On approximations of the redshift-space bispectrum and power spectrum multipoles covariance matrix

Sergi Novell-Masot [ID],[a,b] **Héctor Gil-Marín** [ID][a,b] **and Licia Verde**[a,c]

[a]*Institut de Ciències del Cosmos (ICCUB), University of Barcelona,*
C. Martí i Franquès, 1, E-08028 Barcelona, Spain

[b]*Institut d'Estudis Espacials de Catalunya (IEEC),*
C. Gran Capità, 2-4, E-08034 Barcelona, Spain

[c]*Institució Catalana de Recerca i Estudis Avançats (ICREA),*
Pg. Lluís Companys 23, E-08010 Barcelona, Spain

E-mail: sergi.novell@icc.ub.edu, hectorgil@icc.ub.edu,
liciaverde@icc.ub.edu

ABSTRACT: We investigate, in dark matter and galaxy mocks, the effects of approximating the galaxy power spectrum-bispectrum estimated covariance as a diagonal matrix, for an analysis that aligns with the specifications of recent and upcoming galaxy surveys. We find that, for a joint power spectrum and bispectrum data-vector, with corresponding k-ranges of $0.02 < k\,[h\mathrm{Mpc}^{-1}] < 0.15$ and $0.02 < k\,[h\mathrm{Mpc}^{-1}] < 0.12$ each, the diagonal covariance approximation recovers $\sim 10\%$ larger error-bars on the parameters $\{\sigma_8, f, \alpha_\parallel, \alpha_\perp\}$ with respect to the full covariance case, while still underestimating the corresponding true errors on the recovered parameters by $\sim 10\%$. This is caused by the diagonal approximations weighting the elements of the data-vector in a sub-optimal way, resulting in a less efficient estimator, with poor coverage properties, than the maximum likelihood estimator featuring the full covariance matrix. We further investigate intermediate approximations to the full covariance matrix, with up to $\sim 80\%$ of the matrix elements being zero, which could be advantageous for theoretical and hybrid approaches. We expect these results to be qualitatively insensitive to variations of the total cosmological volume, depending primarily on the bin size and shot-noise, thus making them particularly significant for present and future galaxy surveys.

Contents

1 Introduction

The foundation of cosmological parameters inference rests on the fact that the maximum likelihood estimator (MLE) is the best unbiased estimator in large data sets (e.g., [1] for a cosmologist-friendly presentation). In most applications, and in particular in large-scale structure clustering, the likelihood is assumed to be Gaussian.

In this context, while the power spectrum is the primary large-scale structure clustering statistic, it has been shown that including the redshift space bispectrum significantly enhances the constraining power of galaxy surveys (e.g., [2–4]), by breaking cosmological parameter degeneracies and reducing error-bars on key cosmological parameters.

In principle, even if the initial density field were to be Gaussian, the distribution of the n-point functions, such as the three-point correlation function (3PCF) or the bispectrum (its counterpart in Fourier space) is not Gaussian. However, these summary statistics are usually binned (in k or scales): for large survey volumes, each bin is populated by many data points and, thanks to the central limit theorem, the resulting statistics for the bins approach Gaussianity. In particular, the joint likelihood of the power spectrum and bispectrum multipoles is assumed to be a multi-variate Gaussian, and this approximation has been shown to hold well in practice, in regimes without too strong non-linearities [5–9].

Once a data-vector and a model for the data-vector (also referred to as a signal) are given, the full covariance matrix is the key ingredient for cosmological inference. The evaluation of this matrix beyond the purely linear regime is challenging: analytic expressions quickly

become long and cumbersome to evaluate [9, 10] especially if all real-world effects present in a real survey need to be taken into account. Alternatively, the covariance matrix can be evaluated from multiple mock survey realizations (see for example [11–14] for the specific case involving the redshift space bispectrum multipoles). In this case, the number of realizations needs to be significantly larger than the number of data-vector elements [15–18]. Here the challenge lies in the fact that when including bispectrum multipoles the data-vector can easily reach a length of ~ 1000 elements, requiring sometimes a prohibitive number of mocks.

All these challenges could be significantly eased if the off-diagonal terms of the covariance could be ignored. Some work in the literature has advocated for this approach. For instance [19] used a diagonal covariance with diagonal elements estimated from a set of simulations. A different "flavour" of diagonal covariance matrix is the so-called Gaussian covariance matrix which is not only diagonal but also non-linearities are neglected. Several studies have been using a Gaussian covariance matrix: either modelling it perturbatively [20, 21] or using a Gaussian random field template (e.g., [22–24] for the 3-point correlation function, and [25] for the 2-point correlation function). In these (non-exhaustive) references, the choice of Gaussian covariance is justified by invoking the linearity of the scales present in the analysis. In summary, the motivation behind the choice of a diagonal covariance matrix is not only the simplicity but also that the resulting (forecasted) errors do not seem to be much affected. It would not be remiss to hope that this could be a sufficiently good approximation in the quasi-linear regimes, where all these analyses are based.

For present and upcoming galaxy clustering analyses, such as DESI [26], Euclid [27] or LSST [28], relevant information will be extracted from scales where non-gaussianities are not negligible. Notable efforts in developing analytical templates for higher-order correlators with non-Gaussian contributions have been developed by e.g. [9, 29, 30]. It is then worthwhile and timely to investigate the magnitude of the penalty incurred when using a diagonal approximation for a bispectrum analysis up to mildly non-linear scales.

In the context of a MLE for cosmological inference, using a diagonal covariance matrix would imply adopting an estimator that, while being still (asymptotically) unbiased, is not the best-unbiased estimator — see e.g. [31] for primordial non-Gaussianities studies. In other words, the weighting given to the data-vector elements is not optimal. In this paper, we elucidate the importance of including the off-diagonal elements in the covariance matrix for cosmological inference from the bispectrum summary statistic.

The rest of this paper is organized as follows. In section 2 we present the suites of simulations we use and the set-up. In section 3 we compare the effects of assuming a diagonal covariance matrix, and we review a few intermediate alternative approximations in section 4. In section 5 we offer an analytic description of the effects found on simulations. Finally, we conclude in section 6.

2 Simulations and set-up

We review the main aspects of our analysis set-up and simulation data, which is very similar to what we used in [11] — to which we refer the readers for a more detailed description. The two main simulation sets used in this work are the QUIJOTE suite [32] and the PATCHY galaxy mocks [33]. These are cubic periodic boxes where real-world effects such as window

function, selection function, or systematic weights, are not included. The volume of the QUIJOTE realizations is 1 $(\mathrm{Gpc}\,h^{-1})^3$, whereas the PATCHY galaxy mocks have volume 15.6 $(\mathrm{Gpc}\,h^{-1})^3$ each.

In both cases, the data-vector consists of the power spectrum and bispectrum monopole and quadrupoles: $\mathcal{D} = \{P_0, P_2, B_0, B_{200}, B_{020}\}$. We use the bispectrum multipoles expansion first derived in [34]. Note that we have neglected one bispectrum quadrupole configuration, B_{002}, as it has been shown that it does not add information to our analysis in the scales of interest [11, 35]. We discard the isosceles configurations of the form (k_1, k_1, k_2) of the B_{020} data-vector, due to them being redundant: $B_{200}(k_1, k_1, k_2) = B_{020}(k_1, k_1, k_2)$ for all possible k_1, k_2.

We consider k-ranges of $0.02 < k^P\,[h\mathrm{Mpc}^{-1}] < 0.15$ and $0.02 < k^B\,[h\mathrm{Mpc}^{-1}] < 0.12$, with the superscripts P and B indicating power spectrum and bispectrum respectively. Following ref. [14] the k-vectors are binned with $\Delta k = 1.1 k_\mathrm{f} \approx 0.0069\,h\,\mathrm{Mpc}^{-1}$ (with $k_\mathrm{f} = 2\pi/L_\mathrm{box}$ the fundamental frequency) for QUIJOTE, while for PATCHY the binning is such that $\Delta k = 0.01\,h\,\mathrm{Mpc}^{-1}$.

We model the power spectrum up to two loops in renormalized perturbation theory (RPT) [36, 37], whereas for the bispectrum the recently presented phenomenological model GEO-FPT[1] [11] is employed. The full parameter space is given by $\{\sigma_8, f, \alpha_\parallel, \alpha_\perp, b_1, b_2, A_P, \sigma_P, A_B, \sigma_B\}$. These are, respectively, the amplitude of dark matter fluctuations smoothed by a top-hat filter of $8\,\mathrm{Mpc}h^{-1}$, σ_8, the logarithmic growth rate, f, the dilation parameters along and across the line-of-sight, $\alpha_\parallel$ and $\alpha_\perp$ [38–41]; the linear and non-linear galaxy bias parameters b_1, b_2, where the local Lagrangian bias expansion is assumed [42–44];[2] the phenomenological amplitude parameters that regulate the deviations from shot-noise (A_P, A_B [19, 41]), and the Fingers-of-God damping factors σ_P, σ_B [45], considered independent for the power spectrum and bispectrum. The subset $\{\sigma_8, f, \alpha_\parallel, \alpha_\perp\}$ holds more relevance for cosmology and is the focus of this work, while the rest are treated as nuisance parameters.

We use 8000 realizations of the fiducial set of QUIJOTE dark matter simulations at redshift $z = 0.5$ to estimate the full covariance matrix. This large number of realizations ensures a reliable estimate of the covariance, for a fairly large number of elements in our data-vector (~ 1200). Then, we also consider 2000 (of these 8000) simulations to be analyzed individually, thus obtaining 2000 sets of (single realization) best-fit parameters. We compare the distribution of these best-fit parameters with the MCMC posteriors derived by fitting the averaged data vector of the same 2000 simulations.

Likewise, for the PATCHY mocks we employ the available 2000 realizations at redshift $z = 0.53$; we fit cosmological parameters from them individually, and from their mean power spectrum and bispectrum signal (to reduce cosmic variance). In this case, the full covariance matrix is estimated from 2000 realizations, as this is the maximum available number.

The full power spectrum and bispectrum covariance matrix can be schematically represented in blocks as shown in figure 1. In all cases, at the likelihood evaluation level, we correct for the inverse covariance matrix estimation bias using the Sellentin and Heaven

---

[1]https://github.com/serginovell/geo-fpt.

[2]This expansion fully determines the tidal bias and non-linear bias, respectively b_{s^2}, b_{3nl}, from the linear bias parameter b_1.
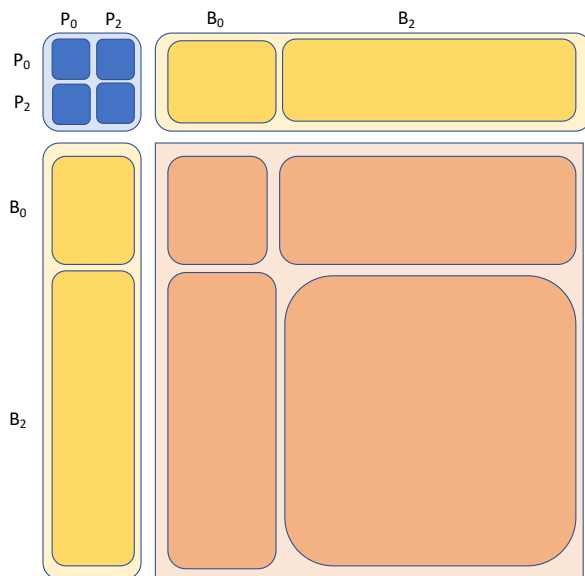
Figure 1. Schematic representation of the full power spectrum-bispectrum covariance matrix. Here P_0 denotes the power spectrum monopole portion of the data-vector, P_2 being the power spectrum quadrupole portion. Similarly, B_0 and B_2 indicate the bispectrum monopole and quadrupole portions of the data-vector respectively. In the *full* approximation the matrix is dense in the sense that all elements of all blocks are non-zero even though many are small. In the *diagonal* approximation the off-diagonal blocks are all zero and the blocks on the diagonal, e.g., the $P_0 - P_0$, $B_0 - B_0$ and $B_2 - B_2$ blocks, are diagonal matrices. Intermediate approximations discussed in section 4 populate the off-diagonal block and the off-diagonal elements of the diagonal blocks according to some case-specific rules. The blue, yellow, and orange sub-matrices will be referred to as PP, PB, and BB blocks, respectively.

prescription [15]. To estimate the magnitude of this effect it is useful to report that the length of the data-vector is 1160 for QUIJOTE and 441 for PATCHY; given the available number of simulations, the corresponding Hartlap correction factor is respectively $H = 0.85$ for QUIJOTE, and $H = 0.78$ for PATCHY.[3]

3 Diagonal approximation: effects in simulations

In order to quantify the effect of assuming a diagonal covariance matrix, we compare the cosmological constraints for an analysis involving the power spectrum and bispectrum monopole and quadrupoles, using both the full covariance matrix and the diagonal approximation. In this approximation, all off-diagonal correlations for the power spectrum, the bispectrum, and the cross power-spectrum bispectrum are set to zero (see figure 1).

