

Reconstruction of the primordial power spectrum of curvature perturbations using multiple data sets

Paul Hunt^a & Subir Sarkar^{b,c}

^aInstitute of Theoretical Physics, Warsaw University, ul Hoża 69, 00-681 Warsaw, POLAND

^bRudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK

^cNiels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, DENMARK

E-mail: Paul.Hunt@fuw.edu.pl, s.sarkar@physics.ox.ac.uk

Abstract. Detailed knowledge of the primordial power spectrum of curvature perturbations is essential both in order to elucidate the physical mechanism (‘inflation’) which generated it, and for estimating the cosmological parameters from observations of the cosmic microwave background and large-scale structure. Hence it ought to be extracted from such data in a model-independent manner, however this is difficult because relevant cosmological observables are given by a convolution of the primordial perturbations with some smoothing kernel which depends on both the assumed world model and the matter content of the universe. Moreover the deconvolution problem is ill-conditioned so a regularisation scheme must be employed to control error propagation. We demonstrate that ‘Tikhonov regularisation’ can robustly reconstruct the primordial spectrum from multiple cosmological data sets, a significant advantage being that both its uncertainty and resolution are then quantified. Using Monte Carlo simulations we investigate several regularisation parameter selection methods and find that generalised cross-validation and Mallows’ C_p method give optimal results. We apply our inversion procedure to data from the Wilkinson Microwave Anisotropy Probe, other ground-based small angular scale CMB experiments, and the Sloan Digital Sky Survey. The reconstructed spectrum (assuming the standard Λ CDM cosmology) is *not* scale-free but has an infrared cutoff at $k \lesssim 5 \times 10^{-4} \text{ Mpc}^{-1}$ (due to the anomalously low CMB quadrupole) and several features with $\sim 2\sigma$ significance at $k/\text{Mpc}^{-1} \sim 0.0013\text{--}0.0025$, $0.0362\text{--}0.0402$ and $0.051\text{--}0.056$, reflecting the ‘WMAP glitches’. To test whether these are indeed real will require more accurate data, such as from the Planck satellite and new ground-based experiments.

Keywords: cosmic microwave background, cosmological parameters, cosmology: theory, large-scale structure of universe, inflation, primordial curvature perturbation

ArXiv ePrint: [1308.2317](https://arxiv.org/abs/1308.2317)

Contents

1	Introduction	2
2	Inversion method	5
2.1	Tikhonov regularisation	5
3	Results	9
3.1	Recovered spectra	9
3.2	Uncorrelated bandpowers	12
3.3	Background parameter errors	14
3.4	Statistical significance of the features	15
4	Conclusions	20
5	Acknowledgments	21
A	Error analysis	21
A.1	Bayesian inference	25
A.2	Performance statistics	26
B	Data	27
B.1	Experimental data sets	27
B.1.1	WMAP	27
B.1.2	Small angular scale CMB experiments	28
B.1.3	SDSS	29
B.2	Likelihood function derivatives	30
B.2.1	WMAP	30
B.2.2	Small angular scale CMB experiments	32
B.2.3	SDSS	32
B.3	Test spectra	33
B.4	Mock data	33
C	Validation of the inversion method	35
C.1	The integral and resolution kernels	35
C.2	Test spectra results	37
C.3	The bias and variance	39
C.4	The MSE and MPE	44
D	Choosing the regularisation parameter	47
D.1	Parameter selection methods	48
D.2	Application to test spectra	50
E	Glossary	55

