Fewer mocks and less noise: Reducing the dimensionality of cosmological observables with subspace projections

Oliver H. E. Philcox,[1,2,3,*] Mikhail M. Ivanov,[4,5] Matias Zaldarriaga,[3] Marko Simonović,[6] and Marcel Schmittfull[3]

[1]*Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08540, USA*
[2]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom*
[3]*School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, New Jersey 08540, USA*
[4]*Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, New York 10003, USA*
[5]*Institute for Nuclear Research of the Russian Academy of Sciences, 60th October Anniversary Prospect, 7a, 117312 Moscow, Russia*
[6]*Theoretical Physics Department, CERN, 1 Esplanade des Particules, Geneva 23, CH-1211 Switzerland*

(Received 11 September 2020; accepted 14 January 2021; published 4 February 2021)

Creating accurate and low-noise covariance matrices represents a formidable challenge in modern-day cosmology. We present a formalism to compress arbitrary observables into a small number of bins by projection into a model-specific subspace that minimizes the prior-averaged log-likelihood error. The lower dimensionality leads to a dramatic reduction in covariance matrix noise, significantly reducing the number of mocks that need to be computed. Given a theory model, a set of priors, and a simple model of the covariance, our method works by using singular value decompositions to construct a basis for the observable that is close to Euclidean; by restricting to the first few basis vectors, we can capture almost all the constraining power in a lower-dimensional subspace. Unlike conventional approaches, the method can be tailored for specific analyses and captures nonlinearities that are not present in the Fisher matrix, ensuring that the full likelihood can be reproduced. The procedure is validated with full-shape analyses of power spectra from Baryon Oscillation Spectroscopic Survey (BOSS) DR12 mock catalogs, showing that the 96-bin power spectra can be replaced by 12 subspace coefficients without biasing the output cosmology; this allows for accurate parameter inference using only ∼100 mocks. Such decompositions facilitate accurate testing of power spectrum covariances; for the largest BOSS data chunk, we find the following: (a) analytic covariances provide accurate models (with or without trispectrum terms); and (b) using the sample covariance from the MultiDark-Patchy mocks incurs a $\sim 0.5\sigma$ shift in Ω_m, unless the subspace projection is applied. The method is easily extended to higher order statistics; the ∼2000-bin bispectrum can be compressed into only ∼10 coefficients, allowing for accurate analyses using few mocks and without having to increase the bin sizes.

DOI: 10.1103/PhysRevD.103.043508

I. INTRODUCTION

Most conventional analyses of cosmological data proceed by measuring a summary statistic, computing a theory model, and comparing the two in a Gaussian likelihood. A key ingredient in this is the inverted covariance matrix (also known as the precision matrix). In the simplest case, this is measured by creating a number of mock datasets, computing the desired statistic in each, and then estimating the sample covariance directly. In order to obtain an unbiased precision matrix estimate, a large number of mocks is required [1], and it has been further shown that noise in the precision matrix leads to stochastic shifts in the best-fit parameters, which is usually treated by inflating the output parameter covariances [2–6]. To reduce these shifts, it is important to use a large number of mocks, though creating such a sample requires considerable computational effort, since the mocks are also required to be accurate. Furthermore, the magnitude of the effect increases with dimensionality; upcoming galaxy surveys such as those of DESI [7], Euclid [8], the Rubin Observatory [9], and the Roman Telescope [10] will provide substantially higher resolution data, with an associated increase in the number of bins.

*ohep2@cantab.ac.uk

There exists significant literature pertaining to alternative methods for galaxy survey covariance matrix generation. On one end of the scale sits purely theoretical models of the covariance, which, for the galaxy power spectrum and its multipoles, is possible via perturbation theory (PT) [11–14]. While these are theoretically well motivated, we are fundamentally limited by the applicability of the perturbative model in the nonlinear regime, and, without simulations, it is unclear how to set model hyperparameters such as galaxy bias and PT counterterms. A natural extension of this therefore is semianalytic models, which combine well-understood theory and free parameters that can be calibrated from a (small) number of simulations. These exist for a range of statistics, including two-point correlation functions [12,15–18], three-point correlation functions [19], power spectra [20,21], and bispectra [22], though the applicability of the model assumptions must be rigorously tested in all circumstances.

An alternative approach is to find different ways of estimating the covariance matrix from simulations, for example, by using noise reduction techniques (such as tapering [23], shrinkage [24], invoking sparsity [25], and combining with theoretical covariances [26]), calibrating the covariance matrix with inexpensive small-volume simulations [27,28], or expanding the precision matrix around some smooth fiducial model [29]. Of particular interest are schemes that reduce the dimensionality of the data vector via some form of compression. As the number of bins falls, precision matrix noise becomes less important, allowing one to use fewer mocks to generate the covariance. The question, however, is how to perform this data compression.

In this work, we develop a method to compress a given data vector by computing an information maximizing subspace from the theoretical model for our data, using techniques developed for gravitational wave analyses [30]. In practice, we take a large sample of $N_{\rm bank}$ points in the underlying (prior-bounded) space of $N_{\rm param}$ cosmological and nuisance parameters, and compute model data vectors (which we denote the "template bank") at each point. This requires no knowledge of the observational data vector, just the theory model and parameter priors. Using a singular value decomposition (SVD) of the resulting $N_{\rm bank} \times N_{\rm bin}$ noise-weighted matrix, we generate a set of basis vectors which are ordered in signal-to-noise and, together, contain all cosmological information. These do not require prior knowledge of the observed data vector. By restricting to the first $N_{\rm SV}$ of these modes, we capture the dominant contributors to the specific experimental likelihood, in a much lower dimensional space. The analysis then proceeds using the projection of the full data vector into this subspace; we note that it is fully applicable to non-Gaussian posteriors (including multimodality) and blind to the choice of input parameters. Assuming one has a somewhat accurate initial ansatz for the covariance, this procedure captures the maximum log-likelihood information possible for a fixed dimensional data vector using linear transformations. Furthermore, it is of low cost, requiring only a set of theory model samples that could easily be evaluated at the start of a Markov chain Monte Carlo (MCMC) routine.

Ours is certainly not the first work performing analysis along these lines, with other examples including bispectrum compression via approximate eigenvalue decompositions [31] and expansions in polynomial basis functions [32–35], as well as the `COSEBI` method for cosmic shear observables [36,37]. Perhaps most notable are the `MOPED` algorithm [38–41] and Karhunen-Loève methods [42], both of which perform linear transformations to compress the data, with the former outputting a set of $N_{\rm param}$ values that preserve the Fisher matrix. While these have been shown to have utility in cosmological analyses [43,44] most define the basis vectors by assuming the posterior surface to be locally Gaussian, i.e., that all information is contained within the second derivatives of the log-likelihood, and hence the Fisher matrix. Further, the compression can be computationally expensive for high-dimensional datasets and is optimal only for the linear response of the model data vector with respect to parameter changes around a fiducial point in model space, without consideration of whether this is valid across the parameter space spanned by the priors. For most cosmological analyses, these issues do not pose a significant problem, especially if the procedure is iterated, but caution is needed in the case of highly nonlinear posteriors. References [45,46] provide an interesting generalization of these approaches to non-Gaussian likelihoods, via compression to a *score function*. The subspace method developed herein instead optimizes the log-likelihood itself, ensuring that the χ^2 factors of the original and compressed likelihoods agree up to some fixed limit across the whole prior domain. Unlike approaches simply decomposing the covariance matrix of the statistic, our approach is tailored to each specific analysis, effectively ensuring that we do not encode information that is well known *a priori*. Notably, this can require using a few *more* modes than parameters to encapsulate nonlinear parameter responses, though the precise number can be robustly set from accuracy considerations. Furthermore, while many methods use an approximate covariance of the compressed data points (e.g., by assuming them to be independent) allowing for analytic posterior sampling, we instead opt to use the full covariance for accuracy, as in Refs. [40,44].

While our approach is fully general and can be applied to any observable, given a theory model, parameter priors, and a fiducial estimate of the covariance, we here consider the analysis of galaxy survey data in redshift space, using the approach of Refs. [47–52]. By creating a template bank from the theory model, we can generate a subspace of a much reduced dimension, which, for a large number of

mocks, gives similarly accurate parameter inference to the full likelihood, and, for a few mocks, substantially reduces the noise-induced shifts, avoiding the need for significant posterior inflation.

The paper has the following structure. In Sec. II, we provide a mathematical derivation of the template bank and subspace decomposition, as well as discussing the corresponding subspace likelihoods. Section III discusses the expected constraints obtained from the subspace analysis and the effects of noise in the data and covariance. We apply the formalism to the galaxy power spectrum in Sec. IV, before discussing extensions to other analyses in Sec. V and concluding with a summary in Sec. VI. Appendixes A and B contain derivations of key results used in Sec. III, with Appendix C presenting additional material related to Sec. IV.

II. METHODOLOGY

A. The template bank and Euclidean subspace

A general cosmological model is specified by a set of parameters θ, which can be fundamental, nuisance, or systematic. While the arguments below strictly apply to any physical model, this work will principally be concerned with the analysis of galaxy power spectra, via the one-loop effective field theory (EFT) model [47–52]. This carries the parameter vector

$$\theta = \{\omega_{\rm cdm}, A_s/A_{s,\rm fid}, h, \ldots\} \times \{b_1, b_2, b_{G_2}, b_4, c_{s,0}, c_{s,2}, P_{\rm shot}\}, \tag{1}$$

where additional parameters can be added to the first set to more finely probe cosmology. Here, we will use only these three for simplicity, but we note that the method can be arbitrarily extended.

The model parameters generate a manifold with which we can associate a vector field of $N_{\rm bin}$-dimensional model spectra, $P_a(\theta)$, where a specifies the component (k-bin and multipole) of the power spectrum evaluated at θ. We may define the inner product between two points $\theta^{(i)}$ and $\theta^{(j)}$ as the noise weighted distance from a suitably defined mean spectrum, $\bar{P}$ (to later be set to the mean power spectrum from the template bank):

$$\langle \delta P^{(i)}, \delta P^{(j)} \rangle_{\mathsf{C}} = \sum_{ab} \delta P_a^{(i)} \mathsf{C}^{-1}_{ab} \delta P_b^{(j)}, \tag{2}$$

where C is some fiducial covariance, encoding the metric, and $\delta P^{(i)} = P(\theta^{(i)}) - \bar{P}$.[1] This is motivated by considerations of the χ^2 of a sample power spectrum $\hat{P}$ with model $P(\theta)$ and covariance C_D, which is simply a squared norm,

$$\hat{\chi}^2(\theta) = \langle \hat{P} - P(\theta), \quad \hat{P} - P(\theta) \rangle_{\mathsf{C}_D}. \tag{3}$$

It is convenient to perform a global linear transform that centers and whitens the power spectra,

$$\delta P(\theta) \to X(\theta), \qquad X_a(\theta) \equiv \sum_{ab} \mathsf{C}_b^{-1/2} [P_b(\theta) - \bar{P}_b] \tag{4}$$

(where $C^{1/2}$ is the Cholesky factorization of C), while preserving the inner product,

$$\begin{aligned} \langle \delta P^{(i)}, \delta P^{(j)} \rangle_{\mathsf{C}} &\equiv \langle X^{(i)} | X^{(j)} \rangle \\ &= \sum_{abcd} X_a^{(i)} \mathsf{C}^{1/2}_{ab} \mathsf{C}^{-1}_{bc} \mathsf{C}^{1/2}_{cd} X_d^{(j)} \\ &= \sum_a X_a^{(i)} X_a^{(j)}. \end{aligned} \tag{5}$$

Note that the metric is Euclidean in this set of coordinates.[2] We can write an arbitrary rotated spectrum $X(\theta)$ in terms of a set of $N_{\rm bin}$ orthonormal basis vectors V_α, i.e.,

$$X_a(\theta) = \sum_\alpha c_\alpha(\theta) V_{\alpha a}, \qquad c_\alpha(\theta) = \sum_a X_a(\theta) V_{\alpha a} \tag{6}$$

with coefficients c_α, where the basis vectors V_α satisfy $\langle V_\alpha | V_\beta \rangle = \delta^K_{\alpha\beta}$. Hence the inner product of any two (rotated) vectors can be written as a product of basis coefficients,

$$\langle X^{(i)} | X^{(j)} \rangle = \sum_\alpha c_\alpha^{(i)} c_\alpha^{(j)}. \tag{7}$$

To compute the basis vectors, V_α, of Eq. (6), we sample a set of $N_{\rm bank}$ spectra from the parameter manifold (known as the "template bank") and execute a singular value decomposition.[3] Practically, this identifies the directions in the vector space which have the largest contributions to the log-likelihood, χ^2. Given a set of points, $\{\theta^{(i)}\}$, on the (bounded) manifold with corresponding spectra $\{P^{(i)}\}$ (each of dimension $N_{\rm bin}$), we first compute the rotated spectra $\{X^{(i)}\}$ at each point using Eq. (4), setting $\bar{P}$ to the mean of all template bank spectra.[4] By stacking the rotated spectra, we can form an $N_{\rm bank} \times N_{\rm bin}$ matrix X_{ia}, which has the SVD,

[1] Note that this assumes the manifold to be Riemannian, such that the "distance" between points is specified by a quadratic form. Below, this will correspond to the assumption of a Gaussian noise model.

[2] When computing the data χ^2 we will later find that the metric becomes *almost* Euclidean, providing the true and fiducial covariances are similar.

[3] Note that we restrict the parameter ranges (i.e., apply broad priors), such that the manifold is of finite extent.