In what follows we compute constraints (i.e., recovered values and estimated errors) for the key cosmological parameters under different approximations and conditions. For both dark matter N-body (QUIJOTE) and galaxy (PATCHY) simulations we compute the four cases that are specified in table 1.

---

[3]While the correction is performed using the Sellentin-Heavens prescription, the Hartlap factor (see equation 17 of [16]) $H = (n - m - 2)/(n - 1)$ where n is the number of simulations and m the size of the data-vector, only indicates an order of magnitude of the noise introduced when inverting the covariance.

	Posterior of the mean data vector	**Distribution of individual best-fits**
Full Covariance Matrix	Posteriors, recovered via MCMC using full covariance matrix, from the mean of the data vectors of 2000 simulations	Distribution of best fit parameters, recovered via MCMC using full covariance matrix, for the 2000 individual simulations
Diagonal Covariance Matrix	Posteriors, recovered via MCMC using diagonal covariance matrix, from the mean of the data vectors of 2000 simulations	Distribution of best fit parameters, recovered via MCMC using diagonal covariance matrix, for the 2000 individual simulations

Table 1. The four methods of inference we perform on both PATCHY and QUIJOTE simulations. For respectively the full covariance and diagonal approximation, we perform both the standard MCMC sampling for the parameters of interest, and the distribution of best-fit parameters obtained individually in each of the 2000 simulations.

Any non-optimal weighting is expected to produce uncertainties on the inferred parameters that are larger than those achievable with an optimal weighting. Moreover, the inferred posterior provided by the MCMC might not yield error-bars that have the correct coverage properties [46].[4] On the other hand, the scatter among individual simulations' best-fit parameters is expected to have better coverage properties for a large enough number of simulations.

Throughout the following subsections, the diagonal covariance approximation is confirmed to underestimate the errors of the recovered cosmological parameters for analysis set-ups similar to the one in section 2. Section 4 tackles the intermediate case where not enough simulations are available in order to estimate or calibrate the full covariance matrix. There, other sparse matrix approximations to the full covariance matrix are then offered, along with their potential for reducing the number of required simulations for covariance estimation.

3.1 Parameter constraints: MCMC and scatter

Given 2000 independent realizations of both the QUIJOTE dark matter N-body simulations and the PATCHY galaxy mocks (presented in section 2), we proceed to compare the recovered parameters $\{\sigma_8, f, \alpha_\parallel, \alpha_\perp\}$ and their estimated uncertainties with two different methods. In particular, we contrast the constraints obtained via the traditional approach of MCMC sampling of the posterior distributions (credible intervals), with the scatter of maximum likelihood values for each realization (confidence intervals). If the statistical approximations adopted — mainly the Gaussian likelihood and optionally the diagonal covariance — are correct, then the MCMC posteriors would have correct coverage properties and thus the two approaches would yield comparable error-bars.

---

[4]The statistical concept of coverage refers to the probability that a confidence interval correctly contains the true values of the parameters being estimated. These confidence intervals do not necessarily coincide with the error-bars derived via MCMC — the *credible intervals* —, which represent the probability range of parameters based on prior beliefs and observed data.

Figure 2 shows that in both sets of simulations, the diagonal covariance MCMC fails to capture the information that is obtained with the scatter of maximum likelihood values, especially for the dilation parameters $\alpha_\parallel, \alpha_\perp$. The diagonal approximation produces *a)* a larger scatter hence larger actual uncertainties; and *b)* an incorrect posterior estimate via the MCMC, which produces an underestimate of the actual errors (i.e., the diagonal approximation yields credible intervals that are smaller than the corresponding confidence intervals and thus have no *coverage*). This behavior is not unexpected at all: the diagonal covariance approximation makes the adopted estimator and likelihood suboptimal as anticipated above and discussed also in [18, 46].

In the QUIJOTE simulations, the scatter of maximum likelihood values in the diagonal approximation is noticeably larger than the MCMC posteriors, while in the PATCHY mocks the effect is milder, mostly affecting the tails of the distribution. This may be caused by the combination of the difference in tracer type, shot-noise and bin-size; in particular, QUIJOTE has negligible shot-noise and a smaller bin size, as specified in section 2 and in [11]. For simplicity, in what follows we focus on the PATCHY mocks.[5]

3.2 Full and diagonal covariance: effects on $\boldsymbol{\alpha_\parallel}$ and $\boldsymbol{\alpha_\perp}$

Motivated by the fact that the effect of the diagonal approximation seems to propagate mainly to the dilation parameters, in the main text we only show and discuss the results for $\{\alpha_\parallel, \alpha_\perp\}$. The σ_8 and f cases are displayed in appendix A.[6]

Figure 3 presents the relationship between the recovered error-bars, obtained via MCMC, using a diagonal ($\sigma_{\rm diag}$) and full ($\sigma_{\rm full}$) covariance, for the dilation parameters. The identity line is also shown for guidance. As expected from figure 2 the diagonal approximation yields larger error-bars ($\sigma_{\rm diag} \sim 1.2\sigma_{\rm full}$). In addition, a weak correlation is observed between the derived errors from the full covariance and from the diagonal part only.

An alternative perspective on the impact of the diagonal covariance matrix approximation can be gained as follows. Let us consider the statistic z, defined as [47],

$$z_{\theta_j} \equiv \frac{\theta_j - \mu_{\theta_j}}{\sigma(\theta_j)}, \tag{3.1}$$

where θ_j denotes one of the parameters of interest, i.e. $\theta_j \in \{\sigma_8, f, \alpha_\parallel, \alpha_\perp\}$. For each i-realization of the 2000 simulations, θ_j^i denotes the best-fitting parameter value recovered via MCMC and $\sigma(\theta_j^i)$ is its error estimated from the MCMC posterior; μ_{θ_j} denotes the mean across all simulations of the recovered parameter θ_j^i. We expand on the features of the z statistic in appendix C.

---

[5]We have checked that the results presented hereafter are qualitatively equivalent in both QUIJOTE and PATCHY simulation sets, with a small difference in magnitude. We choose to present the results from the PATCHY mocks for two reasons: firstly, their closer resemblance to galaxy data compared to the QUIJOTE simulations; and secondly, as demonstrated in figure 2, the PATCHY simulations exhibit superior performance under the diagonal approximation. This implies that, should the diagonal approximation prove unsuitable for PATCHY, its performance in the QUIJOTE simulations would be equally, if not more, limited.

[6]This difference between the two sets of parameters could be due to there being more proportion of information content for $\alpha_\parallel, \alpha_\perp$ in smaller scales (where the diagonal approximation performs the worst) than for f, σ_8.
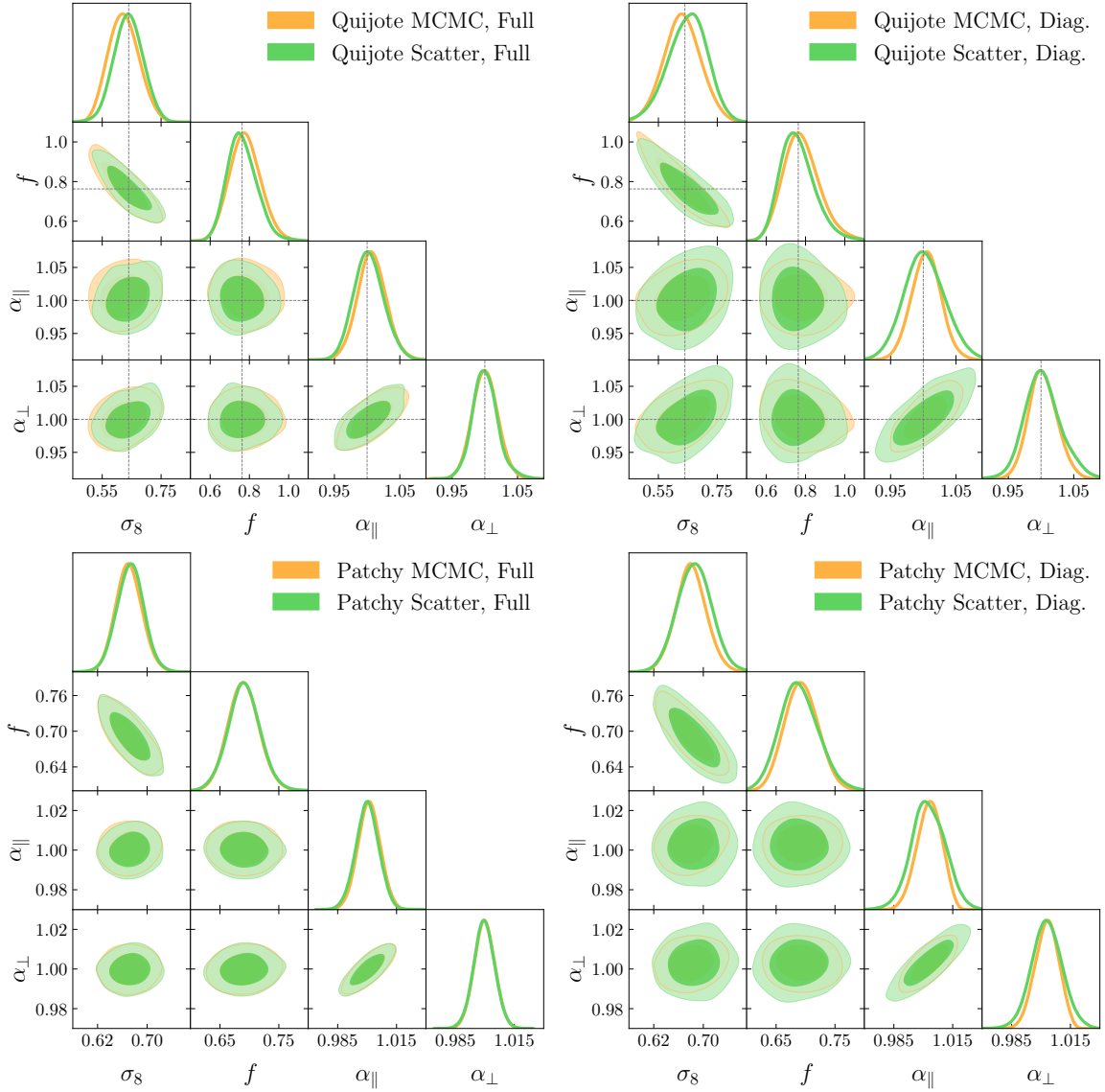
Figure 2. Estimated posteriors for the main cosmological parameters with the data-vector $P_0 + P_2 + B_0 + B_{200} + B_{020}$ from 2000 QUIJOTE fiducial dark matter simulations (upper row) and the 2000 PATCHY galaxy FastMocks (lower row). In the left plots, the full covariance (estimated from the simulations) is employed, while in the right plots, the diagonal approximation is used. The orange contours are *credible intervals* estimated from the posteriors sampled via MCMC, using the mean of the 2000 simulations as a data-vector. The green contours are obtained from the scatter of the maximum likelihood parameters for each of the 2000 simulations individually, hence represent *confidence intervals*. The distribution inferred from the scatter of the maximum likelihood values is smoothed via a kernel density estimator to provide smooth confidence contours to compare with those inferred from the MCMC. The power spectrum is modeled at 2L-RPT and for the bispectrum we employ the GEO-FPT model presented in [11]. The axes size for each set of simulations is kept fixed to facilitate visual comparison of the posteriors between the two covariance approaches. Using the diagonal approximation produces *a)* larger scatter hence larger actual uncertainties, and *b)* causes the MCMC approach to underestimate the (actual) errors.
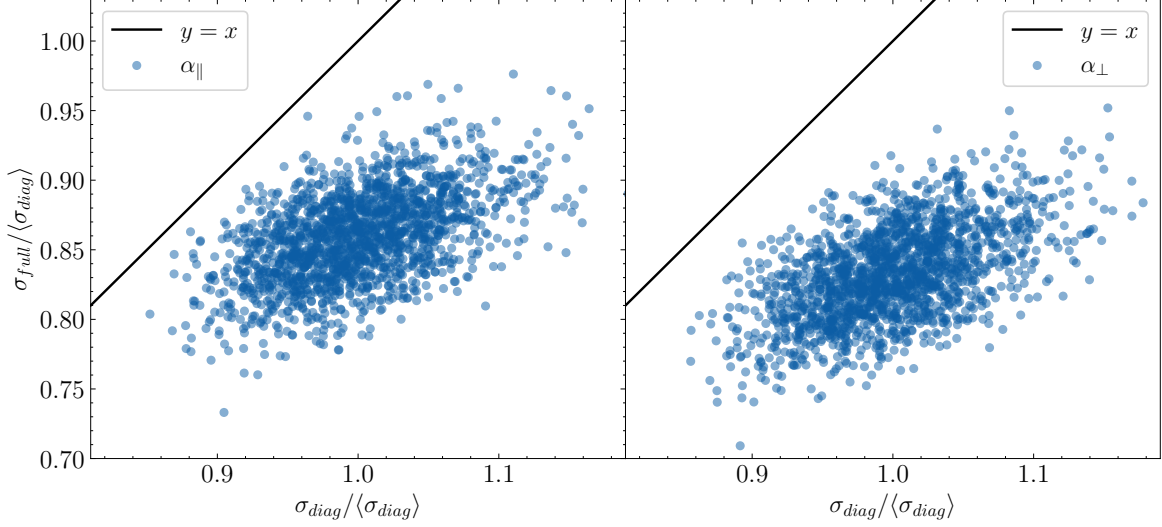
Figure 3. Scatter of the marginalized 1σ regions for $\alpha_\parallel$ (left) and $\alpha_\perp$ (right) from 2000 PATCHY mocks. On the x-axis we show the MCMC-recovered 1σ region using the diagonal covariance, σ_{diag}, on the y-axis the corresponding 1σ region, using the full covariance matrix, σ_{full}. Both are normalized by the mean 1σ region with the diagonal covariance matrix, $\langle\sigma_{\mathrm{diag}}\rangle$ for easy visualization. Each point corresponds to one of the 2000 PATCHY realizations. The identity line $y = x$ is over-plotted, together with the best-fit regression line to the scatter of points. The recovered error-bars from MCMCs are larger when using a diagonal covariance matrix ($\sigma_{\mathrm{diag}} \sim 1.2\sigma_{\mathrm{full}}$).