1 Introduction

The primordial curvature perturbations which seeded the growth of large-scale structure (LSS) is presently our only observational window on the very early universe. Precision observations of anisotropies in the cosmic microwave background (CMB) by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite have greatly enhanced our knowledge of these primordial perturbations. The pattern of acoustic peaks in the CMB angular power spectra measured by WMAP, notably the anticorrelation in the temperature-electric (TE) polarisation spectrum on degree scales, indicate that the initial fluctuations are coherent, predominantly adiabatic and were generated on superhorizon scales [1, 2]. The perturbations also obey Gaussian statistics to a high level of accuracy [3]. Together this provides strong support for an early quasi-de Sitter phase of ‘inflation’, usually assumed to be driven by a scalar field with potential $V(\phi)$, as the origin of the primordial perturbations [4, 5]. However the gravitational dynamics of vacuum energy remains a complete mystery — the ‘Cosmological Constant problem’ [6, 7]. Hence it is essential to gain further insights concerning the underlying *physical* mechanism which generated the primordial curvature perturbations.

Cosmologists have so far attempted only to discriminate between various toy models of $V(\phi)$ on the basis of simple modelling of the primordial power spectrum (PPS) of curvature perturbations — $\mathcal{P}_{\mathcal{R}}(k)$ — usually as a featureless power-law spectrum. This has partly been because of practical limitations in extracting the PPS from CMB and LSS data (to be discussed below), and also because of the prejudice that ‘simple’ models of inflation predict just such a power-law form (with logarithmic corrections). However these are all severely fine-tuned ‘toy models’ so e.g. favouring $V = m^2\phi^2$ over $V = \lambda\phi^4$ [1] does not in fact provide any useful insight into how inflation actually occurred.¹ While detection of tensor perturbations (gravitational waves) would provide important new input, these may be observable through B -mode polarisation of the CMB only if the energy scale of inflation is in a narrow window around the Grand Unified scale of $\mathcal{O}(10^{16})$ GeV, which is certainly not mandatory [9, 10]. Hence scalar density perturbations are likely to remain the only available probe of the early universe, and we must therefore determine their properties in as much detail as is possible, in order to establish how they were generated.

Inhomogeneities after the primordial era depend upon both the initial fluctuations and their evolution. The latter is governed by the cosmological parameters describing the unperturbed background cosmology such as the baryon density Ω_b , the Hubble constant h , *etc.* Consequently it is difficult to disentangle the PPS and background parameters using observables such as CMB anisotropies and galaxy clustering which effectively only sample one time slice. For example, varying a PPS consisting of 75 wavenumber bins can mimic the effect of changes in the background parameters on the CMB temperature angular power spectrum to within the limit of cosmic variance [11]. Therefore the PPS must be assumed to have a simple form in order to avoid the degeneracy between the spectrum and the background parameters. It is usually taken to be a *featureless* power-law, $\mathcal{P}_{\mathcal{R}}(k) \propto k^{n_s-1}$, in the wavenumber k , with the spectral index n_s close to unity (the ‘Harrison-Zeldovich’ scale-invariant spectrum), as is expected in toy inflationary models based on monomial potentials e.g. $V(\phi) \propto \phi^n$ [4]. Then the WMAP data favour the so-called ‘concordance’ Λ CDM cosmology and constrain its parameters to within a few per cent [12]. However the results are very sensitive to the

¹Moreover, these are actually the *same* model in the sense that in the tiny range in ϕ space which corresponds to fluctuations on all observable scales, any monomial potential ϕ^n can be Taylor-expanded as $V_0 + \alpha\phi + \dots$, so both quartic $\lambda\phi^4$ and quadratic $m^2\phi^2$ potentials are in fact ‘linear inflation’ models [8].

assumed PPS. Indeed by incorporating power spectra with ‘bumps’, broken power-laws, or a strongly ‘running’ spectral index, an Einstein-de Sitter model with a $\sim 10\%$ hot dark matter component can still fit both CMB and LSS data [13–15]. Moreover all geometrical measures, i.e. based on an assumed (in this case Friedmann-Robertson-Walker) metric, such as the luminosity distance to Type Ia supernovae or the baryon acoustic oscillation angular scale can also be fitted in such models by adopting a different metric e.g. the radially inhomogeneous Lemaitre-Tolman-Bondi form [16, 17]. Given that the observational situation concerning *dynamical* evidence for dark energy such as the late integrated Sachs-Wolfe (ISW) effect is still confused (e.g. [18]), an accurate estimate of the PPS is clearly essential.