[4] Subtracting the mean is necessary to perform SVDs, but it does not change the analysis, since $\bar{P}$ cancels in the likelihood.

$$X_{ia} = \sum_{\alpha} U_{i\alpha} D_\alpha V_{\alpha a}, \tag{8}$$

where D_α is a rank-$N_{\rm bin}$ diagonal matrix of the singular values (SVs) and the matrices U and V (of dimension $N_{\rm bank} \times N_{\rm bin}$ and $N_{\rm bin} \times N_{\rm bin}$) are unitary, projecting from observations to SVs, and SVs to spectra, respectively.[5] By comparison with Eq. (6), we see that the basis coefficients are given by

$$c_\alpha^{(i)} = U_\alpha^{(i)} D_\alpha. \tag{9}$$

The principal value of this decomposition is that any spectrum can be well approximated by a relatively small number of basis vectors, such that

$$X_a^{(i)} \approx \sum_{\alpha=1}^{N_{\rm SV}} c_\alpha^{(i)} V_{\alpha a} \tag{10}$$

for $N_{\rm SV} < N_{\rm bin}$, where we assume that the SVs are arranged in decreasing order.[6] This is possible since $|c_\alpha| \leq D_\alpha$ as U is unitary, and D_α is small for large α. We henceforth drop the components of U, V, and D with indices above $N_{\rm SV}$. The projection corresponds to projecting the spectra into an $N_{\rm SV}$-dimensional Euclidean subspace, as we approximate

$$\langle X^{(i)} | X^{(j)} \rangle \approx \sum_{\alpha=1}^{N_{\rm SV}} c_\alpha^{(i)} c_\alpha^{(j)}. \tag{11}$$

Here, the first mode ($\alpha = 1$) defines the basis vectors that contribute most to the distance between two points on the manifold (or equivalently, the χ^2 difference), the second mode is the transformation that has the second largest contribution, etc.

We may fix $N_{\rm SV}$ by again considering the inner product. The mean squared distance of all $N_{\rm bank}$ spectra from the mean spectrum $\bar{P}$ is given by the average χ^2,

$$\begin{aligned}\overline{\chi^2} &= \frac{1}{N_{\rm bank}} \sum_i \langle \delta P^{(i)}, \delta P^{(i)} \rangle_{\mathsf{C}} = \frac{1}{N_{\rm bank}} \sum_{ia} X_a^{(i)} X_a^{(i)} \\ &= \frac{1}{N_{\rm bank}} \sum_{ia\alpha\beta} U_{i\alpha} D_\alpha V_{\alpha a} U_{i\beta} D_\beta V_{\beta a} = \frac{1}{N_{\rm bank}} \sum_{i\alpha} U_{i\alpha}^2 D_\alpha^2 \\ &= \frac{1}{N_{\rm bank}} \sum_\alpha D_\alpha^2 \end{aligned} \tag{12}$$

since V and U are unitary. If one uniformly increases the number of template spectra across the manifold-with-boundary by a factor of f, the mean χ^2 (which is a geometric property of the manifold, unrelated to the template bank in the large $N_{\rm bank}$ limit) should be invariant; thus $D_\alpha \propto \sqrt{f}$. We thus conclude that $D_\alpha \propto \sqrt{N_{\rm bank}}$. Assuming $\overline{\chi^2}$ to be a good estimator of the distance between any two spectra in the template bank (and hence any two theory models on the prior-bounded manifold), the value of $N_{\rm SV}$ may be set by excluding those SVs which, in total, contribute less than some fixed $\chi^2_{\rm min}$ to $\overline{\chi^2}$; this fixes the optimal number of SVs, $N^*_{\rm SV}$, to

$$N^*_{\rm SV} = \min \left\{ N_{\rm SV} \left| \sum_{\alpha=N_{\rm SV}+1}^{N_{\rm bin}} D_\alpha^2 \leq \chi^2_{\rm min} N_{\rm bank} \right. \right\}. \tag{13}$$

This condition ensures that the decomposition incurs a χ^2 error below $\chi^2_{\rm min}$, averaged across the prior space. For a perfect measurement, with nondegenerate parameters, we would additionally require $N^*_{\rm SV} \geq N_{\rm param}$, where $N_{\rm param}$ is the dimensionality of the manifold. This would ensure that features of the spectra are not lost under the subspace projection, though, in practice, parameter degeneracies limit this.

A few comments on the SVD are needed. First, we note that the decomposition of Eq. (8) implies

$$X^T X = V^T D^2 V \tag{14}$$

(keeping coordinates implicit for clarity), which is just a diagonalization with eigenvalue matrix D^2. From this, it is clear that the SVD performs an eigendecomposition [or equivalently, a principal component analysis (PCA)] of $X^T X$, the covariance matrix of (whitened) model spectra $X(\theta)$ across the prior manifold.[7] The first basis vector in V (with the largest singular value) therefore identifies the mode of the power spectrum that sources most of the variance between all templates sampled from the model space; the second vector identifies the second largest mode, subject to being orthogonal to the first one, and so on. Keeping only the basis vectors corresponding to the largest singular values [i.e., Eq. (11)] thus encapsulates the main variation in $P(k)$ seen across the sampled model space. Given that we weight by $\mathsf{C}^{-1/2}$, this is equivalent to finding the modes that contribute most to the log-likelihood, χ^2.

Importantly, this approach is different from a standard eigendecomposition or PCA of the *measurement* covariance matrix (e.g., [31,55]), which identifies the orthogonal modes of the statistic that explain most of the variance of the measurement when sampling over different random realizations of the measurement rather than different

[5]Since the SVD algorithm is linear in the number of posterior samples, generating basis vectors is not a computationally intensive step.

[6]By the Eckart-Young theorem, this is the *best* rank-$N_{\rm SV}$ matrix approximation to X [53].

[7]This decomposition is not a new concept; such an approach forms the basis of many dimensionality reduction algorithms used in machine learning. For an additional application of this to approximating cosmological theory models, see Ref. [54].

samples from the model space. Unlike our approach, this PCA is agnostic of the model space, and thus cannot exploit the underlying structure, e.g., that cosmological parameters typically change multiple k-bins simultaneously in a smooth manner (implying that neighboring bins are correlated with respect to parameters, even on very large scales). While the approach of covariance PCAs is to discard *any* information corresponding to modes that are too noisy to be measured well, the effect of using SVDs is instead to discard any information that does not affect *our* analysis, made possible by sampling only a finite region of theory space. While this may lead to a loss of information regarding unsampled parameters, it ensures that the basis vectors do not include much information about parameters for which we have strong prior knowledge. Of course, our compression is optimal only for analyses that utilize the sampled parameters (or some transformations thereof); in practice, we expect the decomposition to still be useful in a larger parameter space, though the optimal analysis would require creation of a new template bank including the additional degrees of freedom.

An additional point of note concerns the choice of parameters, $\{\theta^{(i)}\}$, with which we generate the template bank. A simple choice would be to draw samples uniformly from each cosmological parameter (subject to some limits), though there is freedom in their exact form, for example, whether to use Ω_m or ω_{cdm}, A_s or $\log A_s$. In the general case, this will have a nontrivial impact on the basis vectors, since it will give different weights to different sections of the manifold from which X is sampled. However, we note that the SVD does not receive any information on the input parameters, only the output spectra, and thus, if the locations on the manifold are held fixed, the decomposition is invariant of the choice of coordinate chart. Furthermore, we expect identical results (on average) when drawing samples from any linear transform of the input parameters (assuming that the manifold boundary remains unchanged). In general, however, we suggest using the same choices of parameters and priors (or at least a superset of these) to draw $\{\theta^{(i)}\}$ as will be used in the later MCMC analysis, such that the basis vectors well represent the case in hand. This blindness to the choice of parameters highlights another important note; the method is fully applicable to models which are nonlinear in parameters, and thus give non-Gaussian posteriors. As an example, consider a parameter, ζ, which enters the model quadratically. Based on the above discussion, if the sampling points on the prior manifold are identical, we can parametrize by either ζ^2 or ζ and obtain the same decomposition. In the latter case, the posterior is highly non-Gaussian and indeed bimodal due to the $\zeta \leftrightarrow -\zeta$ symmetry.

While our decomposition is blind to the input parameters, there exist alternative approaches that include such information, for example, partial least squares and canonical correlation analyses [56]. In these formalisms, both the input parameters and model data vector are projected into a latent space; the benefit of this is that it gives the linear combinations of parameters that define the best-constrained features in the data vector. While this sounds appealing, in our context its interpretation is difficult since the optimal parameter sets are nontrivial combinations of cosmological and nuisance parameters, the latter of which are less meaningful.

B. Application to observational likelihoods

While the above is somewhat abstract, it is of use when we consider observational data, since it gives us a rigorous method for which to reduce the dimensionality of the spectra. For a model power spectrum $\hat{P}$, the log-likelihood of parameters θ given data covariance C_D is simply

$$-2\log\mathcal{L}(\theta) = \hat{\chi}^2(\theta) = \sum_{ab}(\hat{P}_a - P_a(\theta))\mathsf{C}^{-1}_{D,ab}(\hat{P}_b - P_b(\theta)), \quad (15)$$

where we have assumed a Gaussian noise model, as is common in cosmology, and ignored the effects of parameter priors. Under the previous rotation by fiducial covariance C [Eq. (4)], this can be written

$$\hat{\chi}^2(\theta) = \sum_{\alpha\beta}(\hat{X}_\alpha - X_\alpha(\theta))[\mathsf{C}^{T/2}\mathsf{C}_D^{-1}\mathsf{C}^{1/2}]_{\alpha\beta}(\hat{X}_\beta - X_\beta(\theta)). \quad (16)$$

Note this is diagonal *only* if the fiducial covariance matches that of the data. When performing inference from a small number of mocks, the data covariance is not well known; thus it is preferred to use a smooth model for C instead, giving a slightly non-Euclidean space. In terms of the basis vectors of Eq. (6), and using modes only up to $N_{\rm SV}$, we obtain

$$\begin{aligned}\hat{\chi}^2(\theta) &\approx \sum_{\alpha=1}^{N_{\rm SV}}\sum_{\beta=1}^{N_{\rm SV}}(\hat{c}_\alpha - c_\alpha(\theta))[V\mathsf{C}^{T/2}\mathsf{C}_D^{-1}\mathsf{C}^{1/2}V^T]_{\alpha\beta} \\ &\quad\times(\hat{c}_\beta - c_\beta(\theta)) \\ &= \sum_{\alpha=1}^{N_{\rm SV}}\sum_{\beta=1}^{N_{\rm SV}}(\hat{c}_\alpha - c_\alpha(\theta))\mathcal{C}^{-1}_{D,\alpha\beta}(\hat{c}_\beta - c_\beta(\theta)). \end{aligned} \quad (17)$$

In the second line, we have noted this likelihood is simply a Gaussian likelihood for the $\hat{c}$ variables, given covariance $\mathcal{C}_D$, which is simply the full covariance C_D projected into the subspace. For $\mathsf{C}_D = \mathsf{C}$, it is simply a unit matrix. Note that this can be written in terms of the inner product of Eq. (5) as

$$\hat{\chi}^2(\theta) = \langle\hat{X} - X(\theta)|\hat{X} - X(\theta)\rangle_D, \quad (18)$$

where the subscript D indicates that this is with respect to the metric $g_{\alpha\beta} = \mathcal{C}^{-1}_{D,\alpha\beta}$, rather than being strictly Euclidean.

Given a set of mocks, we may estimate the projected covariance directly from the measurements of $\hat{c}$ in each.[8] Since this contains fewer bins than the unprojected covariance, it is less susceptible to stochastic shifts arising from covariance matrix noise. The inference proceeds by computing the model power spectrum for each chosen parameter vector, projecting onto the reduced V_α basis, and computing the likelihood of Eq. (17) using the mock-based $\mathcal{C}_D$ subspace covariance. MCMC can then be performed using this subspace likelihood.[9] Subject to a suitably chosen $N_{\rm SV}$, we expect the incurred χ^2 error to be small, and thus the posteriors to be unbiased.

III. PARAMETER SHIFTS AND COVARIANCES

A. Noise-averaged constraints

It is important to show that the θ estimates obtained in this subspace formalism are unbiased and produce a good estimate of the parameter covariance. To do this, we write the data coefficients as $\hat{c} = c(\theta^*) + \hat{n}$ where θ^* are the true parameters and $\hat{n}$ is some noise vector, assumed to be uncorrelated with theory. For the simple case of matching true and fiducial $P(k)$ covariances ($\mathsf{C} = \mathsf{C}_D$), we have a diagonal subspace metric $\mathcal{C}^{-1}_{D,\alpha\beta} = \delta^K_{\alpha\beta}$, which implies that each element of $\hat{n}$ is independent and drawn from a Gaussian of unit variance. In contrast, due to the use of SVD to define the subspace basis vectors, we expect the magnitude of the $c_\alpha(\theta^*)$ coefficients to fall strongly with α; we thus expect the signal-to-noise of the modes to fall monotonically, such that the later components are noise dominated and can be justifiably removed.

The subspace log-likelihood is given by Eq. (17), which we may write as another inner product in the space spanned by $c(\theta)$,

$$\begin{aligned}\hat{\chi}^2(\theta) &= (\hat{c} - c(\theta)|\hat{c} - c(\theta)) \\ &= (c(\theta^*) - c(\theta) + \hat{n}|c(\theta^*) - c(\theta) + \hat{n}).\end{aligned} \tag{19}$$

The observed mean parameter vector $\hat{\theta}$ is defined by maximizing $\hat{\chi}^2$,

$$\left.\frac{\partial\chi^2}{\partial\theta_i}\right|_{\theta=\hat{\theta}} = 0 \Rightarrow \left(\left.\frac{\partial c(\theta)}{\partial\theta_i}\right| c(\theta^*) - c(\hat{\theta}) - \hat{n}\right)\Bigg|_{\theta=\hat{\theta}} = 0, \tag{20}$$

where $i \in \{1, \ldots, N_{\rm param}\}$ indexes the parameter vector.[10] Averaging over noise, this simply requires $c(\theta^*) - c(\hat{\theta}) = 0$, implying that $\hat{\theta} = \theta^*$ [assuming that the mapping $\theta \to c(\theta)$ is injective, which is usually the case in cosmological contexts]. Note this is true for any value of $N_{\rm SV}$; i.e., the subspace decomposition always gives an unbiased estimate of the parameters, as required.

