{GCM}}}{\mathcal{D}_{\text{DCM}}} - 1}, \quad (33)$$

$$\mathcal{E}_{\mathcal{C}} = (1 - \mathcal{C}_{\text{GCM}}^2) (\tanh^{-1} \mathcal{C}_{\text{GCM}} - \tanh^{-1} \mathcal{C}_{\text{DCM}}), \quad (34)$$

and compute the standard deviations of  $\mathcal{E}_{\mathcal{D}}$  and  $\mathcal{E}_{\mathcal{C}}$  to see if their values agree with the results of equations (31) and (32). We find

$$\sigma_{\mathcal{E}_{\mathcal{D}}} = 1.7 \text{ per cent}, \quad (35)$$

$$\sigma_{\mathcal{E}_{\mathcal{C}}} = 1.5 \text{ per cent}. \quad (36)$$

We find a  $2\sigma$  agreement for the diagonal elements and that our method overestimates the difference in the correlations. The latter is somewhat expected as correlations are more difficult to quantify when close to zero and for variables that are strongly degenerated because of non-Gaussianities in the posterior. Regarding the latter effect, we stress again that the results of equations (31) and (32) are based on the Fisher matrices on the parameters, while the ones of equations (35) and (36) are based on the full posterior, which may be significantly non-Gaussian in, especially, the nuisance parameters.

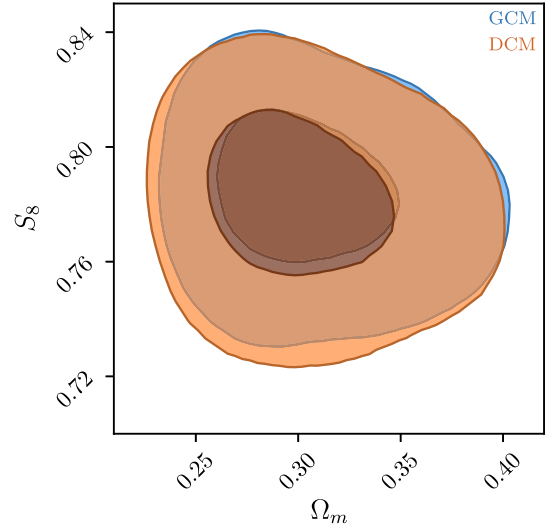
Finally, our algorithm took roughly 0.5 CPUh (2020 laptop) to generate the desired output, while a full cosmological analysis takes roughly 100 times more.

## 6 CONCLUSION

The complexity and the considerable size of covariance matrices of large data sets make them increasingly difficult to analyse. Because of the vast range of values, with their own elements often differing by several orders of magnitude, it is customary to identify the largest elements and the diagonal ones. Should they be similar, then it is likely that the parameter constraints are also compatible; we show here that this is a perilous assumption. Using an invertible transformation, we showed that it is possible to generate very distinct covariance matrices capable of generating the same cosmology, given the same data vector.

On the other hand, we follow the groundwork established by Ferreira et al. (2021), which showed MOPED is capable of reducing the size of the covariance matrix while retaining the necessary information for parameter constraints for cosmic shear statistics. We thus use the compressed matrix formalism to develop a fast and reliable method for the comparison that uses compressed covariance matrices. Together with this work, we release a comprehensible implementation in PYTHON of this method, which is available at [github.com/t-ferreira/Covariance\\_comparison](https://github.com/t-ferreira/Covariance_comparison).

A direct one-to-one element comparison, as is usually done, by considering the perturbed covariance matrices generated in Section 4 would have led one to erroneously expect notable differences in their constraining power as compared to those obtained with the original covariance matrix. This incorrect assumption is avoided if considering MOPED-compressed covariance matrices, since they are identical. This approach thus ensures an equitable comparison since we limit ourselves to the most relevant elements. Using our method, we show that the DESY1 covariance matrix and the GCM, which produce compatible cosmological constraints, feature a difference



**Figure 4.** Constraints on the parameters  $\Omega_m$  and  $S_8$  for the DCM and GCM produced for cosmic shear, as described in Section 2.3. The orange curve is for the DCM, while the blue is for the GCM. In the 16-dimensional parameter space, the volume of the posterior is about 20 per cent smaller for the latter.

of 7.5 per cent for the elements of their correlation matrix and 2.6 per cent for their diagonal elements of the covariance matrix. The latter figure is in agreement with the results from the full Bayesian analysis, which gives a 1.7 per cent difference. Regarding the correlations, the full analysis gives a lower value, 1.5 per cent, possibly because of the difficulty in estimating the correlation for degenerate cases. While our approximate method cannot replace a full Bayesian analysis, it is a powerful and reliable tool for the development and validation of codes that estimate covariance matrices. Our method takes roughly 100 times less CPUh than a full cosmological analysis.