Knowledge of the PPS would also discriminate between models of inflation and probe physics at very high energy scales, particularly if the spectrum is distinctive. Single field slow-roll models predict values of the spectral index and its running $dn_s/d\ln k$ so these parameters can be constrained by fitting to the WMAP and other complementary data (see e.g. [12, 20, 21]). However there are physically better motivated models which produce fully adiabatic perturbations with *broken* scale-invariance: ‘steps’, ‘bumps’ and localised oscillations in the PPS can arise from interruptions to slow-roll evolution caused by phase transitions [22–24], resonant particle production [25–29] or features in the effective inflaton potential such as a ‘kink’ [30–32], ‘step’ [33–39] and others [40–52]. Such PPS features also occur in models with modulated preheating [53, 54], a limited duration of inflation [55–58] and other nonstandard scenarios [59–61]. Undamped PPS oscillations can be generated by trans-Planckian physics [62–66] and in other models that seek to incorporate Planck scale physics [67–72]. Clearly a model-*independent* method of recovering the PPS from observational data is required.

Interest in broken scale-invariant spectra was stimulated by the first WMAP data release which showed a lack of power on large scales in the temperature (TT) angular power spectrum, suggesting a cutoff in the PPS on the present horizon scale. Although the octupole moment was higher in the 3-year data release, the quadrupole is still unexpectedly low in the WMAP 9-year data and outliers or ‘glitches’ (exceeding the cosmic variance) persist around the $\ell = 22$ and $\ell = 40$ multipoles, despite the improved control of experimental systematics and the additional integration time. In addition, a possible anomaly around $\ell = 120$ has been identified [73, 74]. The overall χ^2 assuming a power-law spectrum equals 3336.4 for 3115 degrees of freedom, the probability to exceed this being only 0.3% (stated to be mainly due to the polarized likelihood [75]). Many models of inflation with broken scale-invariance have been compared to the low [76–85] and/or high [86–96] multipole anomalies.

Since the PPS is a free function it must be given a parameterisation. In the model-independent approach the parameterisation is designed to be as general as possible and to fit all conceivable forms of the PPS. Model-independent methods of estimating the PPS fall into two classes, depending on the number of power spectrum parameters. In parametric methods the number of parameters, which is typically around 20 but can be up to 50 [97], is much less than the number of data points. The PPS has been described using wavelets [98–101], principal components [102] and smoothing splines [103–107], in addition to bins in wavenumber with no interpolation (i.e. ‘tophat’ bins) [108], linear interpolation [109–116], cubic spline interpolation [97, 118–121] and power-law bins [122]. The small number of parameters means that accurate confidence limits for the cosmological parameters can be set using Markov Chain Monte Carlo (MCMC) analysis to draw samples from the posterior distribution of the parameters given the data. However from the outset a filter is effectively imposed on the recovered PPS by the choice of parameterisation, which limits the resolution with which the PPS can be recovered. For example, when the spectrum is parameterised

using wavenumber bins or smoothed with a spline, this will obviously miss features in the PPS narrower than the bin width or the spine scale.

In non-parametric methods the number of parameters describing the power spectrum is comparable to, and often greater than, the number of data points. Since the CMB angular power spectrum is given by a convolution of the PPS with a radiative transport kernel, this approach is essentially an exercise in deconvolution. Recovery of the PPS by deconvoluting the CMB angular power spectrum was first discussed in [123] but only the Sachs-Wolfe effect was considered. In the ‘cosmic inversion’ method [124–129] a differential equation for the PPS derived from cosmological perturbation theory was solved iteratively. Regularisation schemes can be employed in non-parametric methods which act as variable filters, governed by a regularisation parameter λ which can be tuned to optimally extract the PPS in the presence of noise. Regularisation methods include truncated singular value decomposition [130], Richardson-Lucy iteration [131–138] and maximum entropy deconvolution [139]. Tikhonov regularisation was previously adopted by e.g. [73, 140–143].