For the parameter covariance, we start from the Fisher matrix, defined by

$$\begin{aligned}2\hat{\mathcal{F}}_{ij} &= \left.\frac{\partial^2\hat{\chi}^2(\theta)}{\partial\theta_i\partial\theta_j}\right|_{\theta=\hat{\theta}} \\ &= \left(\left.\frac{\partial c(\hat{\theta})}{\partial\theta_i}\right| \frac{\partial c(\hat{\theta})}{\partial\theta_j}\right) + \left(\left.\frac{\partial^2 c(\hat{\theta})}{\partial\theta_i\partial\theta_j}\right| c(\theta^*) - c(\hat{\theta}) - \hat{n}\right) \\ &\quad + (i \leftrightarrow j).\end{aligned} \tag{21}$$

Averaging over noise allows us to drop the second set of parentheses (since $\hat{\theta} = \theta^*$). From this, we can see that using less than $N_{\rm bin}$ SVs (i.e., restricting to a subspace) produces a shift in the noise-averaged Fisher matrix,

$$\Delta\mathcal{F}_{ij} = \left(\sum_{\alpha=1}^{N_{\rm SV}}\sum_{\beta=1}^{N_{\rm SV}} - \sum_{\alpha=1}^{N_{\rm bin}}\sum_{\beta=1}^{N_{\rm bin}}\right)\frac{\partial c_\alpha(\theta^*)}{\partial\theta_i}\mathcal{C}^{-1}_{D,\alpha\beta}\frac{\partial c_\beta(\theta^*)}{\partial\theta_j}. \tag{22}$$

This generates a corresponding shift in the noise-averaged parameter covariance $\Phi = \mathcal{F}^{-1}$,

$$\Delta\Phi_{ij} = -\sum_{kl}\mathcal{F}^{-1}_{ik}\Delta\mathcal{F}_{kl}\mathcal{F}^{-1}_{lj} = -\sum_{kl}\Phi_{ik}\Delta\mathcal{F}_{kl}\Phi_{lj}. \tag{23}$$

As shown in Appendix A, $\Delta\Phi_{ij}$ is positive semidefinite; i.e., restricting to a nontrivial subspace can only increase the parameter covariance. While this shift exists, it is expected to be small, since the c_α coefficients rapidly decrease with index α; thus we expect similar behavior for the parameter derivatives and hence contributions to $\Delta\Phi_{ij}$.

B. Noise in the data vector

While the above subsection gives the behavior of the subspace likelihood averaged over noisy realizations, it remains to consider how the best-fit parameter is shifted from the mean for individual noise realizations. Assuming

[8]It is worth noting that the subspace coefficients, c, do not have straightforward correspondences with individual cosmological parameters. As an example, consider an analysis measuring only the primordial amplitude. As required by the SVD, the first coefficient dominates, but will have contributions from *all* nuisance parameters that affect the amplitude, especially those that control the high-k regime, where the error bars are small.

[9]Note that it is not sufficient to simply use the template bank samples to compute the likelihood posterior surface (as is done in gravitational wave analyses, e.g., Ref. [30]). For a posterior that is compact in the prior space, there are very few template bank samples near the minimum point of the likelihood. In this example, we have a minimum log-likelihood of 300 in the data, versus just three in the MCMC output. This could be reduced by using more restrictive priors on the template bank.

[10]Throughout this section we assume that any Gaussian parameter priors (e.g., those on the higher-order bias parameters) are not strongly informative and may be ignored, a justifiable assumption in this work.

the inverse covariance matrix $\mathcal{C}_D$ to be known precisely, Eq. (B5) of Appendix B shows the shift in the best-fit parameter for a given noise realization $\hat{n}$ to be equal to

$$\delta\theta_i = \sum_j \left(\frac{\partial c(\theta^*)}{\partial \theta_j} \middle| \hat{n} \right) \Phi_{ij}, \tag{24}$$

adopting the inner product of Eq. (19) and neglecting (subdominant) noise in the parameter covariance, Φ. Since this is a stochastic quantity, it is best understood by computing the shift covariance, $\langle \delta\theta_i \delta\theta_j \rangle$. Using the definition $\langle \hat{n}_\alpha \hat{n}_\beta \rangle = \mathcal{C}_{D,\alpha\beta}$, the average can be simply obtained,

$$\langle \delta\theta_i \delta\theta_j \rangle = \Phi_{ij}. \tag{25}$$

As noted in the previous section, Φ_{ij} increases as $N_{\rm SV}$ is decreased; thus a reduction in the number of basis vectors necessarily increases the magnitude of best-fit parameter shifts. However, given that $c_\alpha(\theta^*)$ falls rapidly with increasing index α (due to the SVD), we expect higher SVs to have only small contributions to the shift covariance.

C. Noise in the precision matrix

Of greater importance in this work is noise in the precision matrix, which generally appears when estimating $\mathcal{C}_D^{-1}$ using a finite number, $N_{\rm mock}$, of mocks. As shown in Eq. (B8) in Appendix B, a stochastic shift in the precision matrix, $\delta\Psi$, leads to an additional shift in the best-fit parameters

$$\Delta\theta_i = \sum_{\alpha\beta j} \left[\Phi_{ij} \hat{n}_\beta \frac{\partial c_\alpha}{\partial \theta_j} - \sum_{kl} \Phi_{ik} \Phi_{jl} \left(\frac{\partial c}{\partial \theta_j} \middle| \hat{n} \right) \frac{\partial c_\alpha}{\partial \theta_k} \frac{\partial c_\beta}{\partial \theta_l} \right] \delta\Psi_{\alpha\beta}, \tag{26}$$

valid for any likelihood monotonic in $\hat{\chi}^2$. While the exact form is unimportant for our purposes here, we note that it depends on the product $\delta\Psi\hat{n}$; there is no shift from noisy precision matrices if the data is noiseless. For noisy data, this shift is present, though it vanishes on average providing that $\langle \delta\Psi \rangle = \langle \delta\Psi \hat{n} \rangle = 0$.[11] Averaging over statistical noise in both $\hat{n}$ and $\delta\Psi$ (assuming Gaussian statistics), we obtain the shift covariance from precision matrix noise,

$$\langle \Delta\theta_i \Delta\theta_j \rangle = \frac{(N_{\rm mock} - N_{\rm SV})(N_{\rm SV} - N_{\rm param})}{(N_{\rm mock} - N_{\rm SV} - 1)(N_{\rm mock} - N_{\rm SV} - 4)} \Phi_{ij} \tag{27}$$

[11]These conditions are nontrivial in some cases. Systematic parameter biases can be obtained from smooth models of the covariance matrix which have $\langle \delta\Psi \rangle \neq 0$ as well as covariance matrices estimated using the data, which have $\langle \delta\Psi \hat{n} \rangle \neq 0$.

[2,4], which includes the Hartlap factor required to debias the inversion of noisy covariance matrices [1]. In the limit of large $N_{\rm mocks} \gg N_{\rm SV}$, this is approximately

$$\langle \Delta\theta_i \Delta\theta_j \rangle \approx \frac{N_{\rm SV} - N_{\rm param}}{N_{\rm mock}} \Phi_{ij}. \tag{28}$$

It is here that we see the utility of using $N_{\rm SV} < N_{\rm bin}$. This dramatically decreases the parameter shifts obtained using small $N_{\rm mock}$, and, conversely, allows one to perform equally accurate inference using many fewer mocks, greatly improving the computational efficiency for surveys with a large number of bins. The standard way of accounting for such errors is to inflate the output parameter covariances [5] and thus lose constraining power; using fewer SVs reduces this need, implying that the parameter confidence contours will *reduce* as $N_{\rm SV}$ decreases.

The assumption of Gaussianity in the above discussion is not fully valid. As shown in Ref. [6], when the covariance matrix is estimated from a finite number of mocks, one should properly marginalize over its inverse, which leads to the likelihood being replaced by a multivariate t distribution of the following form:

$$-2 \log \mathcal{L}(\theta) \to N_{\rm mock} \log \left(1 + \frac{\hat{\chi}^2(\theta)}{N_{\rm mock} - 1} \right) + \text{const.} \tag{29}$$

with no Hartlap factors required. In general, this correction becomes important when $N_{\rm mock}$ becomes comparable to $N_{\rm SV}$ and alters the tails of the posterior distributions, causing a significant inflation relative to that of the naïve Gaussian likelihood (without the Hartlap correction factor). While Eq. (29) is simple to implement in a sampling code, we principally adopt the Gaussian likelihood in this work, since the alternative complicates some technical points pertaining to the analytic marginalization over nuisance parameters discussed in Appendix C. In fact, this makes very little difference to the output parameter constraints; assuming uninformative priors, the likelihood of Eq. (29) leads to the posterior distribution itself being a multivariate t distribution with $N_{\rm mock} - N_{\rm param}$ degrees of freedom. In general, this is large, hence the posteriors are very close to Gaussian, and we find very similar parameter constraints from the Hartlap-rescaled Gaussian likelihood and that of Ref. [6]. Furthermore, in both choices of likelihood we find stochastic shifts in the *best-fit* parameters arising from covariance matrix noise [Eq. (26)]. Several works advocate for the inclusion of an additional *inflation factor*, m_1, to ensure that the output parameter variances match those one would obtain by repeating the analysis many times allowing for stochastic noise in both the data vector and covariance. Here, we adopt the rescaling factor of Ref. [5], which is derived in the Gaussian limit and discussed further in Sec. IV D 3.

We emphasize that such posterior inflation is not strictly mathematically justified. Indeed, the inflation factor, m_1, essentially captures the difference between the true parameter variance obtained with a correct covariance matrix (i.e., one from infinitely many mocks) and that averaged over a large number of covariance matrix realizations. However, neither a set of such realizations nor the true covariance are available in actual parameter estimates. m_1 should thus be viewed as a phenomenological factor introduced in order to approximately compensate for the stochastic shifts of the best fit by the sampling noise in the covariance matrix. While this is the approach adopted by the BOSS and eBOSS collaborations, it is necessarily inferior to a fully Bayesian treatment.

D. Choice of fiducial covariance matrix

The choice of fiducial covariance C used to generate the SVD subspace warrants discussion. For $N_{\rm SV} = N_{\rm bin}$, this is arbitrary since the mapping from power spectra to coefficient space is invertible; equivalently, χ^2 is independent of C. In the general case it is important, since the fiducial covariance defines the mapping onto the lower-dimensional subspace. If C is equal to the power spectrum data covariance C_D, the subspace metric is Euclidean and all c_α coefficients are uncorrelated. Since we define basis vectors via SVD, we expect the information content of the basis coefficients to decrease monotonically with the index α; having a diagonal covariance ensures that we do not throw away information arising from correlations between low and high basis coefficients. In general, the subspace metric is defined by the projection

$$\mathcal{C}_D^{-1} = V\mathsf{C}^{T/2}\mathsf{C}_D^{-1}\mathsf{C}^{1/2}V^T \tag{30}$$

[Eq. (17)], which becomes more diagonal as C approaches C_D (since V is unitary). For this reason, using a poor estimate of C_D will lead to significant correlations between basis coefficients, and thus reduce the efficacy of the method by introducing a larger shift in the parameter covariance [Eq. (23)] as SVs are removed. It is thus desirable to use some input estimate of C which is close to the truth (to allow for efficient subspace projections) and low-noise (to ensure the basis vectors are smooth). It is worth noting, however, that *any* invertible choice of fiducial covariance gives unbiased parameter estimates in the limit of zero noise; optimizing this simply ensures that the posteriors are less affected by noise for a given (small) number of basis vectors.

IV. APPLICATION TO BOSS DR12

A. Mock datasets and covariances

We apply the formalism of the above section to galaxy power spectra taken from the twelfth data release (DR12) [57] of the Baryon Oscillation Spectroscopic Survey (BOSS), which is part of SDSS-III [58,59]. For simplicity, we use only the largest of the four data chunks, the high-z North Galactic Cap sample, with mean redshift $z = 0.61$ and volume $V = 2.8h^{-3}$ Gpc^3. In this work, all mock data are drawn from the publicly available [60] power spectrum multipoles from 2048 MultiDark-Patchy mocks (hereafter Patchy) [61–63], including 48 k bins in $[0.01, 0.25]h$ Mpc^{-1} for both monopole and quadrupole moments, as in Refs. [47–49,52]. Two choices of mock data are used: a single Patchy realization, which emulates the BOSS sample, and the mean-of-48 realizations, to ensure our analysis is unbiased.

For the covariance matrix, we consider three possible choices: (1) the sample covariance from Patchy mocks; (2) the analytic covariance presented in Ref. [14]; and (3) a simplified Gaussian (plus shot-noise) covariance computed for the best-fit BOSS cosmology. In the former case, we use either $N_{\rm mock} = 2000$ or 125 mocks (excluding those used to define the data vectors), defining the sample covariance via the usual formula

$$\begin{aligned}\mathrm{cov}(P_a, P_b) &\equiv \mathsf{C}_{D,ab} \\ &= \frac{1}{N_{\rm mock}-1}\sum_{i=1}^{N_{\rm mock}}(\hat{P}_a^{(i)} - \bar{P}_a)(\hat{P}_b^{(i)} - \bar{P}_b),\end{aligned} \tag{31}$$

where $\hat{P}_a^{(i)}$ is the power spectrum of the ith mock, subscripts indicate the k bin and multipole, and the overbar represents the mean over all mocks. When forming χ^2 we require the inverse covariance, Ψ_D; to ameliorate noise-induced bias, we include the Hartlap factor [1], defined by

$$\Psi_D = f_{\rm H} \times \mathsf{C}_D^{-1}, \qquad f_{\rm H} = \frac{N_{\rm mock} - N_{\rm bin} - 2}{N_{\rm mock} - 1}. \tag{32}$$

Since the other choices of covariance are analytic, they do not require a Hartlap factor. The first of these uses perturbation theory to compute the off-diagonal trispectrum terms and the contributions from super-survey modes, while the diagonal spectrum is estimated from Patchy. In contrast, the second assumes Gaussianity (and thus no off-diagonal terms before window convolution), with the diagonal computed via the one-loop theory model used in this work. Both include the BOSS window function, as discussed in Ref. [14].