Since the simulations are effectively independent and identically distributed realizations of the underlying fiducial cosmology, if the errors are correctly estimated, the distribution of the 2000 values of z_{θ_j}, by the central limit theorem, should converge to a normal with variance 1, $\mathcal{N}(0, 1)$. Consequently, σ_z (the standard deviation of the z statistic across the 2000 realizations) should converge to 1. Some (small) mis-match is expected, for at least two reasons. The estimated covariance, which enters in the denominator of equation (3.1) has an associated error, which means that the true covariance could have been different than that adopted [16, 48]. The likelihood is non-Gaussian so matching the *rms* with the 1-σ of a Gaussian distribution may not be exact. We return to the first effect below; nevertheless relative statements can still be made.

Figure 4 shows the distribution of $z_{\alpha_\parallel}$ (left panel) and $z_{\alpha_\perp}$ (right panel) for the full (orange) and diagonal (blue) covariance matrix cases. The deviations from $\mathcal{N}(0, 1)$ for z_θ offer a metric to assess the accuracy in the estimation of the errors. From the definition of z_θ in equation (3.1), if the error-bars on the cosmological parameters are consistently underestimated (resp. overestimated), then $\sigma_z > 1$ (resp. $\sigma_z < 1$) — with σ_z being the standard deviation for the distribution of z_θ. For example, for the two dilation parameters we find that for the full covariance $\sigma_z = 0.924$ while for the diagonal approximation $\sigma_z = 1.17\,(1.13)$ for $\alpha_\parallel(\alpha_\perp)$. Hence the true errors appear to be over-estimated by about $\sim 8\%$ when employing the full covariance; when employing the diagonal matrix they appear under-estimated by $17\%(13\%)$ for $\alpha_\parallel(\alpha_\perp)$.
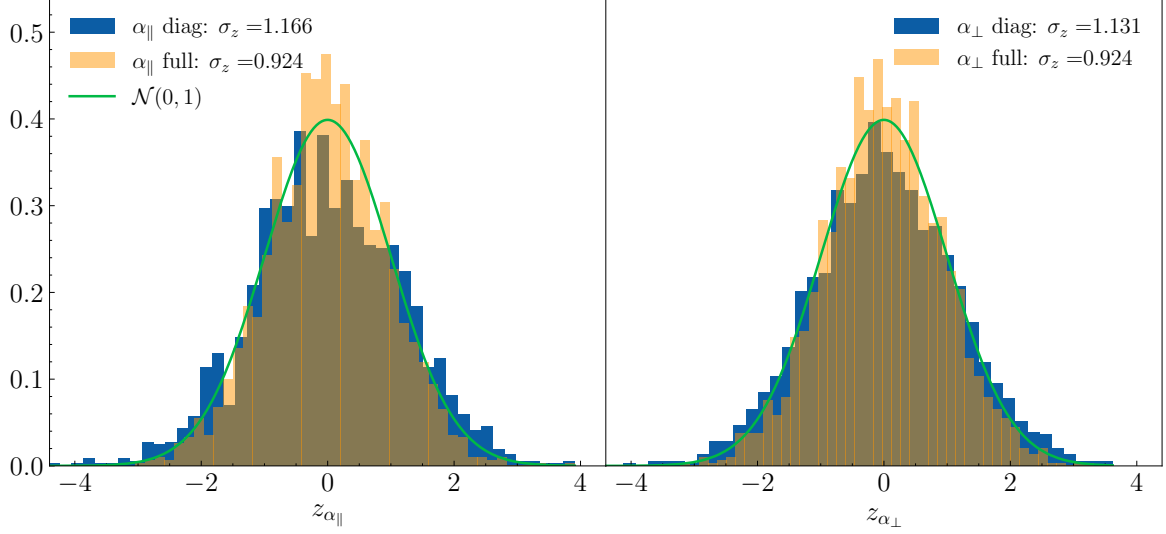
Figure 4. Histogram of the z values for the full and diagonal covariance matrix cases, for the parameters $\alpha_\parallel$ (left panel) and $\alpha_\perp$ (right panel). The standard deviation of z (σ_z) is shown for each case. The solid line, the standard normal distribution $\mathcal{N}(0,1)$, corresponds to the theoretical shape of z where the errors $\sigma(\theta)$ are perfectly estimated. The diagonal approximation (blue) tends to underestimate the error-bars in the cosmological parameters, while the full covariance (orange) instead tends to overestimate them.

This figure confirms the findings of figure 2: the cosmological inference results obtained from running an MCMC assuming a diagonal covariance will underestimate the error-bars, and the constraints will also be sub-optimal. Nevertheless, we find no evidence of a systematic shift in the central values of the recovered parameters: the recovered best fit parameters are unbiased. This again boils down to the diagonal approximation not being a maximum likelihood estimator — the estimator is unbiased but sub-optimal. The inference using the full covariance matrix as estimated from the 2000 simulations and the Sellentin and Heavens correction [15] appears to be, according to the z_θ metric, slightly conservative.

We repeated the same procedure also for the QUIJOTE set ($H = 0.85$) and for the power spectrum part of the signal, P_{02}, in the PATCHY set (the reduced size of the data vector yields $H = 0.99$). In these cases we find $\sigma_z = 1$ (PATCHY, P_{02} only) and $\sigma_z = 0.97$ (QUIJOTE, $P_{02} + B_{02}$) for the α parameters, i.e. the errors are correctly estimated at better than 3%. As expected, the parameters which have markedly non-Gaussian posterior tails (i.e. f, σ_8 from the power spectrum alone), do not perform as well. For comparison diagonal covariance matrices mis-estimate the error by $\sim 20\%$. We tentatively conclude that the appearance of σ_z slightly above unity ($\sim 8\%$) for the full covariance case in figure 4 is correlated to the noise in the estimation of the covariance matrix, which is driven by the number of simulations and the size of the data vector as quantified by the Hartlap factor.

We will not dwell more on this except for drawing the following conclusion: while the full covariance matrix estimated from the available number of simulations is not the ground truth, it is good enough to be treated as our baseline for comparison for various approximations. The diagonal approximation underestimates the uncertainties by $\sim 15\%$.

4   Intermediate approximations

The diagonal approximation for the power spectrum and bispectrum covariance matrix may therefore be a good approximation for Fisher-matrix-based error forecasts, but not accurate enough for precision cosmology from ongoing surveys. We may however investigate whether there are sparse matrix approximations to the full covariance that may potentially require fewer simulations to estimate [49–54] and perform effectively as well as the full covariance.[7] We reiterate that the full covariance matrix estimated from simulations is not the ground truth but it is the best estimate we have which therefore we use as our baseline. There is significant literature associated with modeling with random matrices the properties of noisy estimates of covariance matrices. We will only refer to this tangentially here.

In what follows, we explore a small set of possible approximations, leaving out of the scope of this work valuable contributions that go towards the same direction, such as data-vector compression [55–57], covariance matrix shrinkage [58–60] or resampling [61, 62].

We refer to PP, PB, and BB, respectively, as the blue, yellow, and orange sub-matrices of the full matrix shown in figure 1. We consider the following approximations, all of them being positive-definite, having the full power spectrum auto-covariance PP and of course all the diagonal terms:

- "Auto": include the full PP and BB boxes, setting to zero the PB cross-covariance.

- "Common-k": include the full PP box, and in PB, BB only the terms which correspond to configurations that have a k-vector in common.

- "τ-threshold": beyond the full PP box, only include the terms that are higher than a certain threshold τ in the reduced covariance matrix.

- "(τ)Common-k": apply the τ threshold, as stated above, to the "Common-k" case.

Several additional approximations were initially considered, including the modifications to the "Common-k" approach restricting the BB box to only include the off-diagonal terms corresponding to the same triangle (k_1, k_2, k_3), and also the case with all coefficients for PP and PB and with a diagonal BB box. These approximations yielded matrices that were not positive-definite and hence were discarded. Thresholds over ~ 0.04 in the τ-threshold approximation are also not positive definite for our present analysis. We adopt a threshold of 3%: the resulting matrix offers a balance between being as sparse as possible while maintaining similarity with the full covariance matrix. Note that there may be a connection between this value and the findings of [63], where a level of noise of $\sim 2\%$ is reported in the covariance matrices of the PATCHY simulations.