We too adopt Tikhonov regularisation because it has notable advantages over other non-parametric techniques. There is a relatively simple relationship between the input data and the output estimated PPS (unlike in iterative regularisation methods) which means the results are easy to interpret. Furthermore Tikhonov regularisation is rapid and computationally inexpensive, so that extensive testing of the method on simulated data is feasible.

We improve over previous work in several significant respects. Non-parametric techniques have mainly been applied to CMB TT power spectra data alone (with the exception of [127, 128, 130, 134] where CMB polarisation data was also used). By using additional data sets it should be possible to recover the PPS over a wider wavenumber range with increased accuracy. In the pre-WMAP era, separate estimates of the PPS were produced from a number of data sets including non-CMB data [140]. Here we extend the Tikhonov regularisation method and show how it can be used to obtain a *single* combined high-resolution estimate of the PPS from multiple data sets. We illustrate the method using TT and TE data from WMAP and other CMB experiments, in addition to measurements of the clustering of luminous red galaxies (LRG) in the Sloan Digital Sky Survey (SDSS).

Simply constructing an estimate of the PPS is insufficient; it is equally important to understand the relationship of the estimate to the true PPS.² The estimate will inevitably differ from the true PPS for at least three reasons. The first is that the data are contaminated by noise. We quantify the effects of error propagation on our recovered PPS using Monte Carlo simulation in addition to estimating both Bayesian and frequentist covariance matrices. The second is that the background cosmological model parameters used to calculate the convolution kernels are uncertain. While the effect on the reconstructed spectrum of varying these has been studied earlier [127, 130], we quantify the propagated error from this effect for the first time, using covariance matrices. The third reason is that a finite number of data points contains only a limited amount of information about the true PPS. In practice this means the recovered PPS has limited resolution — it would be unable to distinguish fine features in the true PPS, even with noise-free data. Also for the first time we quantify the resolution of the estimated PPS, using the so-called ‘resolution kernel’.

Nearly all non-parametric inversion algorithms, including Tikhonov regularisation, contain at least one implicit or explicit adjustable regularisation parameter which depends on the signal-to-noise ratio of the data and whose value must be chosen correctly for optimal

²This is known as the ‘appraisal’ of the estimate in the inverse theory literature [144].

results.³ In all previous work the parameter values have either been set by hand or chosen using techniques untested on simulated cosmological data. We pay particular attention to this issue, and study several different objective methods for selecting the Tikhonov regularisation parameter using Monte Carlo simulation.

Almost all of the formal mathematical theory developed on Tikhonov regularisation concerns its application to a single data set, usually with uncorrelated Gaussian errors (see e.g. [145]). However CMB data sets usually have *correlated non*-Gaussian errors. Moreover, the noise distributions at low multipoles depend on the true PPS due to cosmic variance. While Tikhonov regularisation can readily be adapted to multiple data sets with complicated noise properties, it leads to some novel results not previously reported in the literature.

The plan of the paper is as follows. In Section 2 we describe our method, referring to Appendix A where we discuss frequentist error analysis and comment on how it can be adapted to Bayesian inference (A.1). Appendix B presents the data used — from WMAP (B.1.1), ground-based CMB experiments (B.1.2) and SDSS (B.1.3) — and the corresponding likelihood function derivatives (B.2). In Appendix C, we examine the performance of our method using mock data (B.4) for several inflationary model test spectra (B.3) and study how the value of the regularisation parameter affects the inversion. In Appendix D, methods of choosing the optimum regularisation parameter are investigated. We present our main results in Section 3. In addition to the estimated PPS (3.1) we include analysis of the uncorrelated bandpowers (3.2), uncertainties in the extracted cosmological parameters (3.3) and the statistical significance of the PPS features (3.4). In Section 4 we present our conclusions. A glossary of symbols used is provided in Appendix E.