B. Creation of the template bank and subspace coefficients

The template bank is generated following the procedure of Sec. II A. We first draw 10^5 points in the parameter space of Eq. (1), subject to the broad priors,

$$h \in [0.6, 0.74], \quad \omega_{cdm} \in [0.08, 0.16],$$
$$A_s/A_{s,\mathrm{fid}} \in [0.6, 1.4], \quad b_1 \in [1.7, 2.3],$$
$$b_2 \sim \mathcal{N}(0,1), \quad b_{G_2} \sim \mathcal{N}(0,1), \quad b_4 \sim \mathcal{N}(500, 500^2),$$
$$c_{s,0} \sim \mathcal{N}(0, 30^2), \quad c_{s,2} \sim \mathcal{N}(0, 30^2),$$
$$P_{\mathrm{shot}} \sim \mathcal{N}(0, 5000^2), \tag{33}$$

where $\mathcal{N}(\mu, \sigma^2)$ indicates a Gaussian prior and all quantities are in $h\,\mathrm{Mpc}^{-1}$-type units. Note that the flat priors are only for the generation of the template bank and are not used in the later MCMC analysis.[12] For each set of parameters, the power spectrum model is computed via one-loop EFT, as implemented in the `CLASS-PT` code [50], then convolved with the BOSS window function. Note that the priors are purposefully very broad; given more restrictive priors, the domain of the parameter manifold can be reduced, leading to fewer required subspace coefficients.

To compute the basis vectors of the subspace discussed in Sec. II A, we first require the fiducial covariance C, which defines the rotated power spectra X [Eq. (4)]. While setting this equal to the sample covariance would ensure that the subspace metric is diagonal, this is seldom desirable since, unless N_{mock} is very large, it will lead to noisy basis vectors and hence less optimal subspace decompositions. In general, it is important to use a fiducial covariance that is (a) relatively smooth, and (b) provides a fair approximation to the true covariance. For this reason, we principally set C equal to the analytic covariance matrix of Ref. [14], as described above.[13]

Using this, we rotate the template spectra into the X variables and apply SVD as in Eq. (8) to compute the set of basis vectors $\{V_\alpha\}$. This further defines the SVs $\{D_\alpha\}$ which can be used to set N_{SV}.[14] To set N_{SV} we use Eq. (12), enforcing that χ^2 be reproduced to within 0.1, averaged across the prior, which we expect to give reasonable results. This requires $N_{\mathrm{SV}} \approx 12$ (depending on the exact covariance); i.e., all SVs above this index contribute less than 0.1 to the mean χ^2. This corresponds to compressing the data by a factor of 8 and is slightly larger than the number of parameters (10), indicating the effects of nonlinearities in the transformation $\theta \to P(\theta)$ (and hence the likelihood) that cannot be well represented as linear across the prior domain. Alternatively, this signifies a breakdown of the assumption that all information can be encapsulated in the rank-N_{param} Fisher matrix. Below, we will also consider $N_{\mathrm{SV}} = 48$ and the uncompressed power spectrum vector for testing purposes. Had we used more cosmological or nuisance parameters in our model, we would expect N_{SV} to increase correspondingly.

Given the above Patchy mocks, we generate corresponding subspace coefficients $\{\hat{c}_\alpha^{(i)}\}$ using the first N_{SV} basis vectors via Eq. (6). These are then used to compute the sample covariance of the coefficients via the standard formula,

$$\mathrm{cov}(c_\alpha, c_\beta) \equiv \mathcal{C}_{D,\alpha\beta} = \frac{1}{N_{\mathrm{mock}} - 1} \sum_{i=1}^{N_{\mathrm{mock}}} (\hat{c}_\alpha^{(i)} - \bar{c}_\alpha)(\hat{c}_\beta^{(i)} - \bar{c}_\beta) \tag{34}$$

[cf. Eq. (31)], which is used in the subspace likelihood [Eq. (17)]. As for the former covariance, this requires a Hartlap factor to invert; the upside is that the Hartlap factor will be closer to unity here if $N_{\mathrm{SV}} < N_{\mathrm{bin}}$, shrinking the parameter covariances. Additionally, we will find it useful to perform the analysis using analytic covariances within the likelihood (rather than just as the fiducial covariance); in this case, we can form the subspace data covariance from Eq. (30).

C. Posterior sampling methods

With the datasets and basis vectors in hand, posterior surfaces are computed using MCMC methods, here implemented via MontePython v3.3 [64,65]. Given the necessity to run MCMC a significant number of times to validate our method, we implement two approaches to expedite the sampling. First, we note that any parameter which enters the power spectrum model $P(\theta)$ linearly can be analytically marginalized without loss of information [66,67]. This procedure is described in Appendix C and simply leads to those parameters being absorbed into a cosmology-dependent covariance matrix. In our case, the nuisance parameters $b_4, c_{s,0}, c_{s,2}$, and P_{shot} fall in this category and are thus marginalized over directly. Note that this does not affect the SVD decompositions since it is applied only in the final MCMC step.

Second, when computing multiple posteriors at similar locations, it is illogical to rerun the MCMC each time from scratch. Instead, we opt to run the MCMC once, saving the (unprojected) power spectra for each point in parameter space, and then use these samples to reconstruct the posterior for a different analysis, e.g., a different choice of N_{SV}, via standard importance sampling techniques. This is fast (since it does not require a Boltzmann code to be run) and highly accurate for posteriors close to the initial MCMC chain. The latter requirement can be assessed by the *effective sample size* of the resampled chain (e.g., [68]), equal to

[12]The 10^5 samples are likely overkill; the change to the dominant basis vectors is negligible when this is reduced to 10^4 samples and remains small even for $N_{\mathrm{bank}} = 10^3$. Both of these cases require the same number of basis vectors (12) to incur a prior-averaged χ^2 error below 0.1.

[13]Note that having $\mathsf{C} \neq \mathsf{C}_D$ does not bias our inference, though it changes the efficiency of our subspace projection. If we use the same number of SVs as bins, the likelihood is independent of our choice of C. Assuming the sample covariance to be fixed, our analysis is robust to inaccuracies in the fiducial covariance.

[14]See Fig. 6 for a representative plot of the D_α components.

$$\mathrm{ESS} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}, \tag{35}$$

where w_i are the importance sampling weights. For a new posterior, if the effective sample size is too small ($\lesssim 10^4$ samples) when importance sampling from any preexisting set of spectral samples, we simply create a new MCMC chain for this configuration.

D. Results

Below, we present the results of the subspace analyses of the mock power spectrum data. Unless otherwise specified, we use a single realization of mock data, set the fiducial covariance to that of Ref. [14] (including both Gaussian and non-Gaussian components), and use a sample covariance from 2000 Patchy mocks. In all cases, we plot the derived parameters H_0, Ω_m, and σ_8; in the true Patchy cosmology, these are given by 67.77, 0.3071, and 0.8288, respectively.

1. Bias in the noiseless limit

The first important check is whether the subspace analysis gives an unbiased estimate of cosmological parameters in the limit of zero noise in the data vector.

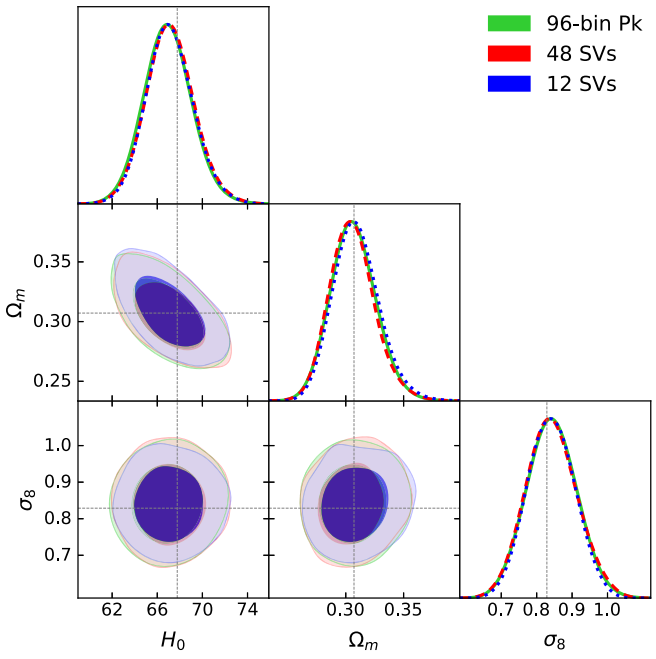


FIG. 1. Corner plot of the MCMC posterior from analyzing the mean of 48 Patchy mocks, using both the standard analysis in 96 power spectrum bins (green, full lines), 48 subspace coefficients (red, dashed lines), and 12 subspace coefficients (blue, dotted lines). The true Patchy cosmology is indicated by the vertical lines in the 1D histograms. All plots are computed using a 2000-mock Patchy covariance matrix, with an analytic covariance used to define the SVD subspace projection, following the method discussed in Sec. II. The posterior is clearly unbiased in all cases.

Figure 1 displays the posterior contours when the above analysis is applied to the mean-of-mocks dataset, comparing 96 uncompressed power spectrum bins, 48 SVs, and 12 SVs. Comparing the modal parameters to the true Patchy cosmology (from which the mock data are generated) shows excellent agreement in all cases; furthermore the posterior shapes are consistent between all datasets, with no noticeable increase in the parameter variances due to the subspace projections. This matches the theoretical calculation of Sec. III A, which asserted that the subspace likelihood would give an unbiased estimate of the mean, but a small increase to the parameter covariance; we here confirm that this increase is negligibly small even when using $N_{\mathrm{SV}} = 12$.

2. Bias from a single realization

Next, we consider the analogous results using data from a single Patchy mock, emulating the true observational sample. This uses the 2000-mock sample covariance; thus we expect the effects of parameter shifts induced by covariance matrix noise to be small. As discussed in Sec. III B, we expect some stochastic parameter shifts as a result of the differing noise properties, since data vectors with lower N_{SV} include fewer noise-dominated modes. These results are shown in Fig. 2 and confirm our suspicions; there is a noticeable shift in parameters (in particular Ω_m) as we move from 96 to 48 to 12 bins in the

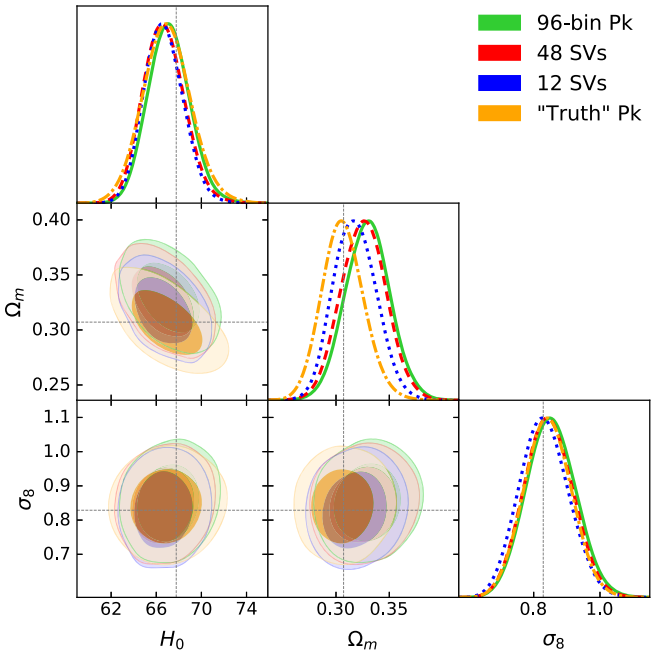
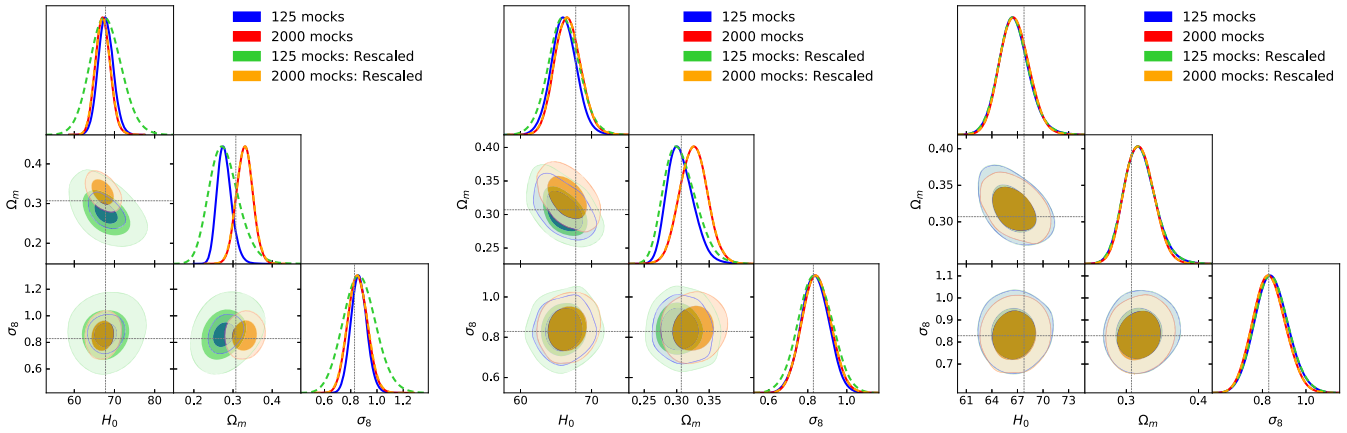


FIG. 2. As Fig. 1, but analyzing a single Patchy mock. We additionally show the "true" results from the mean of 48 mocks in yellow (dot-dashed lines). Here, restricting to fewer basis coefficients does induce a shift in the parameters (as discussed in Sec. III B), due to slightly different noise properties, since fewer noisy modes are included.