A quantitative metric to estimate the "closeness" of each approximation to the reference (full covariance matrix) is the ratio of the eigenvalues. This is reminiscent of the figure-of-merit for the cosmological parameters which is given by the square root of the determinant of the Fisher matrix for the parameters — the determinant is of course the product of the matrix eigenvalues. The analogy arises because the (inverse) Fisher matrix is related to

---

[7]Note that in the last reference, [54], which holds special interest given its formulation in terms of cosmology, the focus is on the sparsity of the precision matrix, the inverse of the covariance matrix.
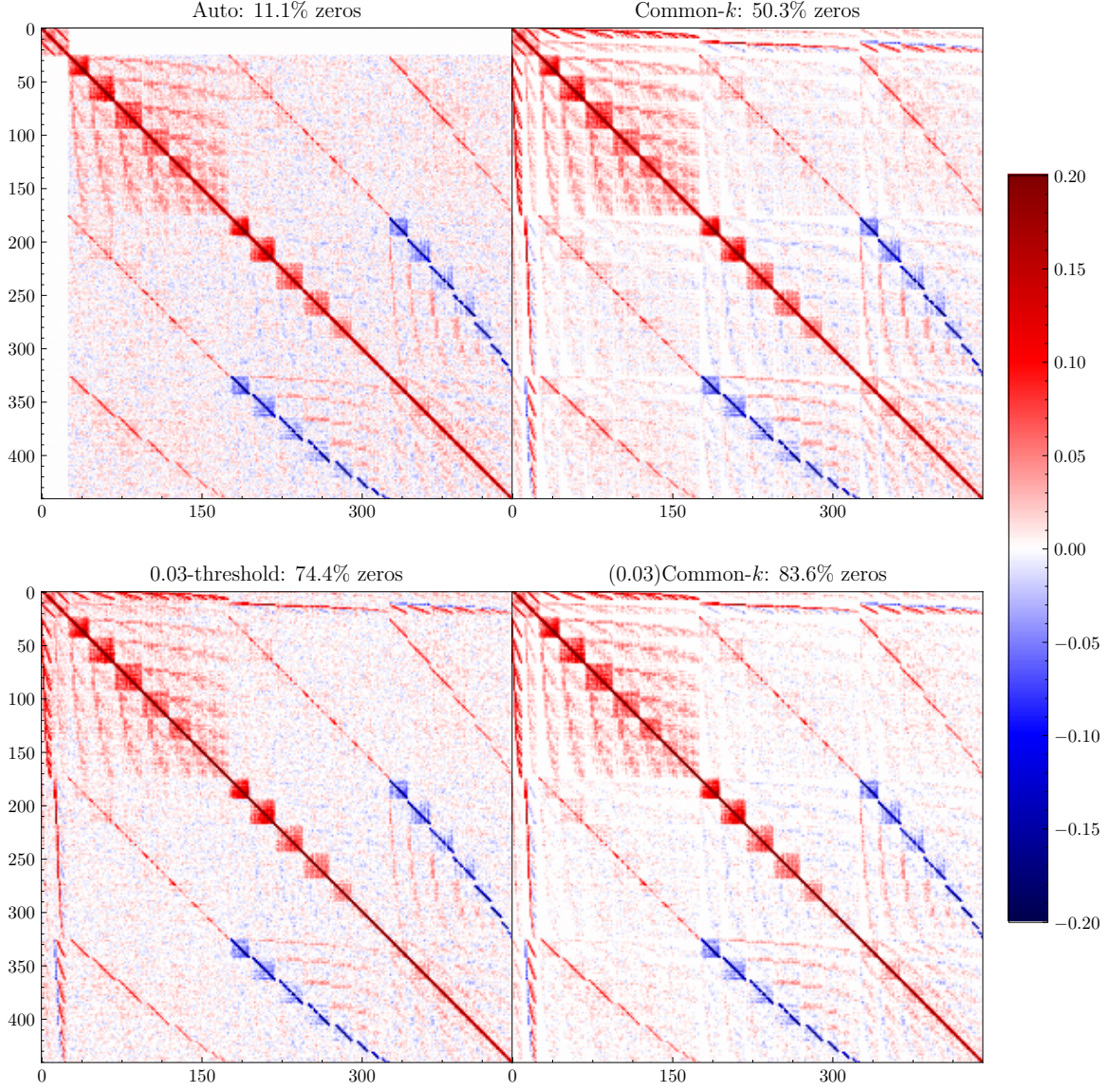
Figure 5. Reduced covariance matrices for each of the intermediate approximations to the covariance matrix. We show as well the sparsity of the matrix, represented by the percentage of coefficients that are set to zero in each case. While the Common-k approximation is the one that is closest to the full case while not underestimating the errors (see figure 6), it still only has half of its coefficients set to zero. The figure has an apparent curvature, caused by removing the duplicate isosceles configurations in the bispectrum data-vector, as explained in section 2.

the cosmological parameters errors while the covariance matrix is related to the data-vector errors. The eigenvalues of the inverse covariance matrix can be seen to represent the signal or information content of the data-vector. By ordering the eigenvalues in decreasing order it is easier to visualize where most of the information content is localized. As expected, the first eigenvalues are those that directly involve the power spectrum. Moreover, to compress the "distance" of an approximation to the full matrix case we also report the logarithm of the ratio of the product of the eigenvalues and the Kullback-Leiber (KL) divergence (or relative entropy, or information gain) between the two distributions that the matrices represent [64, 65]. Assuming that the distributions are multivariate Gaussians, for our case (where the two have the same mean) this quantity reads as [66]

$$D_{KL}(C_{\text{full}}||C) = \frac{1}{2}\left(\text{tr}(C_{\text{full}}C^{-1}) - n + \ln\left(\frac{\det C}{\det C_{\text{full}}}\right)\right), \tag{4.1}$$

with n the dimension of the data-vector.[8] Figure 6 shows the ratios of the (sorted in decreasing order) eigenvalues of the inverse covariance matrix C^{-1} with respect to the full covariance case C_{full}^{-1} for each of the approximations, including the diagonal case. The legend reports the Kullback-Leibler distance, which is complemented by table 2 with other indicators of distance towards the full covariance matrix. All this serves as figures of merit, to assess and rank-order both the closeness to the full covariance matrix and also the expected discrepancy in the recovered error-bars.

The diagonal approximation consistently underestimates the eigenvalues of the inverse matrix. This underestimation is indicative that there is less information that can be captured, which propagates in the error-bars in the parameters, as observed in section 3 and further discussed below in section 5. Additionally, its deviation from the full covariance case — as quantified by means of the Kullback-Leibler divergence (equation (4.1)) — is significantly higher than all remaining approximations.

The "Auto" approximation, by setting to zero the PB boxes, only changes the first ~ 25 eigenvalues, which are those that change most drastically when including the PB boxes. This indicates that the PB cross-correlation is important and should not be neglected. Moreover, this approximation does not yield a sparse matrix (only 11% of the terms are zero) or a matrix that offers any speed-up or computational advantage compared to the full.

The "Common-k" approximation has all eigenvalues much closer to the full covariance case while having $\sim 50\%$ of the terms being zero. Therefore, this or a similar approach (together with suitable sparse matrix covariance methods such as [50, 51]) may be advantageous if it can ease the requirements on the number of simulations needed to estimate the covariance.

The 0.03-threshold features a consistent overestimate of the largest eigenvalues of C^{-1}, especially for the values concerning the power spectrum. Thus, this approximation on its own artificially increases, for some eigenvalues, the information content of the data-vector.

Finally, the (0.03)Common-k approach is a combination of 0.03-threshold and Common-k, which yields a covariance matrix whose eigenvalues are closer to the full case than the 0.03-

---

[8]The KL divergence is not symmetric and here we have explicitly expressed the expected excess surprise from using the distribution given by C_{approx} as a model when the actual distribution is that of C_{full}. If the two distributions are multi-variate Gaussians, as we assume here, and have the same mean (as is the case here, to the best of our knowledge) the KL divergence assumes this simplified form.
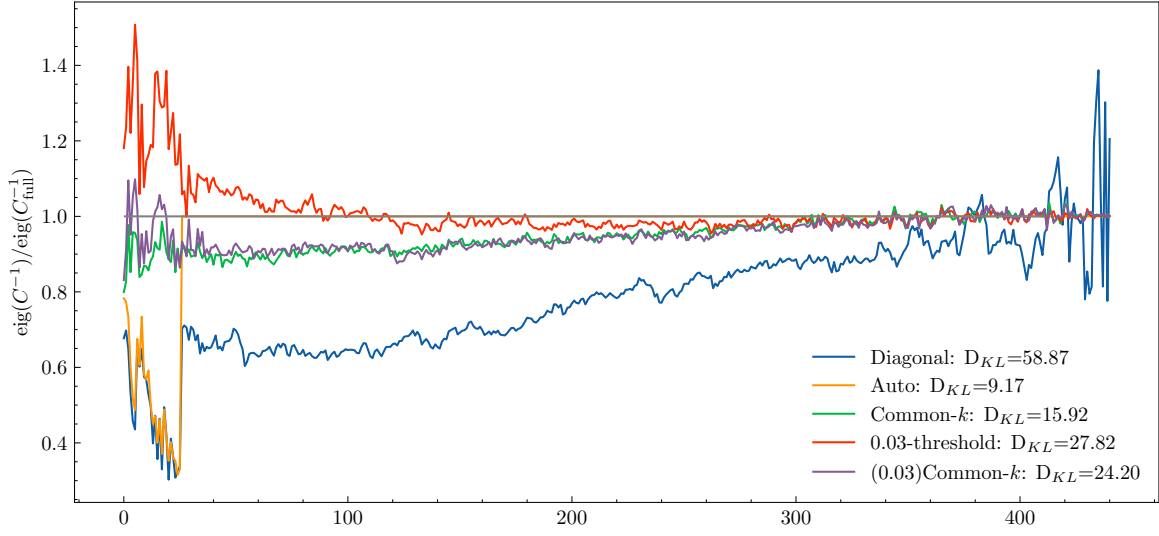
Figure 6. Ratios of the sorted (decreasing order) eigenvalues of the inverse covariance matrix C^{-1} by the inverse full covariance matrix C_{full}^{-1} for all the approximations considered. The eigenvalues of the inverse covariance act as a proxy for the information contained in their direction: the bigger the eigenvalue the more information it potentially contains. The product of the eigenvalues, the determinant, is informative of the size of the recovered error-bars in the derived parameters, so a consistent ratio under unity (as is the case for the diagonal and "Common-k" approaches) typically results in larger error-bars. The Kullback-Leibler divergence — defined in equation (4.1) — of each case with respect to the full covariance estimator is shown in the legend.

Approximation	D_{KL}	$\ln\left(\frac{\det C^{-1}}{\det C_{\text{full}}^{-1}}\right)$	$\left(\frac{\det C^{-1}}{\det C_{\text{full}}^{-1}}\right)^{1/n}$
Diagonal	58.87	-117.73	0.77
Auto	9.17	-18.32	0.96
Common-k	15.92	-23.33	0.95
0.03-threshold	27.82	4.49	1.01
(0.03)Common-k	24.20	-21.66	0.95

Table 2. Different metrics assessing the distance of the proposed approximations to the full covariance matrix. The Kullback-Leibler distance, D_{KL}, is as defined in equation (4.1), with one of its components, the logarithm of the ratio of determinants, shown in the second column. The last column quantifies the geometric mean of the ordered ratios of eigenvalues, which is the nth root of the ratio of the determinants, with n being the dimension of the data-vector. The diagonal covariance matrix performs significantly worse than the other approximations, while we do not see a gain in the Auto approximation — given that it does not reduce significantly the number of non-zero elements of the covariance matrix (see figure 5). Among the remaining approximations, the 0.03-threshold approach seems to artificially obtain more information, as also seen in figure 6, while we find the Common-k and (0.03)Common-k approximations to strike a good balance between closeness to the full covariance and proportion of zero elements (see figure 5).
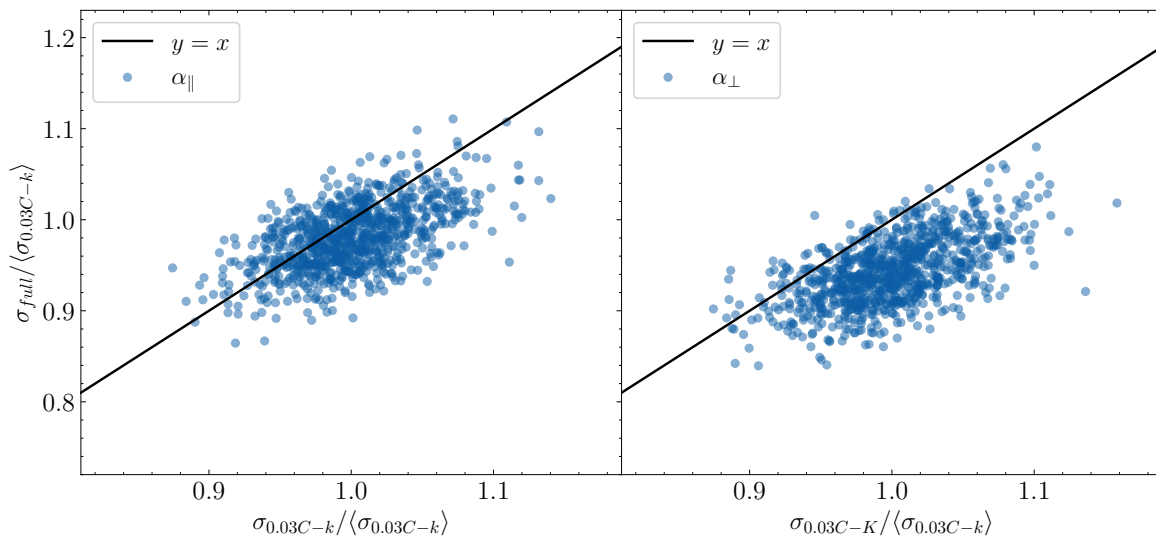
Figure 7. Analogous to figure 3, but for the (0.03)Common-k vs the full cases. Compared to figure 3 the points are scattered closer to the equality marked by the $y = x$ line indicating that the τCommon-k approximation, in the idealized case of periodic box and with the value of τ suitably chosen according to the PATCHY mocks, yields error-bars only marginally larger than the full covariance matrix There is no appreciable difference of central values for $\alpha_\parallel$, and only in $\alpha_\perp$ case there is a small mis-estimate $\sigma_{0.03C-k} \sim 1.05\sigma_{\text{full}}$.

threshold case, while having significantly more sparsity in the matrix (84% of the elements are zero, in the case presented in figure 5). Using this approximation in practice would consist of first estimating the Common-k terms — be it fully estimated from simulations or via theoretical model calibrated to simulations e.g., *à la* [9, 67] — to then establish an appropriate threshold for the set-up at hand, keeping in mind that the threshold should maintain the resulting matrix to be positive-definite. Provided that the choice of threshold is conservative, most of the removed coefficients are likely to be noise or noise-dominated, making it a valid approximation of the covariance matrix as a sparse matrix.