2 Inversion method

2.1 Tikhonov regularisation

The two-power correlation function of the primordial comoving curvature perturbation \mathcal{R} is related to the PPS through:⁴

$$\langle \mathcal{R}^*(\mathbf{k}) \mathcal{R}(\mathbf{k}') \rangle = \delta^3(\mathbf{k} - \mathbf{k}') \frac{2\pi^2}{k^3} \mathcal{P}_{\mathcal{R}}(k). \quad (2.1)$$

Let us assume there are N available cosmological data sets which probe the PPS. The data points are denoted $d_a^{(\mathbb{Z})}$ where the superscript labels the data set (\mathbb{Z}) and the subscript a runs from 1 up to the number $N_{\mathbb{Z}}$ of points in the set.⁵ The data points extracted from measurements of CMB anisotropy, galaxy clustering, Lyman- α forest, cluster abundance and weak lensing obey a first-order integral equation of the form [140]:

$$d_a^{(\mathbb{Z})} = \int_0^\infty \mathcal{K}_a^{(\mathbb{Z})}(\boldsymbol{\theta}, k) \mathcal{P}_{\mathcal{R}}(k) dk + n_a^{(\mathbb{Z})}, \quad (2.2)$$

where the integral kernels $\mathcal{K}_a^{(\mathbb{Z})}$ depend on the background parameters $\boldsymbol{\theta}$, and the noise vectors $\mathbf{n}_{\mathbb{Z}}$ have zero mean and covariance matrices $\mathbf{N}_{\mathbb{Z}} \equiv \langle \mathbf{n}_{\mathbb{Z}} \mathbf{n}_{\mathbb{Z}}^T \rangle$, where T denotes the transpose

³For example, in truncated singular value decomposition and Richardson-Lucy iteration the regularisation parameter is identified with the truncation level and the iteration count respectively.

⁴ Our notation/definition of the PPS is that of [5]. In [3] the PPS is denoted by $\Delta_{\mathcal{R}}^2(k)$.

⁵ For ease of notation a data set superscript is promoted to a subscript when labelling an entire vector or matrix rather than one of its elements e.g. the vector $d_a^{(\mathbb{Z})}$ is simply $\mathbf{d}_{\mathbb{Z}}$ and the matrix $W_{ai}^{(\mathbb{Z})}$ is simply $W_{\mathbb{Z}}$.

matrix. Eq.(2.2) is quite general: it holds on large scales where *linear* cosmological perturbation theory is valid and also applies in the nonlinear regime if a suitable transformation or linearisation can be performed. We assume an estimate $\hat{\boldsymbol{\theta}}$ of the background parameters exists which is *independent* of the N data sets, and has a zero mean uncertainty \mathbf{u} , with elements u_α . Then $\langle u_\alpha n_a^{(\mathbb{Z})} \rangle = 0$ for all elements of the uncertainty and noise vectors as these are uncorrelated by assumption. The covariance matrix for the estimated background parameters is just $\mathbf{U} \equiv \langle \mathbf{u}\mathbf{u}^T \rangle$. Given our estimate of the background parameter set $\hat{\boldsymbol{\theta}}$, the goal is to obtain an estimate $\hat{\mathcal{P}}_{\mathcal{R}}(k)$ of the PPS from the data sets $\mathbf{d}_{\mathbb{Z}}$.

This is an example of an inverse problem (as opposed to the forward problem of calculating the expected data for a given PPS) which generally involve the determination of some unknown model input from indirect measurements. Inverse theory concerns both parameter and function estimation. A ‘well-posed’ problem is defined [146] as having a unique solution which depends continuously on the data, *i.e.* as the error in the data tends to zero, the induced error in the solution also tends to zero. If this is not so, the problem is said to be ‘ill-posed’. Inverse problems are usually ill-posed, and the reconstruction of the PPS is no exception. Ill-posed inverse problems occur also for example in medical and geophysical imaging, remote sensing and image restoration, and an extensive literature is devoted to their solution (for a textbook introduction see e.g. [147, 148]). In cosmology, there have been applications in e.g. gravitational wave detection [149] and in cosmography [150], as well as of course in the present context [140–142].