(a) 96-bin Power Spectrum (b) 48 Subspace Coefficients (c) 12 Subspace Coefficients

FIG. 3. Comparison of the posterior parameter contours from analyses using a sample covariance matrix created from 125 mocks (blue, full lines) or 2000 mocks (red, full lines), for various choices of data compression. Notably, there are significant shifts induced by precision matrix noise, but these are almost completely nulled by using $N_{\rm SV} = 12$, the value indicated by χ^2 constraints. In green and yellow dashed lines we show the corresponding results including the m_1 rescaling factor of Eq. (36) to account for precision matrix noise. With the rescaling, the posteriors with $N_{\rm mock} = 125$ and $N_{\rm mock} = 2000$ are not in tension, though the precision matrix noise leads to a significant loss of constraining power when many bins are used.

data vector. While in this example the posteriors shift in the direction of the noiseless limit, we expect the sign of this shift to vary for different noise realizations. In any case, although the noise properties of the subspace likelihoods differ somewhat from those of the true data, the results of the previous subsection confirm to us that the analysis is not biased as a result.

3. Parameter shifts from precision matrix noise

Perhaps the most important test is whether restricting to a small number of subspace coefficients allows us to use fewer mocks. To test this, we simply run the likelihood analysis with a single mock dataset and two sets of covariance matrices; one generated from 2000 mocks and one with 8 times fewer (matching the factor by which we optimally compress the data). Note that we assume a Gaussian likelihood in both instances (apropos of the discussion in Sec. III C). These results are shown in the first two datasets of Fig. 3, and we immediately note large shifts in the best-fit parameters for the uncompressed likelihood analysis, alongside an inflation of the parameter error bars, matching the predictions of Sec. III C. Notably, these shifts dramatically decrease as we compress the data, and, for $N_{\rm SV} = 12$ (the value suggested by the constraints on $\Delta\chi^2$ in Sec. II A), these are insignificant. This is the main result of this paper; we can obtain robust estimates of cosmological parameters with just $\mathcal{O}(100)$ mocks when using optimal subspace projections.

While the above discussion correctly shows the stochastic shifts induced in the cosmological parameters from precision matrix noise, it does not accurately reflect real analyses, which usually account for this by inflating the parameter error bars by rescaling the Gaussian parameter covariances by the constant factor[15]

$$m_1 = \frac{1 + B(N_{\rm bin} - N_{\rm param})}{1 + A + B(N_{\rm param} + 1)}, \tag{36}$$

defining

$$A = \frac{2}{(N_{\rm mock} - N_{\rm bin} - 1)(N_{\rm mock} - N_{\rm bin} - 4)},$$
$$B = \frac{N_{\rm mock} - N_{\rm bin} - 2}{(N_{\rm mock} - N_{\rm bin} - 1)(N_{\rm mock} - N_{\rm bin} - 4)} \tag{37}$$

[5]. This is more complex than that of Eq. (27) (which is equal to the numerator of m_1 minus one), since noise in the covariance matrix inflates the observed parameter error bars in addition to giving a best-fit parameter shift. When rescaling the observed parameter covariances to account for best-fit variation, this must be accounted for, giving the reduced shift of Eq. (36) [62]. The m_1 numerator [equivalently, Eq. (27)] gives the necessary parameter covariance inflation relative to the $N_{\rm mock} \to \infty$ case, which is a good indicator of the efficacy of the subspace decomposition; for $N_{\rm mock} = 125$ and $N_{\rm param}$, we obtain $1 + B(N_{\rm bin} - N_{\rm param}) \approx 4.3, 1.5, 1.0$ for $N_{\rm bin} = 96, 48, 12$, while $m_1 \approx 3.0, 1.3, 1.0$ for the same data. The results including m_1 are shown in the

[15]It is important to note that we focus only on data vectors that are not used to build the sample covariance. For this reason, we do not include the M_2 factor used in the mock catalog analyses of Refs. [5,69]. Our m_1 factor is equivalent to M_1^2 in the notation of Ref. [69]. Furthermore, we note the discussion at the end of Sec. III C regarding the applicability of the m_1 factor.

third and fourth datasets in Fig. 3; we note that the width of the one-dimensional histograms for the $N_{\rm mock} = 125$, $N_{\rm bin} = 96$ case is inflated by $\sim 2x$, as expected. Following the rescaling, the results for $N_{\rm mock} = 125$ and 2000 are broadly consistent, though this is with a significant loss of precision, or equivalently, an effective survey volume. As above, we note that strictly the likelihood should be replaced by a modified t distribution to properly account for the effects of covariance matrix noise; thus the above rescaling, which is derived in the Gaussian limit, is not fully valid. Switching to the modified likelihood was found to have an insignificant effect in our context, even for $N_{\rm mock} = 125$.

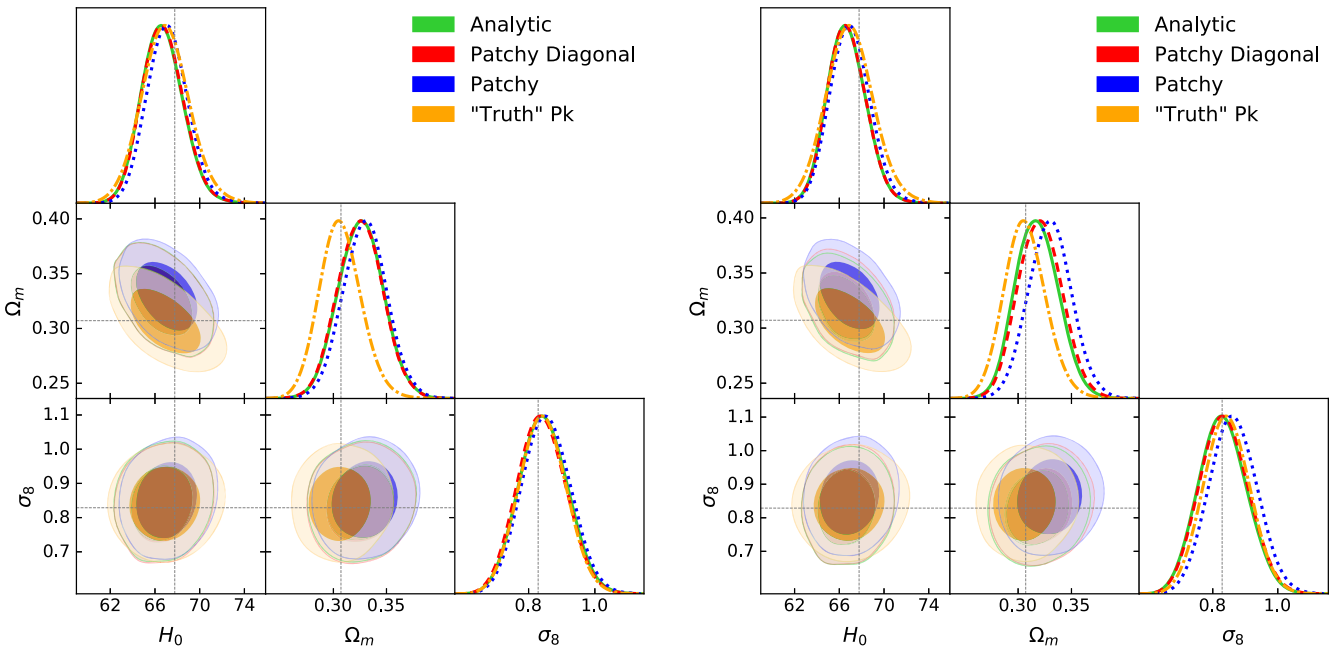
4. Changing the fiducial covariance

An important hyperparameter is the fiducial covariance, C, used to define the subspace. In Sec. III D, it was stated that, while any (invertible) choice of fiducial covariance would lead to an unbiased estimate of cosmology when averaged over noise, noisy estimates, or those far from the true covariance, would lead to greater noise-induced shifts for small numbers of SVs. To test this, Fig. 4 compares the posterior predictions from a single mock power spectrum analyzed with three fiducial covariances: (1) the analytic covariance discussed above (with off-diagonal terms); (2) the diagonal of the Patchy covariance matrix from 2000 mocks; and (3) the full Patchy covariance matrix. In all cases, the data covariance is held constant.

Notably, we observe no significant differences between posteriors when using $N_{\rm SV} = 48$ but larger shifts with $N_{\rm SV} = 12$. In particular, we note a significant shift from using the Patchy fiducial covariance; this is attributed to the large off-diagonal noise present therein, which leads to less efficient subspace decompositions. This conclusion is strengthened by the results with the diagonal of the covariance matrix; while this does not accurately represent the true covariance (since window effects induce nontrivial mode coupling and hence off-diagonal terms are present even if one assumes Gaussianity), it has low noise and is consistent with the analytic prediction. We thus conclude that it is important to have a relatively smooth estimate of the fiducial covariance (though not necessarily one that matches the true covariance), else the low-SV results will be significantly affected by noise.

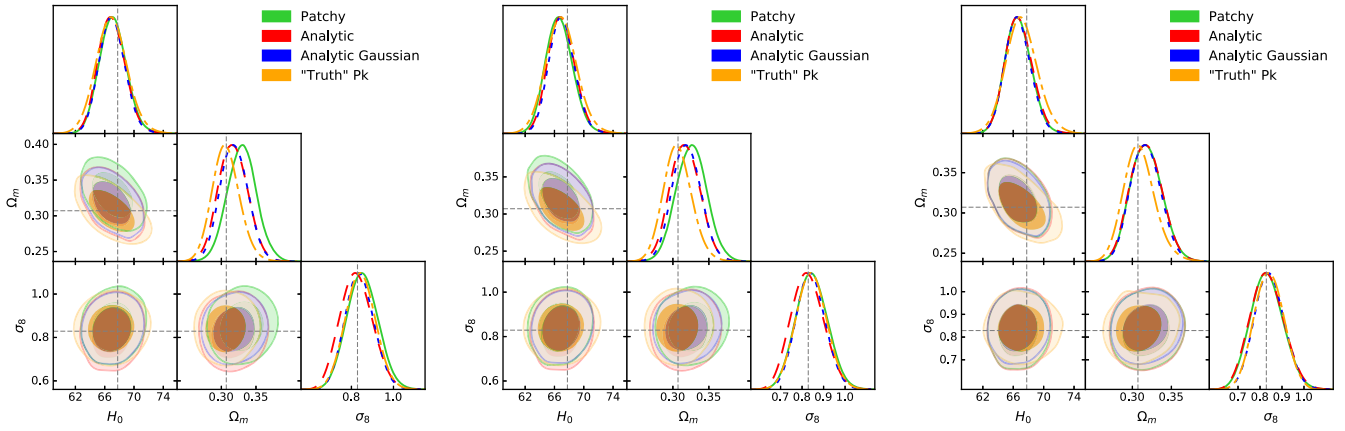
5. Changing the data covariance

Can consistent results be obtained from analyses using different data covariances? This question extends beyond subspace analyses, but can be well probed in our formalism, since we can separate out the effects of precision-matrix noise by reducing the number of basis coefficients. In Fig. 5 we display the results from a selection of covariance matrices:



(a) 48 Subspace Coefficients (b) 12 Subspace Coefficients

FIG. 4. Comparison of the parameter posteriors estimated using different choices of the fiducial covariance matrix, C, to define the subspace decomposition. Results are plotted for the analytic covariance of Ref. [14] (green, full lines), the diagonal of the sample covariance from 2000 Patchy mocks (red, dashed lines), and the full Patchy covariance (blue, dotted lines). We additionally show the posterior from the mean-of-mocks analysis of Fig. 1 (yellow, dot-dashed lines) and note that all likelihoods use the full 2000-mock Patchy covariance in the likelihood. For 48 SVs, there is little difference between choices of fiducial covariance, but some shifts for 12 SVs. We attribute this to noise in the SVD basis vectors induced by using noisy fiducial covariances (i.e., Patchy).



(a) 96-bin Power Spectrum (b) 48 Subspace Coefficients (c) 12 Subspace Coefficients

FIG. 5. Estimated posteriors from analyses using different choices of covariance matrix in the likelihood. Data are shown for the 2000-mock Patchy covariance (green, full lines), the analytic covariance of Ref. [14] (red, dashed lines), and an analytic Gaussian covariance (blue, dotted lines) alongside the mean covariance of Fig. 1 (yellow, dot-dashed lines). In all cases, the fiducial covariance (used to define the subspace basis vectors) is set to the analytic covariance of Ref. [14], and we analyze a single mock dataset with differing compressions. Using 12 SVs, the posteriors are identical for each covariance, though there are larger noise-induced discrepancies when using a greater numbers of bins.

the 2000-mock Patchy covariance, the analytic prescription of Ref. [14] (including trispectrum and supersample terms), and an analytic Gaussian covariance. In the conventional analysis using 96 power spectrum bins, there are seen to be significant ($\sim 0.5\sigma$) shifts in Ω_m when different covariances are used, especially between Patchy and the analytic prescriptions.[16] Notably, adding the trispectrum terms to the covariance does *not* appear to induce a significant parameter shift. As $N_{\rm SV}$ decreases, these differences become small, and, by 12 SVs, we report no obvious deviations in the best-fit parameters when using different covariances.

From these results, one may conclude that, for a BOSS-like sample volume and tracer density, the trispectrum terms in the covariance do not alter the output parameter posteriors. Further, there is a bias induced by using the publicly available Patchy covariance, which disappears as the number of bins is reduced. This can thus be attributed to residual noise in the covariance matrix, and must carefully be taken into account to avoid biasing the output cosmology, for example, by using the subspace-based analysis. These conclusions agree with the results of Ref. [70], which analyzed the BOSS data using the perturbation theory covariance matrices [14].