We repeat the tests carried out in section 3 for the "Common-k" and "(0.03)Common-k" approximations for 1000 PATCHY realizations, finding a remarkable agreement with the full covariance matrix in both cases (see figure 7, where we report the correlation of errors between the full and "(0.03)Common-k" covariances). Both approaches result in the *rms* of the z_θ statistic defined in equation (3.1) being equal or below 1 ($\sim$ 1–0.93, comparable with the full case of figures 4 and 11) for the parameters $\theta = \{\sigma_8, f, \alpha_\parallel, \alpha_\perp\}$, which is a signal that neither approximation underestimates the error-bars on the key cosmological parameters. We finally show the MCMC constraints and coverage properties for the "Common-k" and "(0.03)Common-k" approximations in figure 8.

This figure indicates that the "Common-k" or the "τCommon-k" approximations (for a suitable choice of τ and in particular one that leaves the matrix positive definite) are useful approximations that, in a cosmological parameter inference setting, could provide results virtually indistinguishable from using the full covariance matrix.
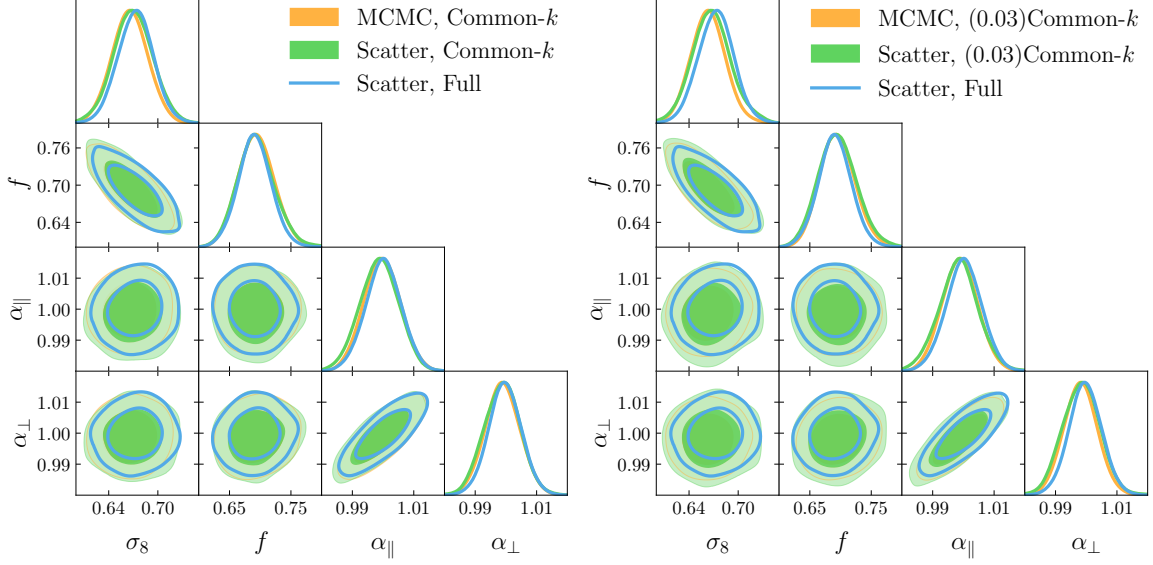
Figure 8. Analogous to figure 2, but focusing on the performance of respectively the "Common-k" and "(0.03)Common-k" approximations, compared to the full covariance case. Contrasting with the behaviour of the diagonal approximation displayed in figure 2, the proposed approximations do not exhibit a significant increase in errors with respect to the full covariance, while maintaining good coverage properties.

5 Understanding and modeling the effect of the off-diagonal terms

Before we conclude, we attempt to provide an analytic description of the effects illustrated above. Although the content of this section is well-known standard material, having it here helps the reader to put it in context and to better understand the results of the other sections of this paper.

A covariance matrix can always be decomposed in the sum of a (block) diagonal part $\mathbb{D}$ and an off-diagonal part ϵ so that

$$\mathbb{C} = \mathbb{D} + \epsilon \tag{5.1}$$

For example, $\mathbb{C}$ could be the full power spectrum plus bispectrum multipoles covariance matrix. See for example [9] for an analytic expression for the covariance terms and schematic decomposition in blocks. The effect of the off-diagonal terms within the power spectrum block of the covariance has been extensively explored in the literature and does not concern us here. We focus on the off-diagonal pure bispectrum terms and power spectrum-bispectrum terms; the power spectrum contribution can thus be considered as a block in a block diagonal-dominated matrix.

The magnitude of the ϵ_{ij} terms compared to the diagonal contribution to the covariance of the bispectrum has been studied in appendix A of [11]. There it is shown that $r_{ij} = \epsilon_{ij}/\sqrt{\mathbb{D}_i \mathbb{D}_j}$ when averaged over off-diagonal blocks involving the bispectrum, takes values up to 0.01, but when not averaged it is as high as 0.6.

If it could be assumed that $\mathbb{C}$ was diagonal-dominant (that is $\epsilon_{ij} \ll \sqrt{\mathbb{D}_i \mathbb{D}_j}$), then a possible approximation for the inverse of $\mathbb{C}$ could be [68, 69]

$$\mathbb{C}^{-1} \simeq \mathbb{D}^{-1} - \mathbb{D}^{-1} \epsilon \mathbb{D}^{-1} \equiv \mathbb{D}^{-1} - \mathbb{M} \equiv \mathcal{C}^{-1} \tag{5.2}$$

where the last equalities define the matrices $\mathbb{M}$ and $\mathcal{C}^{-1}$. For the covariance matrix involving the bispectrum we find that for the PATCHY mocks the median residual between $\mathbb{C}^{-1}$ and $\mathcal{C}^{-1}$ is -0.62, which we illustrate in figure 12 in appendix B. Hence, the diagonal-dominant approximation is completely invalid for the data-vector of interest, which already signals the importance of the off-diagonal terms of the covariance matrix. We, therefore, define the off-diagonal contribution to the covariance as the matrix $\mathbb{O}$ fulfilling $\mathbb{C}^{-1} \equiv \mathbb{D}^{-1} - \mathbb{O}$.

Note that equation (5.2) is more general than stated above since $\mathbb{D}$ does not need to be diagonal for the approximation to be valid provided that the matrix ϵ is a perturbation to the matrix $\mathbb{D}$. For example, the matrix $\mathbb{D}$ could be the thresholded covariance matrix, and $\epsilon = \mathbb{C} - \mathbb{D}$.[9]

The effect of the off-diagonal components on the χ^2 and on the response of the χ^2 to a change of parameters can be modeled as follows. Let us define the vector $\mathbf{u} = \mathbf{d} - \mathbf{t}$ where $\mathbf{d}$ denotes the data (measurements) vector and $\mathbf{t}$ the theoretical model. Then

$$\chi^2_{\mathrm{Diag}} = \sum_i u_i^2 \mathbb{D}_{ii}^{-1} = \sum_i \frac{u_i^2}{\mathbb{D}_{ii}} \tag{5.3}$$

$$\chi^2_{ND} = -\sum_{i,j} u_i \mathbb{O}_{ij} u_j \tag{5.4}$$

and of course $\chi^2 = \chi^2_{\mathrm{Diag}} + \chi^2_{ND}$. Equation (5.4) (and the following equations involving $\mathbb{O}$), if $r_{ij} \ll 1$ for $i \neq j$, could be approximated by substituting $\mathbb{M}$ for $\mathbb{O}$. In the applications considered here, this approximation is not sufficient.

For the PATCHY analysis from section 3, we find the typical contribution of the off-diagonal part of the covariance to χ^2 (evaluated at the best-fit cosmological parameters) to be of $|\chi^2_{ND}| \sim 0.04\chi^2$. This is a small proportion, but it should be noted that the importance and role of the off-diagonal terms in the covariance do not necessarily correspond to their effect on the χ^2, but rather to its response, $\Delta\chi^2$.

The response of the likelihood and the χ^2 to a change in cosmological parameters from, say, the best-fit values, that induces a change Δt on the theory model from the best-fit theory model, can be written as:

$$\Delta\chi^2_{\mathrm{Diag}} = -2 \sum_i u_i \mathbb{D}_{ii}^{-1} \Delta t_i = -2 \sum_i \frac{u_i \Delta t_i}{\mathbb{D}_{ii}} \tag{5.5}$$

$$\Delta\chi^2_{ND} = -2 \sum_{i,j} u_i \mathbb{O}_{ij} \Delta t_j \tag{5.6}$$

where again $\Delta\chi^2 = \Delta\chi^2_{\mathrm{Diag}} + \Delta\chi^2_{ND}$. Figure 3 shows that there is a weak correlation between $\Delta\chi^2_{\mathrm{Diag}}$ and $\Delta\chi^2$, and that $\Delta\chi^2_{ND}$ is not negligible compared to $\Delta\chi^2_{\mathrm{Diag}}$.

---

[9]We have checked that for the thresholds considered in this specific application this approximation is not good enough.

For an idealized case where $\mathbf{d}$ is noiseless and $\mathbf{t}$ is a perfect description of $\mathbf{d}$, for example when doing Fisher matrices forecasts, $\mathbf{u} = \mathbf{d} - \mathbf{t} = 0$: the size of the errors can be estimated from the second derivative of the log-likelihood (see below). In any practical realization $\mathbf{u} \neq 0$.

Note that in the case of the bispectrum covariance, the relative importance of $(\Delta)\chi^2_{ND}$ and $(\Delta)\chi^2_{\mathrm{Diag}}$ to $(\Delta)\chi^2$ do not depend on the survey volume used to rescale the covariance. Hence (relative) results obtained for $1\,\mathrm{Gpc}^3$ (which is much smaller than the volume of ongoing surveys) also hold for stage IV dark energy experiments with volumes $\sim 50\,\mathrm{Gpc}^3$. As shown even visually for example in [9], the matrix elements of $\mathbb{D}$ and ϵ and therefore the matrix elements of $\mathbb{D}^{-1}$ and $\mathbb{O}$ both scale with the survey volume coherently, provided everything else including the choice of binning and the effect of the survey window remain unaltered. However, different survey volumes or different survey geometries can enable different binning choices in k-space and thus different correction factors A_P and A_B (using nomenclature from [9]) and different scaling in the diagonal with respect to the off-diagonal elements. The effect of different shot noise levels might also rescale the diagonal and non-diagonal contributions to $(\Delta)\chi^2$ differently. Moreover, the realization of $\mathbf{d}$ (and $\mathbf{u}$) will be more or less noisy depending on the volume of the realization.