For any finite number of data points there is an infinite-dimensional null space of functions $\mathcal{N}_{\mathcal{R}}(k)$ such that

$$0 = \int_0^\infty \mathcal{K}_a^{(\mathbb{Z})}(\boldsymbol{\theta}, k) \mathcal{N}_{\mathcal{R}}(k) dk \quad (2.3)$$

for all of the integral kernels [151]. The available data cannot distinguish between two power spectra that differ by a member of the null space (e.g. because of finite resolution or even because the difference would be obvious only where there is no data). Clearly some additional criterion must be used to narrow down the infinite number of power spectra consistent with the data, the vast majority of which may well be unphysical. From our present (admittedly crude) understanding of inflation the PPS is expected to be smooth in some sense with little fine detail. Hence we assume that the simplest solution is the preferred one *i.e.* the optimal estimate of the PPS is the *smoothest* of the spectra matching the data. While other spectra may fit the data equally well, they will inevitably be more complex and likely possess misleading features. Thus our prior knowledge (prejudice) concerning the PPS is combined with the information contained in the data to obtain the optimal estimate.

We approximate the PPS as a piecewise function given by a sum of N_j basis functions $\phi_i(k)$, weighted by coefficients p_i giving the power in each bin:

$$\mathcal{P}_{\mathcal{R}}(k) = \sum_{i=1}^{N_j} p_i \phi_i(k). \quad (2.4)$$

The basis functions are defined using a grid of wavenumbers $\{k_i\}$ as

$$\phi_i(k) \equiv \begin{cases} 1, & k_i < k \leq k_{i+1}, \\ 0, & \text{elsewhere.} \end{cases} \quad (2.5)$$

We set $N_j = 2000$ and use a logarithmically spaced grid between $k_1 = 7 \times 10^{-6} \text{ Mpc}^{-1}$ and $k_{N_j+1} = 0.7 \text{ Mpc}^{-1}$. Substituting eq.(2.4) into eq.(2.2) yields

$$\mathbf{d}_a^{(\mathbb{Z})} = \sum_i W_{ai}^{(\mathbb{Z})}(\boldsymbol{\theta}) p_i + \mathbf{n}_a^{(\mathbb{Z})}, \quad (2.6)$$

where the $N_{\mathbb{Z}} \times N_j$ matrices $W_{ai}^{(\mathbb{Z})}(\boldsymbol{\theta})$ ($W_{\mathbb{Z}}$ for short) depend on the background parameters:

$$W_{ai}^{(\mathbb{Z})}(\boldsymbol{\theta}) = \int_{k_i}^{k_{i+1}} \mathcal{K}_a^{(\mathbb{Z})}(\boldsymbol{\theta}, k) dk. \quad (2.7)$$

Vectors of length $N_d \equiv \sum_{\mathbb{Z}} N_{\mathbb{Z}}$ assembled from the individual data and noise vectors are denoted by $\mathbf{d} \equiv (\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_N^T)^T$ and $\mathbf{n} \equiv (\mathbf{n}_1^T, \mathbf{n}_2^T, \dots, \mathbf{n}_N^T)^T$ respectively. We also define the $N_d \times N_j$ matrix $\mathbf{W} \equiv (\mathbf{W}_1^T, \mathbf{W}_2^T, \dots, \mathbf{W}_N^T)^T$ so that

$$\mathbf{d} = \mathbf{W}(\boldsymbol{\theta}) \mathbf{p} + \mathbf{n}, \quad (2.8)$$

where \mathbf{p} is the vector with elements p_i defining the PPS.