Finally, it is relevant to point out that the comparison presented here is done between analytical covariance matrices, and that drawing  $\delta$  from a Gaussian distribution proved suitable for reproducing the differences between their parameter constraints. In the case of covariance matrices obtained using different methods, such as comparing an analytical and a simulated one, the Wishart distribution is a likely better candidate for generating  $\delta$  (Taylor, Joachimi & Kitching 2013).

## ACKNOWLEDGEMENTS

It is a pleasure to thank Scott Dodelson for useful comments and discussions. TF thanks Tilman Troester for helpful comments on how to use KCAP. TF also thanks CAPES and FAPES for financial support. VM thanks CNPq and FAPES for partial financial support.

## DATA AVAILABILITY

The algorithm used as well as the data underlying this paper can be found in [github.com/t-ferreira/Covariance\\_comparison](https://github.com/t-ferreira/Covariance_comparison), where we have included a notebook example to reproduce our results.

## REFERENCES

- Alonso D., 2018, *MNRAS*, 473, 4306  
 Alsing J., Wandelt B., 2018, *MNRAS*, 476, L60

- Asgari M. et al., 2021, *A&A*, 645, A104  
 Blanchard A. et al., 2020, *A&A*, 642, A191  
 Chisari N. E. et al., 2019, *ApJS*, 242, 2  
 Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, *The Messenger*, 154, 32  
 Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601  
 Ferreira T., Zhang T., Chen N., Dodelson S., 2021, *Phys. Rev. D*, 103, 103535  
 Friedrich O. et al., 2021, *MNRAS*, 508, 3125  
 Heavens A., Jimenez R., Lahav O., 2000, *MNRAS*, 317, 965  
 Heavens A. F., Sellentin E., Jaffe A. H., 2020, *MNRAS*, 498, 3440  
 Hikage C. et al., 2019, *PASJ*, 71, 1  
 Hildebrandt H. et al., 2017, *MNRAS*, 456, 1454  
 Hoyle B. et al., 2018, *MNRAS*, 478, 592  
 Joachimi B., Schneider P., Eifler T., 2008, *A&A*, 477, 43  
 Joachimi B. et al., 2020, *A&A*, 646, A129  
 Kilbinger M., 2015, *Rep. Prog. Phys.*, 78, 086901  
 Köhlinger F. et al., 2017, *MNRAS*, 471, 4412  
 Krause E., Eifler T., 2017, *MNRAS*, 470, 2100  
 Krause E. et al., 2017, preprint ([arXiv:1706.09359](https://arxiv.org/abs/1706.09359))  
 Powell M. J. D., 1964, *Comput. J.*, 7, 155  
 Schneider P., van Waerbeke L., Mellier Y., 2002, *A&A*, 389, 729  
 Secco L. F. et al., 2022, *Phys. Rev. D*, 105, 023515  
 Taylor A., Joachimi B., Kitching T., 2013, *MNRAS*, 432, 1928  
 Tegmark M., Taylor A., Heavens A., 1997, *ApJ*, 480, 22  
 Troxel M. A. et al., 2018, *Phys. Rev. D*, 98, 043528  
 Wright A. H. et al., 2019, *A&A*, 632, A34  
 Zuntz J. et al., 2015, *Astron. Comput.*, 12, 45  
 Zuntz J. et al., 2018, *MNRAS*, 481, 1149

## APPENDIX A: COSMIC SHEAR STATISTICS

The mapping of the light distribution of the source to image coordinates is done via the magnification tensor, which is the inverse of the Jacobian matrix,

$$\mathbf{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (\text{A1})$$

with  $\kappa$  being the isotropic decrease or increase of the observed size of the source image and the anisotropic deformation  $\gamma = \gamma_1 + i\gamma_2$  being the shear. In Fourier space, these two quantities are related via

$$\tilde{\gamma}(\ell) = \frac{(\ell_1 + i\ell_2)^2}{\ell^2} \tilde{\kappa}(\ell) = e^{2i\beta} \tilde{\kappa}(\ell), \quad (\text{A2})$$

where  $\beta$  is the polar angle of the wavevector  $\ell = (\ell_1, \ell_2)$ .