The decomposition in terms of a matrix and a perturbation can help us understand also the behaviour of the eigenvalues. Let us consider the sum of matrices $\mathbb{C} = \mathbb{D} + \mathbb{O}$, with their eigenvalues being respectively: $c_1 \geq \ldots \geq c_n$; $\mathbb{D}$: $d_1 \geq \ldots \geq d_n$; $\mathbb{O}$: $o_1 \geq \ldots \geq o_n$.

Weyl's inequality states that: $d_i + o_n \leq c_i \leq d_i + o_1$. This inequality is valid for any sum of symmetric matrices, i.e., holds even if $\mathbb{D}$ is not diagonal. If $\mathbb{O}$ is a perturbation of $\mathbb{D}$ of order ϵ then $|c_i - d_i| \leq \epsilon$. In fact, if we identify $\mathbb{D}$ with the thresholded matrix $\mathbb{T}$ and $\mathbb{O}$ with $\mathbb{C} - \mathbb{T} = \epsilon$, the thresholded matrix $\mathbb{T}$ will have each eigenvalue bounded below (resp. above) by the sum of the eigenvalue of $\mathbb{C}$ and the minimum (resp. maximum) eigenvalue of $\mathbb{O}$. In this case, the matrix $\mathbb{O}$ has approximately random values on the coefficients that are not set to zero, and then the minimum and maximum eigenvalues will be respectively negative and positive. Therefore, the matrix $\mathbb{T} = \mathbb{C} + \mathbb{O}$ can have eigenvalues lower than $\mathbb{C}$.

5.1 Relation to Fisher forecasts

We can now see in a transparent way the effect of the off-diagonal covariance terms in a Fisher-forecast analysis. In this approach to forecasts, the Fisher information matrix $\mathbb{F}$ yields statistical error estimates on the model's parameters $\{\theta_\alpha, \theta_\beta \ldots\}$. The Greek indices hereafter will run over the model's parameters. Given $\mathcal{L} = -\ln L$ where L denotes the likelihood, which under the Gaussian approximation $\mathcal{L} = \chi^2/2$,

$$\mathbb{F}_{\alpha\beta} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle = \sum_{ij} \partial_\alpha t_i \mathbb{C}^{-1}_{ij} \partial_\beta t_j = \sum_i \partial_\alpha t_i \mathbb{D}^{-1}_{ii} \partial_\beta t_i - \sum_{ij} \partial_\alpha t_i \mathbb{O}_{ij} \partial_\beta t_j \equiv \mathbb{F}'_{\alpha\beta} - \mathcal{F}_{\alpha\beta} \quad (5.7)$$

where linear dependence of the theory on the parameters is assumed and we have decomposed the Fisher matrix in a diagonal part $\mathbb{D}$ and a purely off-diagonal contribution $\mathbb{O}$. In the above equation $\mathbb{O}$ can be replaced by $\mathbb{M}$ only when the off-diagonal contributions to $\mathbb{C}$ are small enough. The marginal error on a given parameter is thus $\sigma_\alpha = (\mathbb{F}^{-1})^{1/2}_{\alpha\alpha}$.

$$\sigma^2_\alpha = [(\mathbb{F}' - \mathcal{F})^{-1})]_{\alpha\alpha} = [\mathbb{F}'^{-1}]_{\alpha\alpha} + [\mathbb{F}'^{-1}\mathcal{F}(\mathbb{F}'')^{-1}]_{\alpha\alpha} \equiv \sigma^2_{\mathrm{Diag},\alpha} + \Delta\sigma^2_{ND\alpha} \qquad (5.8)$$

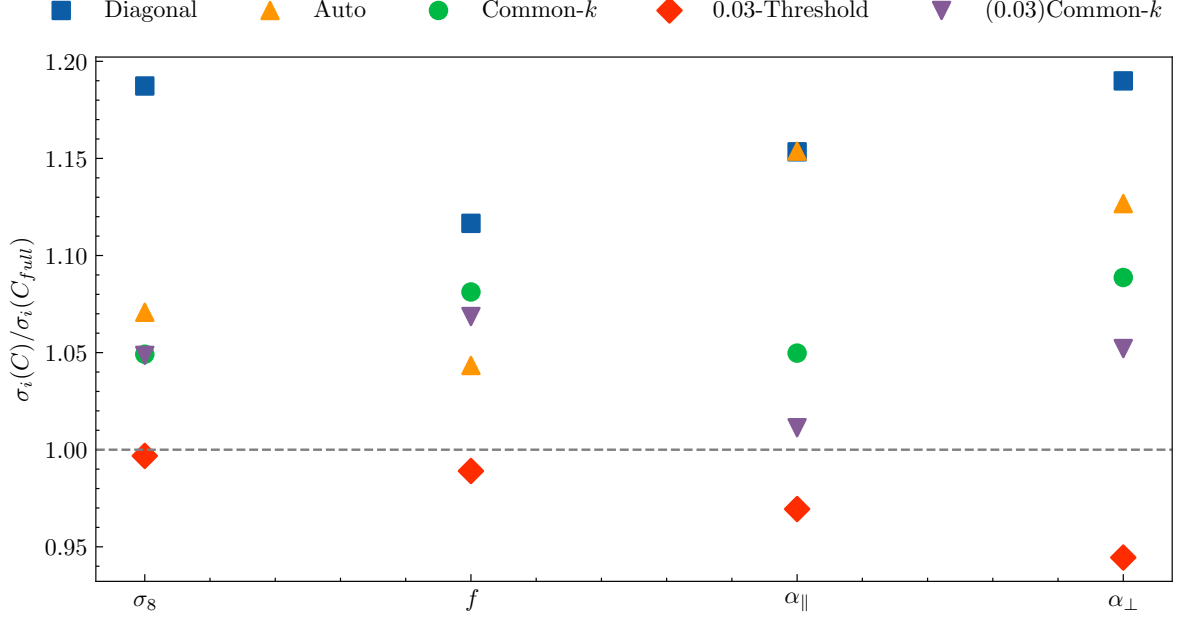**Figure 9.** Fisher forecasts for the parameters of interest $\{\sigma_8, f, \alpha_\parallel, \alpha_\perp\}$ for all the approximations considered in this work. These are shown, for each of the parameters, as the ratio between forecasted errors using the covariance approximation over the errors obtained with the full covariance. All but the "0.03-thresholded" case result in higher values of σ. The diagonal case gives the overall largest errors (as expected). The "Auto" case is shown for reference as it ignores the PB terms but includes all the off-diagonal BB terms and thus does not offer any significant advantages or speed up in its evaluation.

where $\mathbb{F}'' = (\mathbb{F}' - \mathcal{F})$ (a recursive expression) and $\Delta\sigma^2_{ND\alpha}$ is the correction to the diagonal errors induced by the off-diagonal terms. If $\mathbb{F}''$ is approximated by $\mathbb{F}'$ one is assuming that the conditional errors still coincide with the marginalized errors, i.e., $\Delta\sigma^2_{ND} = 0$ (as it happens for a diagonal Fisher matrix). The resulting (Fisher) error ellipses will display approximately correct degeneracies, but the values for the marginalized errors will be underestimated.

In our specific case, we perform Fisher forecasts for the cosmological parameters of interest with the full matrix and various approximations to it. It is then possible to compare σ^2_{Diag} with σ^2 for the parameters of interest, finding that the Fisher forecasts for the diagonal covariance matrix are 10–20% larger than those of the full covariance matrix — as illustrated by figure 9.

We find that, for the analysis set-up featuring the PATCHY mocks, the median element of $\mathcal{F}/\mathbb{F}'$ is 0.43, so the approximation $\mathbb{F}'' \simeq \mathbb{F}'$ will introduce a sizeable error in the forecasted marginalized errors.

In a realistic scenario, the vector **u** is non-zero, and the estimator will only be efficient if the full covariance is used, which provides the optimal weighting of the data-vector. Ignoring the off-diagonal terms effectively weights the data-vector in a sub-optimal way, yielding higher uncertainty on the recovered parameter values. While the case using the full covariance approximates the Cramer-Rao bound, the case using the diagonal covariance approximation loses information, resulting in a less efficient maximum likelihood estimator with (possibly underestimated) bigger error-bars.

6 Discussion and conclusions

The inclusion of the bispectrum multipoles in cosmological parameters estimations from the large scale structure improves significantly cosmological constraints (see e.g. [11] for a quantitative assessment). But this information gain comes at a cost. Besides the additional effort needed to model the bispectrum signal, in the joint likelihood analysis of the redshift space multipoles of the power spectrum and bispectrum, the evaluation of the full covariance matrix can be challenging. The inclusion of the bispectrum multipoles increases dramatically the data-vector size compared to a power spectrum-only analysis. This increases the requirement on the number of mock simulations needed to evaluate reliably the covariance matrix from simulations.

All these challenges could be significantly eased if suitable approximations of the covariance matrix are found.

We have shown that approximating the full covariance matrix blocks that include the bispectrum with a purely diagonal matrix means using an unbiased but sub-optimal estimator. This causes the true errors on the recovered cosmological parameters to be not optimal, and the inferred error-bars to be underestimated, with the constraints not having "coverage". The standard likelihood analysis with the full covariance matrix, if anything, slightly overestimates the errors estimated from the scatter among recovered parameters from many realizations. However, this is likely an apparent error overestimation due to the fact that the scatter among simulations is done with an approximate covariance matrix assuming fully Gaussian statistics and does not fully propagate the uncertainty in the covariance estimation itself [15]; in fact, the magnitude of the apparent overestimation is correlated with the square-root of the Hartlap factor.

We have quantified these effects on a suite of simulated boxes of dark matter and more realistic dark matter tracers, although the effects reported are expected to be even larger in more realistic set-ups such as with survey window and selection functions. The covariance matrix estimated from simulations is *dense*: there are many small off-diagonal terms but none of them is exactly zero. While "thou shall not ignore the bispectrum covariance off-diagonal terms", other approximations may be advantageous. In particular, approximations that make the matrix sparse instead (where many off-diagonal terms are zero instead of being very small but non-zero) can be particularly interesting. The requirements on the number of simulations, for example, may be, potentially, relaxed.

Also, analytic expressions (which can be long and cumbersome to evaluate for all terms) can be evaluated only for the few terms that really matter and more easily calibrated on simulations (following for example the prescription of [9]). The best approximations for the block involving the bispectrum we recommend using are either considering to be non-zero only the terms that have a k mode in common, or applying a suitable thresholding operation to the matrix. The first approach, "Common-k", keeps $\sim 45\%$ of the bispectrum auto-covariance terms, while the second approach, "(τ)Common-k", keeps between 15 and 20% of the bispectrum auto-covariance terms, depending on the k-binning. Such approximations produce results for cosmological inference virtually indistinguishable from, and with the same coverage properties, as those obtained with the full covariance matrix.

We envision that these findings will be useful for the joint redshift space power spectrum and bispectrum analyses from forthcoming surveys, once real-world effects such as window and selection functions and systematics weights are included in the modeling.

Acknowledgments

A MCMC and simulation scatter error for σ_8 and f

Figures 10 and 11 are the counterpart of the figures in section 3 for the remaining parameters of interest, σ_8 and f. A qualitatively similar behaviour as for the parameters $\alpha_\parallel, \alpha_\perp$ is obtained. The only minor differences (appearing in figure 10) are that σ_8 features less correlation between σ_{diag} and σ_{full} and that the diagonal covariance errors on f are closer to the case with the full covariance — in a factor of $\sim 10\%$.

B Visualisation of the limitations of the diagonal approximation

In figure 12 we further quantify the error introduced by assuming the covariance matrix to be diagonally dominant. As said in section 5, in such a case, the inverse covariance matrix could be approximated as in equation (5.2): $\mathcal{C}^{-1} = \mathbb{D}^{-1} - \mathbb{D}^{-1}\epsilon\mathbb{D}^{-1}$. We show that this would introduce unacceptable biases to the estimation of the inverse covariance matrix, which would propagate to the cosmological analysis. Hence, the off-diagonal terms ought not to be ignored.