V. DISCUSSION

A. Dependence of number of SVs on priors and the survey volume

In Sec. IV B, we found that, for the BOSS sample used in this work, 12 SVs were needed to ensure that the mean subspace χ^2 was consistent with the usual power spectrum χ^2 to within $\chi^2_{\rm min} = 0.1$. This is not a general statement, since it depends on the fiducial covariance matrix, the cosmological model, and our choice of priors. In the above, we opted to use broad priors implying a lack of knowledge of the posterior; tighter priors lead to a smaller variation of χ^2 across the parameter manifold, thus requiring fewer SVs. A simple test of this is to repeat the SVD for power spectrum samples drawn from the *posterior* rather than the prior template bank. This is simply done and represents the minimum number of SVs which can safely be used, since it is the number one would obtain if they started with full knowledge of the posterior space. In this case, we find that only 8 subspace coefficients are required, rather than 12. This is *less* than the number of model parameters, indicating that the parameters are partially degenerate. A hybrid approach might also be possible; i.e., one could approximate the posterior as a multivariate Gaussian and draw template banks instead from this distribution. While this would result in basis vectors that are more tailored to the posterior peak and thus somewhat fewer SVs, it will fail if the posterior is not well modeled as a Gaussian (noting that a Gaussian posterior is not necessarily implied by a Gaussian likelihood). For the sake of generality, we have not implemented such an approach here.

The required number of SVs also has dependence on the observational volume. If the survey size is increased by a factor f, we expect the covariance to fall by a factor f, and thus the SVD matrix D_α to scale as $f^{1/2}$. From Eq. (13), we may mimic this by choosing $N_{\rm SV}$ such that the cumulative signal-to-noise of excluded SVs is less than $\chi^2_{\rm min}/f$ rather than $\chi^2_{\rm min}$. For the priors considered herein, increasing the survey volume by a factor of 10 (making it comparable to

[16]These shifts are larger than those of Ref. [70]; this is due to our restriction to a single data chunk, and the different random catalog used in the former work.

the DESI volume), while making the unrealistic assumption that there is still only one tomographic bin, the number of required SVs increases from 12 to 16. This increase is due to the nonlinear parameter dependencies becoming more important (with respect to noise), though we note it to still be a small number compared to the 96 power spectrum bins, due to the steepness of the D_α.

B. Application to bispectra

While the main focus of this paper has been the galaxy power spectrum, the subspace decomposition is fully general and can be applied to any observable, given a theory model and a set of priors. Of particular interest is the galaxy bispectrum, since this statistic generically has a large number of bins, which has limited its applicability in previous mock-based approaches.[17] Any formalism that is able to substantially compress the bispectrum, while retaining its information content, is thus of great importance, allowing mock-based analyses to take place in reasonable computation times.[18] To test the applicability of our method to the bispectrum, we may simply ask the question: how many SVs are needed to reproduce the bispectrum likelihood to within $\Delta\chi^2 = 0.1$?

For this forecast, we consider a simplified scenario; a survey with the effective volume and redshift of the largest BOSS data chunk, but with a periodic box geometry. We assume the same power spectrum model as above (one-loop effective field theory), keeping the k-binning constant (i.e., with 48 k bins for each of the monopole and quadrupole). For the (redshift-space monopole) bispectrum, we use tree-level theory as in Ref. [72], for k in $[0.01, 0.15]h$ Mpc^{-1}. This gives a total of 2135 bispectrum bins and 96 power spectrum bins. To generate the subspace basis vectors we require also a fiducial model for the joint covariance (Sec. II A); for this we assume a Gaussian covariance for both observables (matching that of Ref. [72]), noting that the exact choice of fiducial covariance is of limited importance in the analysis. For simplicity, we assume the power spectrum and bispectrum to be uncorrelated, i.e., that they are observed in separate regions of the sky. By instead inserting a Gaussian model for the cross covariance [73], we have shown that this assumption does not have a significant impact on the observed χ^2 or number of SVs.

A set of 10^4 template bank samples is then computed from the prior and the SVD performed. As in Sec. II A, the output SVs can be used to assess the impact on the prior-averaged χ^2 from including only a subset of all bins; in Fig. 6, we plot the χ^2 deficit for analyses using (a) only the power spectrum, (b) only the bispectrum, and (c) their combination. As previously noted, there is a steep dependence of $\Delta\chi^2$ on the number of SVs, with negligible contributions from the higher basis vectors. As before, we consider the optimal $N_{\rm SV}$ to be the minimum number which yields a total (prior-averaged) χ^2 error below 0.1; this is found to be 12 for the power spectrum, 9 for the bispectrum, and 21 for their combination.[19] Further, any correlations between observables are expected to reduce the combined number. The efficacy of this method is impressive; we can compress a combined data vector with 2231 bins into just 21 (even in the presence of nonlinearities in the likelihood), which would allow easy analysis with $\mathcal{O}(100)$ mocks. This has further implications for the *measurement* of bispectra from data; by including the optimal basis vectors directly in the bispectrum estimators, the relevant compressed statistics can be computed in far fewer operations than before, an important quality given the nontrivial computation time required for bispectrum estimation.

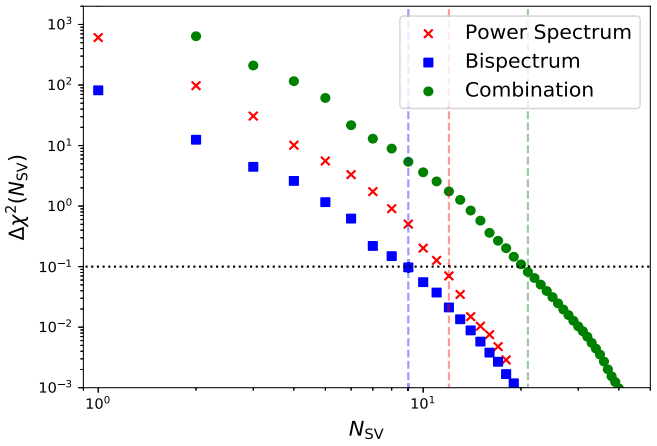


FIG. 6. Impact on χ^2 from using a finite number of basis vectors in the likelihood for mock analyses using the power spectrum, the bispectrum, and the combination of two probes (as described in the main text). For each case we plot the difference between the true χ^2 (from all $N_{\rm bin}$ bins) and that using $N_{\rm SV}$ SVs. The horizontal line indicates $\Delta\chi^2 = 0.1$; i.e., beyond this point, the sum of all remaining SVs contribute less than 0.1 to the total log-likelihood. For the power spectrum analysis, 12 basis vectors are required to reach this limit (compared to 96 bins), while for the bispectrum, only 9 are needed (compared to 2135 bins). In combination, one needs 21 SVs, though this is an upper bound as the two observables are assumed to be uncorrelated.

Similar analysis is possible for the combination of full-shape analyses, such as that presented herein, with Baryon Acoustic Oscillations (BAO) information from reconstructed power spectra (as in Ref. [49]). This is straightforward to test; we simply generate a template bank of the

[17]As an example, Ref. [71] had to use broad k bins to ensure that the covariance could still be estimated from the 2048 available mock catalogs.

[18]See Ref. [31] for an approximate bispectrum compression method using bispectrum eigenmodes.

[19]If we restrict only to modes with $k < 0.1h$ Mpc^{-1} for the tree-level bispectrum, we find that only 5 (17) SVs are needed for the bispectrum-only (combined) analysis.

power spectrum multipoles coupled with the Alcock-Paczynski parameters, $\alpha = \{\alpha_\parallel, \alpha_\perp\}$, and apply the SVD, supplementing the fiducial covariance with the measured joint covariance of α and multipoles discussed in Ref. [49]. Here, we find that only a single additional extra SV is required to accurately approximate the (significantly more constraining) χ^2.

VI. SUMMARY

In this work, we have introduced a formalism to reduce the dimensionality of cosmological observables by linearly projecting the observables into subspaces that maximize the prior-averaged χ^2. Given only a theoretical model and a set of priors, a set of model predictions can be computed and used to define a quasi-Euclidean metric space via a singular value decomposition. This provides a natural compression for the observables; by restricting to the first $N_{\rm SV}$ basis vectors of the subspace, the dimensionality of the data is significantly reduced while the likelihood remains almost unchanged. While we assume the likelihood itself to be Gaussian, it can be non-Gaussian in the parameters themselves (and indeed multimodal), since our decomposition is blind to the model parametrization. Observables such as the power spectrum and bispectrum can be represented with $\mathcal{O}(10)$ subspace coefficients, incurring χ^2 errors of <0.1 over the broad prior manifold, even in the presence of nonlinearities in the likelihood. While the method has some dependence on an assumed fiducial covariance, this does not bias the inference, and requires only a simple (low-noise) estimate.

The principal appeal of this concerns covariance matrices; likelihood analyses are inherently sensitive to precision matrix noise induced by the use of finite numbers of mocks, which is drastically reduced by data compression. Indeed, we prove that restricting to low-dimensional subspaces significantly reduces the induced parameter shifts, while keeping the noise-averaged constraints unchanged (except for a slight increase in parameter covariances, which is shown to be negligible in practice). Indeed, due to the posterior inflation required to ensure that the estimators remain appropriate in the presence of noise, our subspace projection gives more precise parameter constraints when the number of mocks is small.

Our method is validated by application to mock BOSS DR12 power spectrum multipoles. In particular, we show that the 96-bin power spectrum data can be robustly compressed to a set of 12 subspace coefficients, allowing for accurate parameter estimations using only 125 mocks. The required number of coefficients has only weak dependence on the prior space; adding maximally restrictive priors would allow the number of coefficients to be reduced to only 8, in our 10-parameter model. While we have assumed a simple ΛCDM model in this work, this is by no means a restriction; the formalism can apply to any extension for which model spectra can be evaluated and further include arbitrary parameters describing systematics. While one should recompute the basis vectors given a new cosmological model, this is not a major concern, since creating the template bank requires computing far fewer spectra than needed for a well-converged MCMC analysis.

It is useful to compare our methodology to the SVD-based techniques used to de-noise the power spectrum and bispectrum covariance matrices in Refs. [74–76]. In this approach one diagonalizes the covariance matrix and uses only the highest signal-to-noise (S/N) eigenvalues in the likelihood analysis. In contrast to S/N, we use $\Delta\chi^2$ from variations of all relevant parameters as a criterion to define the subspace projections, which allows us to better capture the eigenmodes that are sensitive to features in the power spectrum shape. Indeed, we have seen that our method can fully extract the information encoded in the BAO wiggles, which contribute very little to the overall S/N, but dominate the distance constraints.

With the addition of just a small number of extra coefficients, the method can also apply to larger volumes (where nonlinearities in the likelihood become more important, and thus Fisher-based compressions become suboptimal), as well as more complex statistics, such as bispectra or combination with BAO constraints. For the BOSS sample analyzed herein, we need only 12 SVs to encapsulate 96 power spectrum bins; this rises to 16 for a ten-times larger survey (such as DESI), and only a further eight are required to incorporate the ∼2000 bin bispectrum. An additional possibility is for projected statistics such as lensing; the large number of bins across different redshifts in 3×2 pt analyses could be substantially reduced using SVs. In general, the exact number of subspace coefficients depends on both the chosen parameter and the experimental setup (e.g., survey redshift and volume) and should thus be computed separately for each experiment, but this is a trivial operation to implement once the SVD has been performed.

The methods developed herein provide a useful sandbox in which to investigate an important question in cosmology: how parameter estimates depend on power spectrum covariances. Focusing on the largest data chunk of BOSS, we find that the inclusion of off-diagonal trispectrum terms in the covariance does not affect the parameter estimates, though this will generically depend on the shot noise and k binning. Further, in the uncompressed likelihood, there is a $\sim 0.5\sigma$ stochastic shift in the best-fit posteriors for Ω_m using a covariance drawn from MultiDark-Patchy mocks compared to that using analytic covariances, away from the true cosmology and in the direction of increasing tension with *Planck*. In contrast, using a small number of subspace coefficients, all covariances yield consistent results, indicating that (a) there is residual noise in the mock-based covariance that shifts the inferred cosmological parameters even if 2000 mock catalogs are used, and (b) analytic prescriptions can be

accurately applied. Subspace projection provides a simple way of ameliorating this, through limiting the effects of covariance matrix noise. Furthermore, the ability to perform analyses using significantly fewer mocks paves the way to allowing the covariance matrix to be parameter dependent; the computational cost of performing repeated analyses gradually updating the covariance matrix becomes far more manageable.

Approaches such as that of this work are vital when using statistics beyond the power spectrum. In particular, bispectrum analyses should no longer be considered mock-limited, opening up new and exciting avenues into the exploration of cosmological datasets. The future grows brighter for higher-order statistics.