The convergence can also be interpreted according to the projected matter density, with its power spectrum being associated with that of the matter density contrast,  $P_\delta$  (Schneider, van Waerbeke & Mellier 2002). Using several approximations, such as the Limber projection, which uses only modes that lie in the plane of the sky, the small-angle approximation, and the flat-sky limit, where we replace spherical harmonics by Fourier transforms, we can write

$$P_\kappa(\ell) = \frac{9}{4} \Omega_m^2 \left( \frac{H_0}{c} \right)^4 \int_0^{\chi_{\text{lim}}} d\chi \frac{g^2(\chi)}{a^2(\chi)} P_\delta(k = \frac{\ell}{f_K(\chi)}, \chi). \quad (\text{A3})$$

The integral goes to the limiting comoving distance of the galaxy sample,  $\chi_{\text{lim}}$ , and we identify  $H_0$  as the value of the Hubble constant today,  $c$  as the speed of light,  $\chi$  as the comoving coordinate,  $a$  as the

scale factor,  $f_K$  as the comoving angular distance, and the reduced shear,  $g$ , given by

$$g = \frac{\gamma}{1 - \kappa}. \quad (\text{A4})$$

The real-space shear two-point correlation function is the main cosmic shear observable because it can be readily obtained by averaging over the multiplied ellipticities of galaxy pairs. We can decompose the shear into its tangential component,  $\gamma_t$ , and the cross-component,  $\gamma_\times$ , which are defined as

$$\gamma_t = -\text{Re}(\gamma e^{-2i\phi}), \quad \gamma_\times = -\text{Im}(\gamma e^{-2i\phi}), \quad (\text{A5})$$

where  $\phi$  is the polar angle of the separation vector  $\theta$ . The shear correlation functions are then

$$\begin{aligned} \xi_+(\theta) &= \langle \gamma \gamma^* \rangle(\theta) = \langle \gamma_t \gamma_t \rangle(\theta) + \langle \gamma_\times \gamma_\times \rangle(\theta), \\ \xi_-(\theta) &= \text{Re}(\langle \gamma \gamma^* \rangle(\theta) e^{-4i\phi}) = \langle \gamma_t \gamma_t \rangle(\theta) - \langle \gamma_\times \gamma_\times \rangle(\theta). \end{aligned} \quad (\text{A6})$$

Its power spectrum is given by the Fourier transform of the correlation function,

$$\begin{aligned} \langle \tilde{\gamma}(\ell) \tilde{\gamma}^*(\ell') \rangle &= (2\pi)^2 \delta_D(\ell - \ell') P_\kappa(\ell), \\ \langle \tilde{\gamma}(\ell) \tilde{\gamma}(\ell') \rangle &= (2\pi)^2 \delta_D(\ell - \ell') e^{4i\phi} P_\kappa(\ell), \end{aligned} \quad (\text{A7})$$

with  $\delta_D$  being the Dirac function and using equation (A2) to write  $P_\kappa(\ell) = P_\gamma(\ell)$ . Finally, if we choose  $\theta = (\theta, 0)$ , we have

$$\begin{aligned} \xi_+(\theta) &= \langle \gamma(\theta) \gamma^*(\theta) \rangle \\ &= \int \frac{d^2\ell}{(2\pi)^2} \int \frac{d^2\ell'}{(2\pi)^2} e^{i\ell' \cdot \theta} \langle \tilde{\gamma}(\ell) \tilde{\gamma}^*(\ell') \rangle \\ &= \int \frac{d^2\ell'}{(2\pi)^2} e^{i\ell' \cdot \theta} \int d^2\ell \delta_D(\ell - \ell') P_\kappa(\ell) \\ &= \int \frac{d^2\ell}{(2\pi)^2} e^{i\ell \cdot \theta} P_\kappa(\ell), \end{aligned} \quad (\text{A8})$$

where we make the substitution  $\ell' \rightarrow \ell$ . If we expand  $\int d^2\ell = \int \ell d\ell \int d\phi$ ,

$$= \int_0^\infty \frac{\ell d\ell}{(2\pi)^2} P_\kappa(\ell) \int_0^{2\pi} e^{i\ell\theta \cos\phi} d\phi. \quad (\text{A9})$$

Substituting

$$\int_0^{2\pi} e^{i\ell\theta \cos\phi} d\phi = 2\pi J_0(\ell\theta), \quad (\text{A10})$$

we arrive at the well-known result,

$$\xi_+(\theta) = \int_0^\infty \frac{\ell d\ell}{2\pi} J_0(\ell\theta) P_\kappa(\ell). \quad (\text{A11})$$

Similarly, we obtain

$$\xi_-(\theta) = \int_0^\infty \frac{\ell d\ell}{2\pi} J_4(\ell\theta) P_\kappa(\ell). \quad (\text{A12})$$

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.