A complementary way to observe the effect of assuming a diagonal covariance is shown in the right panel of figure 12. There, the residual between the inverse Fisher matrices when assuming a diagonal and full covariance are displayed, which are notated respectively as $\mathbb{F}$ and $\mathbb{F}'$ following the conventions of section 5. As in figure 9, we see how using diagonal covariance results in a forecasted error in the 1D marginalized constraints of $\{\sigma_8, f, \alpha_\parallel, \alpha_\perp\}$ between 10 and 20% higher than using the baseline full covariance. In this figure we can additionally see that the cross-correlations between parameters will be affected to a higher degree in percentage than the auto-correlations.
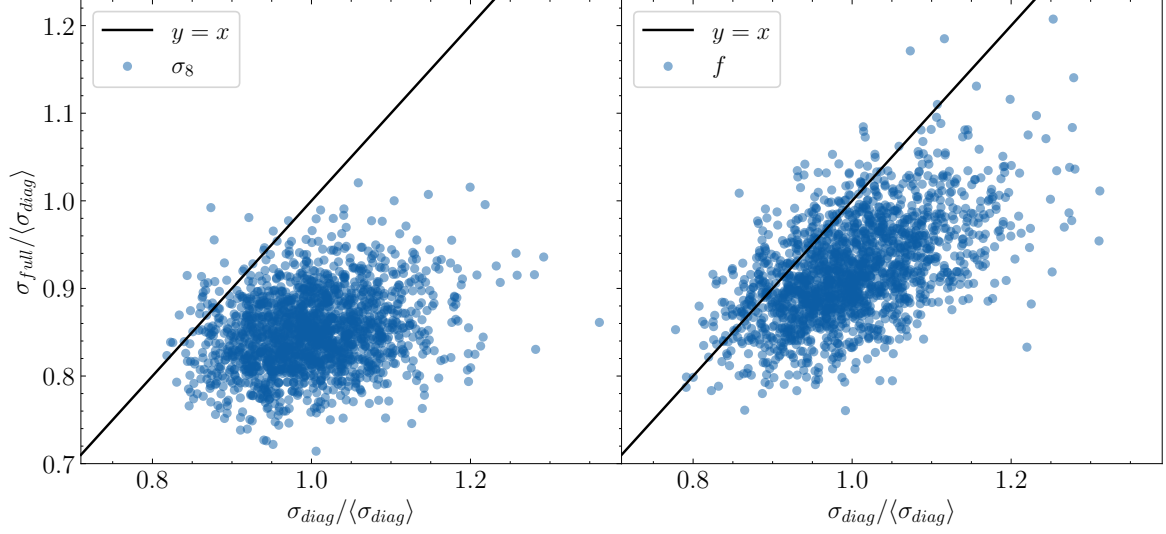
Figure 10. Analogous plot to figure 3, for the case of the parameters σ_8 and f. Similar results are achieved, with the slight variation that the errors on f feature a smaller ratio, $\sigma_{\text{diag}} \sim 1.1\sigma_{\text{full}}$, and that σ_8 has less correlation between the diagonal and full covariance recovered errors.
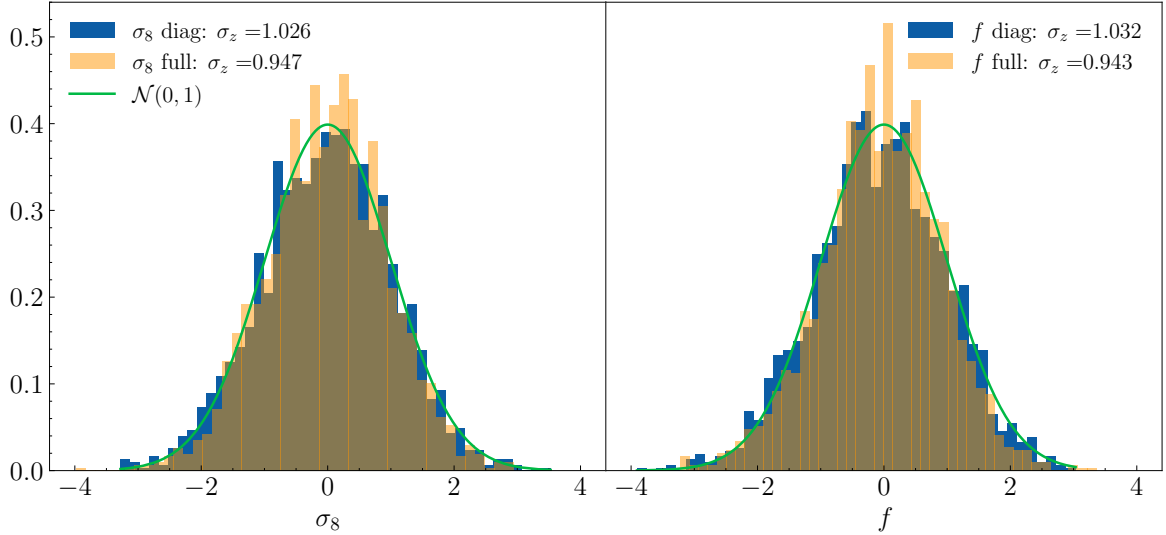

Figure 11. Analogous plot to figure 4, for the parameters σ_8 and f. For these parameters, both the error underestimation (resp. overestimation) for the diagonal (resp. full) covariance is less marked.

C Additional comments on the z statistic

Let us further unpack the z statistic defined in equation (3.1) as

$$z_{\theta_j} \equiv \frac{\theta_j - \mu_{\theta_j}}{\sigma(\theta_j)}. \tag{C.1}$$

The samples of z can be seen as a 2D array with indices representing the mock realization (i) and the parameter (j). For each single parameter j and each of the 2000 mock realizations, we conduct parameter inference for each, deriving a best-fit parameter (θ_j^i) and an *estimate*
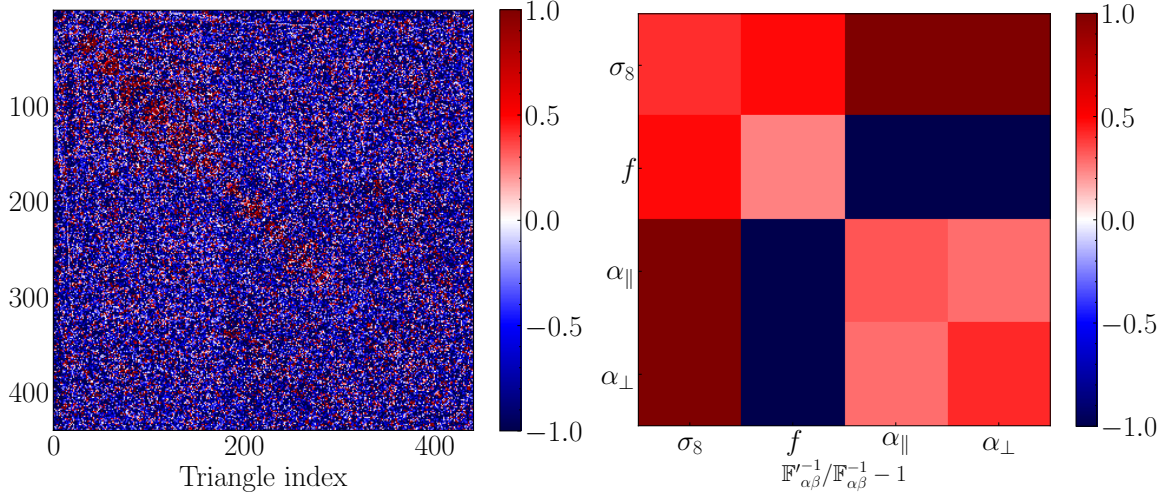
Figure 12. Discrepancy in the obtained inverse covariance matrix by assuming that the covariance matrix is diagonal-dominant, $\mathcal{C}^{-1} = \mathbb{D}^{-1} - \mathbb{D}^{-1}\epsilon\mathbb{D}^{-1}$, and the accurate estimation of C^{-1}. In the left plot we show the residuals $(C^{-1} - \mathcal{C}^{-1})/C^{-1}$, element-wise. The residuals for the inverse Fisher matrix obtained from the diagonal approximation ($\mathbb{F}'$) and from the full covariance matrix ($\mathbb{F}$) are shown in the right plot. In both, it is clearly visible how the diagonal-dominant approximation is invalid for the analysis present in this work, signaling the importance of the off-diagonal terms.

of its error $(\sigma^i(\theta_j))$. It is important to note that using an incorrect covariance in this inference results in an estimator that is unbiased but not optimal, leading to potentially inaccurate error estimates.

The average over 2000 realizations of the best fit parameters has a very little error and can be taken to correspond to the true values μ_j. So, the scatter of the 2000 best fit values θ is an estimate of the error which, for a sufficiently large number of realizations, should be very close to the true error — furthermore, by construction it has good coverage properties. This approach is a standard *rms* estimate and does not involve Hartlap or Sellentin-Heavens corrections as no matrix inversion is needed. For 2000 realizations we expect the error on the error to be quite small and at all effects it can be taken as the true error.

According to the central limit theorem, the distribution of z_{θ_j} should approach a Gaussian form. The width of this distribution will be unity if $\sigma^i(\theta_j)$ accurately estimates the true error on θ_j, i.e. if it has equivalent coverage to the scatter of θ_j^i among the 2000 realizations. However, the extent to which the central limit theorem is applicable remains uncertain, and the assumption of a Gaussian likelihood with fixed covariance for bispectrum data is known to be an approximation and might be slightly sub-optimal. Additionally, the Sellentin-Heavens correction can further deviate an initially Gaussian likelihood from Gaussianity.

Our findings in section 3 indicate that using a diagonal covariance approximation tends to underestimate errors by about 15%. Conversely, employing the full covariance estimated from simulations, while using the Sellentin-Heavens correction and assuming a Gaussian likelihood for the data vector, may slightly overestimate the Gaussianized errors as determined from the scatter under the central limit theorem.

However, some mismatch in these estimations is anticipated for at least two reasons: firstly, the estimated covariance has an associated error, since this estimate might vary if

different sets of mocks are used. This variance is in part why the Sellentin-Heavens correction effectively increases the error estimates. Secondly, the likelihood is not perfectly Gaussian, and matching the *rms* to a 1σ of a Gaussian distribution may result in a slight discrepancy.

References

[1] M. Tegmark, A. Taylor and A. Heavens, *Karhunen-Loeve eigenvalue problems in cosmology: How should we tackle large data sets?*, *Astrophys. J.* **480** (1997) 22 [astro-ph/9603021] [INSPIRE].

[2] L. Verde et al., *The 2dF Galaxy Redshift Survey: The bias of galaxies and the density of the Universe*, *Mon. Not. Roy. Astron. Soc.* **335** (2002) 432 [astro-ph/0112161] [INSPIRE].

[3] R. Scoccimarro, H.A. Feldman, J.N. Fry and J.A. Frieman, *The bispectrum of IRAS redshift catalogs*, *Astrophys. J.* **546** (2001) 652 [astro-ph/0004087] [INSPIRE].

[4] H. Gil-Marín et al., *The power spectrum and bispectrum of SDSS DR11 BOSS galaxies — II. Cosmological interpretation*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 1914 [arXiv:1408.0027] [INSPIRE].

[5] J. Carron, *On the assumption of Gaussianity for cosmological two-point statistics and parameter dependent covariance matrices*, *Astron. Astrophys.* **551** (2013) A88 [arXiv:1204.4724] [INSPIRE].

[6] S. Dodelson and M.D. Schneider, *The Effect of Covariance Estimator Error on Cosmological Parameter Constraints*, *Phys. Rev. D* **88** (2013) 063537 [arXiv:1304.2593] [INSPIRE].

[7] A. Repp, I. Szapudi, J. Carron and M. Wolk, *The Impact of Non-Gaussianity upon Cosmological Forecasts*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 3533 [arXiv:1506.00083] [INSPIRE].

[8] N. Bellomo et al., *Beware of commonly used approximations. Part I. Errors in forecasts*, *JCAP* **10** (2020) 016 [arXiv:2005.10384] [INSPIRE].

[9] D. Gualdi and L. Verde, *Galaxy redshift-space bispectrum: the Importance of Being Anisotropic*, *JCAP* **06** (2020) 041 [arXiv:2003.12075] [INSPIRE].

[10] M. Biagetti, L. Castiblanco, J. Noreña and E. Sefusatti, *The covariance of squeezed bispectrum configurations*, *JCAP* **09** (2022) 009 [arXiv:2111.05887] [INSPIRE].