ACKNOWLEDGMENTS

We thank Jay Wadekar and Roman Scoccimarro for providing us with analytic covariance matrices and sharing a draft of Ref. [70]. We additionally thank Alex Hall, Alan Heavens, Dragan Huterer, David Spergel, Masahiro Takada, Jay Wadekar, Ben Wandelt, Martin White, and the anonymous referee for comments on the manuscript. O. H. E. P. acknowledges funding from the WFIRST program through NNG26PJ30C and NNN12AA01C and thanks the Max Planck Institute for Astrophysics for hospitality when this work was being finalized. M. I. is partially supported by the Simons Foundations *Origins of the Universe* program. M. S. acknowledges support from the Corning Glass Works Fellowship and the National Science Foundation. M. Z. is supported by NSF Grant No. PHY-1820775, the Canadian Institute for Advanced Research (CIFAR) Program on Gravity and the Extreme Universe, and the Simons Foundation Modern Inflationary Cosmology initiative. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is [77]. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

APPENDIX A: INCREASE IN PARAMETER COVARIANCE FROM SUBSPACE PROJECTION

We here demonstrate that the shift in the parameter covariance [Eq. (23)] is positive semidefinite, implying that the restriction to a subspace increases the covariance on average. For convenience, and without loss of generality, we adopt the parameter set ψ in which the Fisher matrix is diagonal (this can always be found by diagonalization). We further rotate the coefficients $c(\psi) \to \tilde{c}(\psi) = \mathcal{C}_D^{-1/2} c(\psi)$ such that the subspace metric is diagonal. In this basis, the full Fisher matrix is simply

$$\mathcal{F}_{ij} = \frac{\delta^K_{ij}}{\sigma_i \sigma_j} \Rightarrow \Phi_{ij} = \delta^K_{ij} \sigma_i \sigma_j, \quad \Delta\Phi_{ij} = -\sigma_i^2 \sigma_j^2 \Delta\mathcal{F}_{ij}, \tag{A1}$$

where σ_i^2 is the variance of parameter ψ_i. The Fisher matrix in the subspace defined by $N_{\rm SV} \leq N_{\rm bin}$ SVs is given by

$$\begin{aligned}\hat{\mathcal{F}}_{ij} &= \sum_{\alpha=1}^{N_{\rm SV}} \sum_{\beta=1}^{N_{\rm SV}} \frac{\partial \tilde{c}_\alpha(\psi^*)}{\partial \psi_i} \frac{\partial \tilde{c}_\beta(\psi^*)}{\partial \psi_j} \\ &\leq \sqrt{\sum_{\alpha=1}^{N_{\rm SV}} \left(\frac{\partial \tilde{c}_\alpha(\psi^*)}{\partial \psi_i}\right)^2 \sum_{\beta=1}^{N_{\rm SV}} \left(\frac{\partial \tilde{c}_\beta(\psi^*)}{\partial \psi_j}\right)^2},\end{aligned} \tag{A2}$$

using the Cauchy-Schwarz inequality in the second line. Since $N_{\rm SV} \leq N_{\rm bin}$ and the derivatives are real, each component of the sum is positive; hence

$$\hat{\mathcal{F}}_{ij} \leq \sqrt{\sum_{\alpha=1}^{N_{\rm bin}} \left(\frac{\partial \tilde{c}_\alpha(\psi^*)}{\partial \psi_i}\right)^2 \sum_{\beta=1}^{N_{\rm bin}} \left(\frac{\partial \tilde{c}_\beta(\psi^*)}{\partial \psi_j}\right)^2} = \frac{1}{\sigma_i \sigma_j}, \tag{A3}$$

and thus

$$\Delta\mathcal{F}_{ij} \equiv \hat{\mathcal{F}}_{ij} - \mathcal{F}_{ij} \leq \frac{1}{\sigma_i \sigma_j}(1 - \delta^K_{ij}). \tag{A4}$$

Contracting this with an arbitrary vector x, one can easily show this to be negative semidefinite,

$$\sum_{ij} x_i \Delta\mathcal{F}_{ij} x_j = \left[\sum_i \left(\frac{x_i}{\sigma_i}\right)^2 - \sum_i \frac{x_i^2}{\sigma_i^2}\right] \leq 0, \tag{A5}$$

where we have again invoked the Cauchy-Schwarz inequality. From Eq. (A1), the parameter covariance shift $\Delta\Phi$ is thus positive semidefinite.

APPENDIX B: PARAMETER SHIFTS FROM NOISY DATA AND NOISY COVARIANCES

We here derive the expected stochastic shift in the best fit parameters from both noisy data and a noisy covariance matrix estimate. Much of these derivations parallel Refs. [2,4], but we include them here for completeness. To begin, consider the observed χ^2 using noisy precision matrix $\hat{\Psi} = \Psi + \delta\Psi$ ($\Psi \equiv \mathcal{C}_D^{-1}$ in the notation of Sec. II) and data $\hat{c} = c(\theta^*) + \hat{n}$ where θ^* are the true parameters and n is some vector of noise (with each element drawn from a unit Gaussian for a diagonal subspace metric),

$$\hat{\chi}^2(\theta) = (c_\alpha(\theta) - \hat{c}_\alpha)\hat{\Psi}_{\alpha\beta}(c_\beta(\theta) - \hat{c}_\beta), \tag{B1}$$

assuming implicit index summation for brevity. The best-fit parameter vector, $\hat{\theta}$, is found by minimization,

$$\left.\frac{\partial\hat{\chi}^2}{\partial\theta_i}\right|_{\hat{\theta}} = 0 \Rightarrow \left.\frac{\partial c_\alpha}{\partial\theta_i}\right|_{\hat{\theta}} \hat{\Psi}_{\alpha\beta}(c_\beta(\hat{\theta}) - \hat{c}_\beta) = 0 \tag{B2}$$

for all i. Assuming the deviation from the true parameters, θ^*, to be small, we can Taylor expand to first order in $\delta\theta \equiv \hat{\theta} - \theta^*$,

$$\begin{aligned} c_\alpha(\hat{\theta}) &= c_\alpha(\theta^*) + \left.\frac{\partial c_\alpha}{\partial\theta_i}\right|_{\theta^*}\delta\theta_i, \\ \left.\frac{\partial c_\alpha}{\partial\theta_i}\right|_{\hat{\theta}} &= \left.\frac{\partial c_\alpha}{\partial\theta_i}\right|_{\theta^*} + \left.\frac{\partial^2 c_\alpha}{\partial\theta_i\partial\theta_j}\right|_{\theta^*}\delta\theta_j. \end{aligned} \tag{B3}$$

Inserting these into Eq. (B2) and simplifying, we obtain

$$\begin{aligned} &\left\{\frac{\partial c_\alpha}{\partial\theta_i}(\Psi_{\alpha\beta} + \delta\Psi_{\alpha\beta})\frac{\partial c_\beta}{\partial\theta_j} - \frac{\partial^2 c_\alpha}{\partial\theta_i\partial\theta_j}\Psi_{\alpha\beta}\hat{n}_\beta\right\}\delta\theta_j \\ &\quad = \frac{\partial c_\alpha}{\partial\theta_i}(\Psi_{\alpha\beta} + \delta\Psi_{\alpha\beta})\hat{n}_\beta, \end{aligned} \tag{B4}$$

implicitly assuming all terms to be evaluated at θ^*. A further simplification is obtained by noting that the term in curly parentheses in Eq. (B4) is simply the Fisher matrix of a noisy realization, $\hat{\mathcal{F}}_{ij} = \mathcal{F}_{ij} + \delta\mathcal{F}_{ij}$. The parameter shift is thus

$$\delta\theta_i = (\mathcal{F} + \delta\mathcal{F})^{-1}_{ij}\frac{\partial c_\alpha}{\partial\theta_j}(\Psi_{\alpha\beta} + \delta\Psi_{\alpha\beta})\hat{n}_\beta, \tag{B5}$$

agreeing with standard results [e.g., Eq. (23) of Ref. [2]]. To simplify this further, note that

$$(\mathcal{F} + \delta\mathcal{F})^{-1}_{ij} = \Phi_{ij} - \Phi_{ik}\delta\mathcal{F}_{kl}\Phi_{lj} \tag{B6}$$

at first order, where the $\Phi = \mathcal{F}^{-1}$ is the standard parameter covariance.

A number of well-known results are apparent from Eq. (B5): (a) the best-fit parameters are shifted by noisy data; (b) the amplitude of the shift is proportional to the parameter covariance; (c) noise on the precision matrix gives an additional shift in parameters, though is only present when the data is itself noisy. Note we have made no assumptions thus far on the number of SVs used in the α, β summation; thus the above results indicate that, for noise-free data, we will *never* have a bias from using too few SVs.

To quantify the extent of these parameter shifts, it is easiest to consider two regimes separately, first setting $\delta\Psi = 0$. The covariance of $\delta\theta$ is then

$$\begin{aligned} \langle\delta\theta_i|_{\delta\Psi=0}\delta\theta_j|_{\delta\Psi=0}\rangle &= \Phi_{ik}\Phi_{jl}\frac{\partial c_\alpha}{\partial\theta_k}\frac{\partial c_\beta}{\partial\theta_l}\langle\hat{n}_\alpha\hat{n}_\beta\rangle \\ &\equiv \Phi_{ik}\Phi_{jl}\mathcal{F}_{kl} = \Phi_{ij}, \end{aligned} \tag{B7}$$

using $\langle\hat{n}_\alpha\hat{n}_\beta\rangle = \mathcal{C}_{D,\alpha\beta} = \Psi^{-1}_{\alpha\beta}$ and ignoring noise contributions to $\delta\mathcal{F}_{ij}$. This is a standard result; the covariance in the parameter shift $\delta\theta$ is just the inverted Fisher matrix. Combining Eqs. (B5) and (B6), we find that the precision matrix noise induces an *extra* shift in the best-fit parameter vector

$$\begin{aligned} \Delta\theta_i &\equiv \delta\theta_i - \delta\theta_i|_{\delta\Psi=0} \\ &= \left[\Phi_{ij}\hat{n}_\beta\frac{\partial c_\alpha}{\partial\theta_j} - \Phi_{ik}\Phi_{jl}\Psi_{\gamma\delta}\hat{n}_\gamma\frac{\partial c_\alpha}{\partial\theta_k}\frac{\partial c_\beta}{\partial\theta_l}\frac{\partial c_\delta}{\partial\theta_j}\right]\delta\Psi_{\alpha\beta}. \end{aligned} \tag{B8}$$

When the error in the precision matrix arises from estimating the sample covariance with too few mocks, the covariance of the extra parameter shift can be shown to be

$$\langle\Delta\theta_i\Delta\theta_j\rangle = \frac{(N_{\rm mock} - N_{\rm SV})(N_{\rm SV} - N_{\rm param})}{(N_{\rm mock} - N_{\rm SV} - 1)(N_{\rm mock} - N_{\rm SV} - 4)}\Phi_{ij} \tag{B9}$$

[2], where $N_{\rm param}$ is the length of the parameter vector.[20] This just inflates the usual parameter estimate by a constant factor, which, for $N_{\rm mock} \gg N_{\rm SV}$ is well approximated by $(N_{\rm SV} - N_{\rm param})/N_{\rm mock}$.

APPENDIX C: ANALYTIC MARGINALIZATION OF THE POWER SPECTRUM LIKELIHOOD

For efficient sampling of high-dimensional parameter spaces, it is useful to marginalize over nuisance parameters analytically, as in Refs. [66,67]. While one may perform this approximately for any parameters, those which enter the likelihood quadratically can be marginalized exactly, given some choice of prior, assuming that the likelihood is Gaussian. Here, we consider the log-likelihood of Eq. (17) for parameters $\theta = \{\psi, \eta\}$, where nuisance parameters η

[20] Note that this includes the Hartlap factor required to invert the noisy covariance without bias.

enter the model $c(\theta)$ linearly. As shown in Ref. [66], marginalizing over η, assuming Gaussian priors, we obtain

$$\hat{\chi}^2_{D,M}(\psi) = \sum_{\alpha\beta}(\hat{c}_\alpha - c_\alpha(\psi))\mathcal{C}^{-1}_{D,M,\alpha\beta}(\psi)(\hat{c}_\beta - c_\beta(\psi)) + \log|\mathcal{C}_{D,M}(\psi)|, \tag{C1}$$

where the model is evaluated at the mean values of η. Now the covariance matrix inherits dependence on cosmology via

$$\mathcal{C}_{D,M,\alpha\beta}(\psi) = \mathcal{C}_{D,\alpha\beta} + \sum_{ij}\frac{\partial c_\alpha(\psi)}{\partial\eta_i}\frac{\partial c_\beta(\psi)}{\partial\eta_j}\mathsf{P}_{ij}, \tag{C2}$$

where P_{ij} is the prior parameter covariance (usually diagonal) and i, j run over the elements of η.

In our context, the counterterms $c_{s,0}$, $c_{s,2}$, b_4, and $P_{\rm shot}$ enter the power spectrum model linearly [and thus also the window convolved $P(k)$ and rotated $c(\theta)$ coefficients]. Analytic marginalization is possible via the methods above, using the broad Gaussian parameter priors of Ref. [47]. While the likelihood becomes more complex in this case (and requires additional window function derivatives for the parameter derivatives), it is still fast to compute and the parameter space can be reduced from ten to six elements.

In some scenarios, one has a parameter, say ζ, that is additionally constrained to be positive (for example, the shot noise, if an estimate has not already been subtracted). Analytic marginalization can still be performed in this case, though the expression for the log-likelihood is more complex,

$$\begin{aligned}\hat{\chi}^2_{D,M}(\psi)|_{\zeta>0} &= \sum_{\alpha\beta}(\hat{c}_\alpha - c_\alpha(\psi))\tilde{\mathcal{C}}_{D,M,\alpha\beta}(\psi)(\hat{c}_\beta - c_\beta(\psi)) \\ &+ \log|\tilde{\mathcal{C}}_{D,M}(\psi)| \\ &- 2\log\left[1 + \mathrm{erf}\left(\sqrt{\frac{J_1(\psi)+\sigma_\zeta^{-2}}{2}}\right)\left(\frac{J_2(\psi)}{J_1(\psi)+\sigma_\zeta^{-2}} + \bar{\zeta}\right)\right],\end{aligned} \tag{C3}$$

where $\tilde{C}_{D,M}$ is as Eq. (C2) including ζ in the nuisance parameters η and we define

$$\begin{aligned}J_1(\psi) &= \sum_{\alpha\beta}\frac{\partial c_\alpha(\psi)}{\partial\zeta}\frac{\partial c_\beta(\psi)}{\partial\zeta}\mathcal{C}^{-1}_{D,M,\alpha\beta}, \\ J_2(\psi) &= \sum_{\alpha\beta}\frac{\partial c_\alpha(\psi)}{\partial\zeta}\mathcal{C}^{-1}_{D,M,\alpha\beta}(\hat{c}_\beta - c_\beta(\psi)),\end{aligned} \tag{C4}$$

where $\bar{\zeta}$ and σ_ζ define the Gaussian prior on the positive parameter and $\mathcal{C}_{D,M}$ does *not* include ζ.