[11] S. Novell-Masot, D. Gualdi, H. Gil-Marín and L. Verde, *GEO-FPT: a model of the galaxy bispectrum at mildly non-linear scales*, *JCAP* **11** (2023) 044 [arXiv:2303.15510] [INSPIRE].

[12] A. Oddo et al., *Cosmological parameters from the likelihood analysis of the galaxy power spectrum and bispectrum in real space*, *JCAP* **11** (2021) 038 [arXiv:2108.03204] [INSPIRE].

[13] D. Gualdi, S. Novell-Masot, H. Gil-Marín and L. Verde, *Matter trispectrum: theoretical modelling and comparison to N-body simulations*, *JCAP* **01** (2021) 015 [arXiv:2009.02290] [INSPIRE].

[14] D. Gualdi, H. Gil-Marín and L. Verde, *Joint analysis of anisotropic power spectrum, bispectrum and trispectrum: application to N-body simulations*, *JCAP* **07** (2021) 008 [arXiv:2104.03976] [INSPIRE].

[15] E. Sellentin and A.F. Heavens, *Parameter inference with estimated covariance matrices*, *Mon. Not. Roy. Astron. Soc.* **456** (2016) L132 [arXiv:1511.05969] [INSPIRE].

[16] J. Hartlap, P. Simon and P. Schneider, *Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix*, *Astron. Astrophys.* **464** (2007) 399 [astro-ph/0608064] [INSPIRE].

[17] W.J. Percival et al., *The Clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Including covariance matrix errors*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 2531 [arXiv:1312.4841] [INSPIRE].

[18] W.J. Percival, O. Friedrich, E. Sellentin and A. Heavens, *Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix*, *Mon. Not. Roy. Astron. Soc.* **510** (2022) 3207 [arXiv:2108.10402] [INSPIRE].

[19] H. Gil-Marín et al., *The power spectrum and bispectrum of SDSS DR11 BOSS galaxies — I. Bias and gravity*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 539 [arXiv:1407.5668] [INSPIRE].

[20] M.M. Ivanov et al., *Precision analysis of the redshift-space galaxy bispectrum*, *Phys. Rev. D* **105** (2022) 063512 [arXiv:2110.10161] [INSPIRE].

[21] M.M. Ivanov et al., *Cosmology with the galaxy bispectrum multipoles: Optimal estimation and application to BOSS data*, *Phys. Rev. D* **107** (2023) 083515 [arXiv:2302.04414] [INSPIRE].

[22] Z. Slepian and D.J. Eisenstein, *Computing the three-point correlation function of galaxies in $\mathcal{O}(N^2)$ time*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 4142 [arXiv:1506.02040] [INSPIRE].

[23] Z. Slepian et al., *Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies*, *Mon. Not. Roy. Astron. Soc.* **469** (2017) 1738 [arXiv:1607.06097] [INSPIRE].

[24] Z. Slepian et al., *The large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies*, *Mon. Not. Roy. Astron. Soc.* **468** (2017) 1070 [arXiv:1512.02231] [INSPIRE].

[25] X. Xu et al., *A 2% Distance to $z = 0.35$ by Reconstructing Baryon Acoustic Oscillations — II: Fitting Techniques*, *Mon. Not. Roy. Astron. Soc.* **427** (2012) 2146 [arXiv:1202.0091] [INSPIRE].

[26] DESI collaboration, *The DESI Experiment Part I: Science,Targeting, and Survey Design*, arXiv:1611.00036 [INSPIRE].

[27] EUCLID collaboration, *Euclid Definition Study Report*, arXiv:1110.3193 [INSPIRE].

[28] LSST collaboration, *LSST: from Science Drivers to Reference Design and Anticipated Data Products*, *Astrophys. J.* **873** (2019) 111 [arXiv:0805.2366] [INSPIRE].

[29] N.S. Sugiyama, S. Saito, F. Beutler and H.-J. Seo, *Perturbation theory approach to predict the covariance matrices of the galaxy power spectrum and bispectrum in redshift space*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) 1684 [arXiv:1908.06234] [INSPIRE].

[30] N.S. Sugiyama, S. Saito, F. Beutler and H.-J. Seo, *A complete FFT-based decomposition formalism for the redshift-space bispectrum*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 364 [arXiv:1803.02132] [INSPIRE].

[31] T. Flöss, M. Biagetti and P.D. Meerburg, *Primordial non-Gaussianity and non-Gaussian covariance*, *Phys. Rev. D* **107** (2023) 023528 [arXiv:2206.10458] [INSPIRE].

[32] F. Villaescusa-Navarro et al., *The Quijote simulations*, *Astrophys. J. Suppl.* **250** (2020) 2 [arXiv:1909.05273] [INSPIRE].

[33] F.-S. Kitaura et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release*, *Mon. Not. Roy. Astron. Soc.* **456** (2016) 4156 [arXiv:1509.06400] [INSPIRE].

[34] R. Scoccimarro, H.M.P. Couchman and J.A. Frieman, *The bispectrum as a Signature of Gravitational Instability in Redshift-Space*, *Astrophys. J.* **517** (1999) 531 [astro-ph/9808305] [INSPIRE].

[35] G. D'Amico et al., *The BOSS bispectrum analysis at one loop from the Effective Field Theory of Large-Scale Structure*, *JCAP* **05** (2024) 059 [arXiv:2206.08327] [INSPIRE].

[36] M. Crocce and R. Scoccimarro, *Renormalized cosmological perturbation theory*, *Phys. Rev. D* **73** (2006) 063519 [astro-ph/0509418] [INSPIRE].

[37] H. Gil-Marín et al., *Perturbation theory approach for the power spectrum: from dark matter in real space to haloes in redshift space*, *JCAP* **11** (2012) 029 [arXiv:1209.3771] [INSPIRE].

[38] C. Alcock and B. Paczynski, *An evolution free test for non-zero cosmological constant*, *Nature* **281** (1979) 358 [INSPIRE].

[39] W.E. Ballinger, J.A. Peacock and A.F. Heavens, *Measuring the cosmological constant with redshift surveys*, *Mon. Not. Roy. Astron. Soc.* **282** (1996) 877 [astro-ph/9605017] [INSPIRE].

[40] BOSS collaboration, *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles*, *Mon. Not. Roy. Astron. Soc.* **443** (2014) 1065 [arXiv:1312.4611] [INSPIRE].

[41] H. Gil-Marín et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies*, *Mon. Not. Roy. Astron. Soc.* **465** (2017) 1757 [arXiv:1606.00439] [INSPIRE].

[42] T. Baldauf, U. Seljak, V. Desjacques and P. McDonald, *Evidence for Quadratic Tidal Tensor Bias from the Halo Bispectrum*, *Phys. Rev. D* **86** (2012) 083540 [arXiv:1201.4827] [INSPIRE].

[43] S. Saito et al., *Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum*, *Phys. Rev. D* **90** (2014) 123522 [arXiv:1405.1447] [INSPIRE].

[44] S. Brieden, H. Gil-Marín and L. Verde, *PT challenge: validation of ShapeFit on large-volume, high-resolution mocks*, *JCAP* **06** (2022) 005 [arXiv:2201.08400] [INSPIRE].

[45] J.C. Jackson, *Fingers of God: A critique of Rees' theory of primoridal gravitational radiation*, *Mon. Not. Roy. Astron. Soc.* **156** (1972) 1P [arXiv:0810.3908] [INSPIRE].

[46] E. Sellentin and J.-L. Starck, *Debiasing inference with approximate covariance matrices and other unidentified biases*, *JCAP* **08** (2019) 021 [arXiv:1902.00709] [INSPIRE].

[47] H. Gil-Marín, *How to optimally combine pre-reconstruction full shape and post-reconstruction BAO signals*, *JCAP* **05** (2022) 040 [arXiv:2203.05581] [INSPIRE].

[48] D. Kodwani, D. Alonso and P. Ferreira, *The effect on cosmological parameter estimation of a parameter dependent covariance matrix*, *Open J. Astrophys.* **2** (2019) 3 [arXiv:1811.11584] [INSPIRE].

[49] P.J. Bickel and E. Levina, *Covariance regularization by thresholding*, arXiv:0901.3079 [DOI:10.1214/08-AOS600].

[50] T. Cai and W. Liu, *Adaptive Thresholding for Sparse Covariance Matrix Estimation*, arXiv:1102.2237.

[51] T.T. Cai and H.H. Zhou, *Optimal rates of convergence for sparse covariance matrix estimation*, arXiv:1302.3030 [DOI:10.1214/12-AOS998].

[52] J. Friedman, T. Hastie and R. Tibshirani, *Sparse inverse covariance estimation with the lasso*, *Biostatistics* **9** (2008) 432 [arXiv:0708.3517].

[53] F. Oztoprak, J. Nocedal, S. Rennie and P.A. Olsen, *Newton-like methods for sparse inverse covariance estimation*, in *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc. (2012) [PDF].

[54] N. Padmanabhan, M. White, H.H. Zhou and R. O'Connell, *Estimating sparse precision matrices*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1567 [`arXiv:1512.01241`] [ɪɴSPIRE].

[55] A. Heavens, E. Sellentin, D. de Mijolla and A. Vianello, *Massive data compression for parameter-dependent covariance matrices*, *Mon. Not. Roy. Astron. Soc.* **472** (2017) 4244 [`arXiv:1707.06529`] [ɪɴSPIRE].

[56] D. Gualdi et al., *Geometrical compression: a new method to enhance the BOSS galaxy bispectrum monopole constraints*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) L29 [`arXiv:1901.00987`] [ɪɴSPIRE].

[57] D. Gualdi et al., *GEOMAX: beyond linear compression for three-point galaxy clustering statistics*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) 776 [`arXiv:1912.01011`] [ɪɴSPIRE].

[58] C. Stein, B. Efron and C. Morris, *Improving the usual estimator of a normal covariance matrix*, Department of Statistics, Stanford University (1972).

[59] A. Taylor, B. Joachimi and T. Kitching, *Putting the Precision in Precision Cosmology: How accurate should your data covariance matrix be?*, *Mon. Not. Roy. Astron. Soc.* **432** (2013) 1928 [`arXiv:1212.4359`] [ɪɴSPIRE].

[60] B. Joachimi, *Non-linear shrinkage estimation of large-scale structure covariance*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) L83 [`arXiv:1612.00752`] [ɪɴSPIRE].

[61] B. Efron, *The Jackknife, the Bootstrap and other resampling plans*, SIAM (1982).

[62] M.D. Schneider, S. Cole, C.S. Frenk and I. Szapudi, *Fast generation of ensembles of cosmological N-body simulations via mode-resampling*, *Astrophys. J.* **737** (2011) 11 [`arXiv:1103.2767`] [ɪɴSPIRE].

[63] J. Hou, Z. Slepian and R.N. Cahn, *Measurement of parity-odd modes in the large-scale 4-point correlation function of Sloan Digital Sky Survey Baryon Oscillation Spectroscopic Survey twelfth data release CMASS and LOWZ galaxies*, *Mon. Not. Roy. Astron. Soc.* **522** (2023) 5701 [`arXiv:2206.03625`] [ɪɴSPIRE].

[64] C.E. Shannon, *A mathematical theory of communication*, *Bell Syst. Tech. J.* **27** (1948) 379.

[65] S. Kullback and R.A. Leibler, *On Information and Sufficiency*, *Annals Math. Statist.* **22** (1951) 79 [ɪɴSPIRE].

[66] M. Tumminello, F. Lillo and R.N. Mantegna, *Kullback-Leibler distance as a measure of the information filtered from multivariate data*, *Phys. Rev. E* **76** (2007) 031123 [`arXiv:0706.0168`].

[67] A. Fumagalli et al., *Fitting covariance matrix models to simulations*, *JCAP* **12** (2022) 022 [`arXiv:2206.05191`] [ɪɴSPIRE].

[68] J. Sherman and W.J. Morrison, *Adjustment of an inverse matrix corresponding to a change in one element of a given matrix*, *Annals Math. Statist.* **21** (1950) 124.

[69] M.A. Woodbury, *Inverting modified matrices*, Department of Statistics, Princeton University (1950).