[1] 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**, 399 (2007).

[2] S. Dodelson and M. D. Schneider, The effect of covariance estimator error on cosmological parameter constraints, Phys. Rev. D **88**, 063537 (2013).

[3] A. Taylor and B. Joachimi, Estimating cosmological parameter covariance, Mon. Not. R. Astron. Soc. **442**, 2728 (2014).

[4] A. Taylor, B. Joachimi, and T. Kitching, Putting the precision in precision cosmology: How accurate should your data covariance matrix be?, Mon. Not. R. Astron. Soc. **432**, 1928 (2013).

[5] W. J. Percival, A. J. Ross, A. G. Sánchez, L. Samushia, A. Burden, R. Crittenden *et al.*, The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Including covariance matrix errors, Mon. Not. R. Astron. Soc. **439**, 2531 (2014).

[6] E. Sellentin and A. F. Heavens, Parameter inference with estimated covariance matrices, Mon. Not. R. Astron. Soc. **456**, L132 (2016).

[7] M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein *et al.*, The DESI Experiment, a whitepaper for Snowmass 2013, arXiv:1308.0847.

[8] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole *et al.*, Euclid definition study report, arXiv:1110.3193.

[9] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman *et al.*, LSST: From science drivers to reference design and anticipated data products, Astrophys. J. **873**, 111 (2019).

[10] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue *et al.*, Wide-field infrarred survey telescope-astrophysics focused telescope assets WFIRST-AFTA 2015 Report, arXiv:1503.03757.

[11] J. N. Grieb, A. G. Sánchez, S. Salazar-Albornoz, and C. Dalla Vecchia, Gaussian covariance matrices for anisotropic galaxy clustering measurements, Mon. Not. R. Astron. Soc. **457**, 1577 (2016).

[12] M. Lippich, A. G. Sánchez, M. Colavincenzo, E. Sefusatti, P. Monaco, L. Blot *et al.*, Comparing approximate methods for mock catalogues and covariance matrices—I. Correlation function, Mon. Not. R. Astron. Soc. **482**, 1786 (2019).

[13] 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. R. Astron. Soc. **497**, 1684 (2020).
[14] D. Wadekar and R. Scoccimarro, The galaxy power spectrum multipoles covariance in perturbation theory, Phys. Rev. D **102**, 123517 (2020).
[15] G. M. Bernstein, The variance of correlation function estimates, Astrophys. J. **424**, 569 (1994).
[16] R. O'Connell, D. Eisenstein, M. Vargas, S. Ho, and N. Padmanabhan, Large covariance matrices: Smooth models from the two-point correlation function, Mon. Not. R. Astron. Soc. **462**, 2681 (2016).
[17] R. O'Connell and D. J. Eisenstein, Large covariance matrices: Accurate models without mocks, Mon. Not. R. Astron. Soc. **487**, 2701 (2019).
[18] O. H. E. Philcox, D. J. Eisenstein, R. O'Connell, and A. Wiegand, RASCALC: A jackknife approach to estimating single- and multitracer galaxy covariance matrices, Mon. Not. R. Astron. Soc. **491**, 3290 (2020).
[19] O. H. E. Philcox and D. J. Eisenstein, Estimating covariance matrices for two- and three-point correlation function moments in Arbitrary Survey Geometries, Mon. Not. R. Astron. Soc. **490**, 5931 (2019).
[20] D. W. Pearson and L. Samushia, Estimating the power spectrum covariance matrix with fewer mock samples, Mon. Not. R. Astron. Soc. **457**, 993 (2016).
[21] M. Colavincenzo, E. Sefusatti, P. Monaco, L. Blot, M. Crocce, M. Lippich *et al.*, Comparing approximate methods for mock catalogues and covariance matrices—III: Bispectrum, Mon. Not. R. Astron. Soc. **482**, 4883 (2019).
[22] L. Blot, M. Crocce, E. Sefusatti, M. Lippich, A. G. Sánchez, M. Colavincenzo *et al.*, Comparing approximate methods for mock catalogues and covariance matrices II: Power spectrum multipoles, Mon. Not. R. Astron. Soc. **485**, 2806 (2019).
[23] D. J. Paz and A. G. Sánchez, Improving the precision matrix for precision cosmology, Mon. Not. R. Astron. Soc. **454**, 4326 (2015).
[24] B. Joachimi, Non-linear shrinkage estimation of large-scale structure covariance, Mon. Not. R. Astron. Soc. **466**, L83 (2017).
[25] N. Padmanabhan, M. White, H. H. Zhou, and R. O'Connell, Estimating sparse precision matrices, Mon. Not. R. Astron. Soc. **460**, 1567 (2016).
[26] A. Hall and A. Taylor, A Bayesian method for combining theoretical and simulated covariance matrices for large-scale structure surveys, Mon. Not. R. Astron. Soc. **483**, 189 (2019).
[27] C. Howlett and W. J. Percival, Galaxy two-point covariance matrix estimation for next generation surveys, Mon. Not. R. Astron. Soc. **472**, 4935 (2017).
[28] A. Klypin and F. Prada, Dark matter statistics for large galaxy catalogues: Power spectra and covariance matrices, Mon. Not. R. Astron. Soc. **478**, 4602 (2018).
[29] O. Friedrich and T. Eifler, Precision matrix expansion—efficient use of numerical simulations in estimating errors on cosmological parameters, Mon. Not. R. Astron. Soc. **473**, 4150 (2018).
[30] J. Roulet, L. Dai, T. Venumadhav, B. Zackay, and M. Zaldarriaga, Template bank for compact binary coalescence searches in gravitational wave data: A general geometric placement algorithm, Phys. Rev. D **99**, 123022 (2019).
[31] R. Scoccimarro, The bispectrum: From theory to observations, Astrophys. J. **544**, 597 (2000).
[32] J. R. Fergusson and E. P. S. Shellard, Shape of primordial non-Gaussianity and the CMB bispectrum, Phys. Rev. D **80**, 043510 (2009).
[33] J. R. Fergusson, M. Liguori, and E. P. S. Shellard, General CMB and primordial bispectrum estimation: Mode expansion, map making, and measures of F_{NL}, Phys. Rev. D **82**, 023502 (2010).
[34] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard, Rapid separable analysis of higher order correlators in large-scale structure, Phys. Rev. D **86**, 063511 (2012).
[35] M. M. Schmittfull, D. M. Regan, and E. P. S. Shellard, Fast estimation of gravitational and primordial bispectra in large scale structures, Phys. Rev. D **88**, 063512 (2013).
[36] P. Schneider, T. Eifler, and E. Krause, COSEBIs: Extracting the full E-/B-mode information from cosmic shear correlation functions, Astron. Astrophys. **520**, A116 (2010).
[37] T. Eifler, E. Krause, P. Schneider, and K. Honscheid, Combining probes of large-scale structure with COSMOLIKE, Mon. Not. R. Astron. Soc. **440**, 1379 (2014).
[38] A. F. Heavens, R. Jimenez, and O. Lahav, Massive lossless data compression and multiple parameter estimation from galaxy spectra, Mon. Not. R. Astron. Soc. **317**, 965 (2000).
[39] C. Reichardt, R. Jimenez, and A. F. Heavens, Recovering physical parameters from galaxy spectra using MOPED, Mon. Not. R. Astron. Soc. **327**, 849 (2001).
[40] A. F. Heavens, E. Sellentin, D. de Mijolla, and A. Vianello, Massive data compression for parameter-dependent covariance matrices, Mon. Not. R. Astron. Soc. **472**, 4244 (2017).
[41] A. F. Heavens, E. Sellentin, and A. H. Jaffe, Extreme data compression while searching for new physics, Mon. Not. R. Astron. Soc. **498**, 3440 (2020).
[42] M. Tegmark, A. N. Taylor, and A. F. Heavens, Karhunen-Loève eigenvalue problems in cosmology: How should we tackle large data sets?, Astrophys. J. **480**, 22 (1997).
[43] H. Prince and J. Dunkley, Data compression in cosmology: A compressed likelihood for Planck data, Phys. Rev. D **100**, 083502 (2019).
[44] D. Gualdi, M. Manera, B. Joachimi, and O. Lahav, Maximal compression of the redshift-space galaxy power spectrum and bispectrum, Mon. Not. R. Astron. Soc. **476**, 4045 (2018).
[45] J. Alsing and B. Wandelt, Generalized massive optimal data compression, Mon. Not. R. Astron. Soc. **476**, L60 (2018).
[46] T. Charnock, G. Lavaux, and B. D. Wandelt, Automatic physical inference with information maximizing neural networks, Phys. Rev. D **97**, 083004 (2018).
[47] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Cosmological parameters from the BOSS galaxy power spectrum, J. Cosmol. Astropart. Phys. 05 (2020) 042.
[48] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Cosmological parameters and neutrino masses from the final Planck and full-shape BOSS data, Phys. Rev. D **101**, 083504 (2020).

[49] O. H. E. Philcox, M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Combining full-shape and BAO analyses of galaxy power spectra: A 1.6% CMB-independent constraint on H_0, J. Cosmol. Astropart. Phys. 05 (2020) 032.
[50] A. Chudaykin, M. M. Ivanov, O. H. E. Philcox, and M. Simonović, Non-linear perturbation theory extension of the Boltzmann code CLASS, Phys. Rev. D **102**, 063533 (2020).
[51] T. Nishimichi, G. D'Amico, M. M. Ivanov, L. Senatore, M. Simonović, M. Takada, M. Zaldarriaga, and P. Zhang, Blinded challenge for precision cosmology with large-scale structure: Results from effective field theory for the redshift-space galaxy power spectrum, Phys. Rev. D **102**, 123541 (2020).
[52] O. H. E. Philcox, B. D. Sherwin, G. S. Farren, and E. J. Baxter, Determining the Hubble constant without the sound horizon: Measurements from galaxy surveys, arXiv: 2008.08084 [Phys. Rev. D (to be published)].
[53] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, Psychometrika **1**, 211 (1936).
[54] W. A. Fendt and B. D. Wandelt, Pico: Parameters for the Impatient Cosmologist, Astrophys. J. **654**, 2 (2007).
[55] I. Kayo, M. Takada, and B. Jain, Information content of weak lensing power spectrum and bispectrum: Including the non-Gaussian error covariance matrix, Mon. Not. R. Astron. Soc. **429**, 344 (2013).
[56] M. Tenenhaus, La régression PLS: Théorie et pratique, Editions technip, 1998.
[57] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek *et al.*, The clustering of galaxies in the completed SDSS-III baryon oscillation spectroscopic survey: Cosmological analysis of the DR12 galaxy sample, Mon. Not. R. Astron. Soc. **470**, 2617 (2017).
[58] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey *et al.*, The baryon oscillation spectroscopic survey of SDSS-III, Astron. J. **145**, 10 (2013).
[59] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson *et al.*, SDSS-III: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems, Astron. J. **142**, 72 (2011).
[60] https://fbeutler.github.io/hub/boss_papers.html.
[61] F. S. Kitaura, G. Yepes, and F. Prada, Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing., Mon. Not. R. Astron. Soc. **439**, L21 (2014).
[62] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín *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. R. Astron. Soc. **456**, 4156 (2016).
[63] S. A. Rodríguez-Torres, C.-H. Chuang, F. Prada, H. Guo, A. Klypin, P. Behroozi *et al.*, The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release, Mon. Not. R. Astron. Soc. **460**, 1173 (2016).
[64] B. Audren, J. Lesgourgues, K. Benabed, and S. Prunet, Conservative constraints on early cosmology with Monte Python, J. Cosmol. Astropart. Phys. 02 (2013) 001.
[65] T. Brinckmann and J. Lesgourgues, MontePython 3: Boosted MCMC sampler and other features, Phys. Rev. D **97**, 063506 (2018).
[66] A. N. Taylor and T. D. Kitching, Analytic methods for cosmological likelihoods, Mon. Not. R. Astron. Soc. **408**, 865 (2010).
[67] S. L. Bridle, R. Crittenden, A. Melchiorri, M. P. Hobson, R. Kneissl, and A. N. Lasenby, Analytic marginalization over CMB calibration and beam uncertainty, Mon. Not. R. Astron. Soc. **335**, 1193 (2002).
[68] L. Martino, V. Elvira, and F. Louzada, Effective Sample Size for Importance Sampling based on discrepancy measures, Signal Process. **131**, 386 (2017).
[69] F. Beutler, H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein *et al.*, The clustering of galaxies in the completed SDSS-III baryon oscillation spectroscopic survey: Anisotropic galaxy clustering in Fourier space, Mon. Not. R. Astron. Soc. **466**, 2242 (2017).
[70] D. Wadekar, M. M. Ivanov, and R. Scoccimarro, Cosmological constraints from BOSS with analytic covariance matrices, Phys. Rev. D **102**, 123521 (2020).
[71] H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres, and M. D. Olmstead, 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. R. Astron. Soc. **465**, 1757 (2017).
[72] A. Chudaykin and M. M. Ivanov, Measuring neutrino masses with large-scale structure: Euclid forecast with controlled theoretical error, J. Cosmol. Astropart. Phys. 11 (2019) 034.
[73] R. E. Smith, R. K. Sheth, and R. Scoccimarro, An analytic model for the bispectrum of galaxies in redshift space, Phys. Rev. D **78**, 023523 (2008).
[74] D. J. Eisenstein and M. Zaldarriaga, Correlations in the spatial power spectrum inferred from angular clustering: Methods and application to apm, Astrophys. J. **546**, 2 (2001).
[75] R. Scoccimarro, The bispectrum: From theory to observations, Astrophys. J. **544**, 597 (2000).
[76] E. Gaztanaga and R. Scoccimarro, The 3-point function in large scale structure: Redshift distortions and galaxy bias, Mon. Not. R. Astron. Soc. **361**, 824 (2005).
[77] http://www.sdss3.org/