Standard methods of statistical inference such as maximum likelihood analysis perform poorly when applied to ill-posed inverse problems. Instead, specialist techniques must be employed. To illustrate the difficulties involved, we use a superficially appealing estimate of the PPS labelled $\hat{\mathbf{p}}_0$ defined in the following manner. Recall that the rank of a matrix is equal to the number of nonzero singular values, and is also equal to the number of linearly independent column and row vectors. If the rank ρ of \mathbf{W} is equal to N_j then $\hat{\mathbf{p}}_0$ is the unique maximum of the likelihood function of the data given the PPS and the background parameters. In general the matrix \mathbf{W} has a $(N_j - \rho)$ -dimensional null space of vectors \mathbf{v}_{nl} such that $\mathbf{W}\mathbf{v}_{\text{nl}} = 0$. Hence if $\rho < N_j$ there are infinitely many solutions which maximise the likelihood and differ by members of the null space, so that the likelihood function has no well-defined peak. We introduce a function $R(\mathbf{p})$ which quantifies the ‘roughness’ of \mathbf{p} . Then if $\rho < N_j$ the estimate $\hat{\mathbf{p}}_0$ is taken as the vector with the minimum R value (i.e. the ‘smoothest’) of those that maximise the likelihood.

For multiple data sets with Gaussian noise the likelihood function $\mathcal{L}(\mathbf{p}, \boldsymbol{\theta} | \mathbf{d})$ is given by

$$-2 \ln \mathcal{L}(\mathbf{p}, \boldsymbol{\theta} | \mathbf{d}) = [\mathbf{W}(\boldsymbol{\theta}) \mathbf{p} - \mathbf{d}]^T \mathbf{N}^{-1} [\mathbf{W}(\boldsymbol{\theta}) \mathbf{p} - \mathbf{d}] \quad (2.9)$$

up to an unimportant normalisation constant, where $\mathbf{N} \equiv \text{diag}(\mathbf{N}_1, \mathbf{N}_2, \dots, \mathbf{N}_N)$ is a $N_d \times N_d$ block diagonal matrix assembled from the individual data covariance matrices. It is diagonal because the individual data sets are assumed to be *independent*. If in addition the roughness function has the quadratic form $R(\mathbf{p}) = \mathbf{p}^T \boldsymbol{\Gamma} \mathbf{p}$ where $\boldsymbol{\Gamma}$ is a positive definite matrix, then $\hat{\mathbf{p}}_0 = \mathbf{W}_{\mathbf{N}\boldsymbol{\Gamma}}^\dagger \mathbf{d}$ for $\rho \leq N_j$. Here $\mathbf{W}_{\mathbf{N}\boldsymbol{\Gamma}}^\dagger$ is the weighted Moore-Penrose inverse of \mathbf{W} , also known as the weighted generalised inverse [152, 153] and can be calculated using the weighted singular value decomposition of \mathbf{W} developed in [154]. This is written in terms of the $N_d \times N_d$ matrix \mathbf{J} and the $N_j \times N_j$ matrix \mathbf{K} which satisfy $\mathbf{J}^T \mathbf{N} \mathbf{J} = \mathbf{I}$ and $\mathbf{K}^T \boldsymbol{\Gamma}^{-1} \mathbf{K} = \mathbf{I}$ and has the form

$$\mathbf{W} = \mathbf{J} \begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{K}^T, \quad (2.10)$$

where $\boldsymbol{\Lambda} \equiv \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_\rho)$ with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_\rho > 0$. The $\{\sigma_i\}$ are called the weighted singular values of \mathbf{W} and are just the square-roots of the nonzero eigenvalues of

$\Gamma^{-1}\mathbf{W}^T\mathbf{N}\mathbf{W}$. The weighted Moore-Penrose inverse is then given by:

$$\mathbf{W}_{\mathbf{N}\Gamma}^\dagger = \Gamma^{-1}\mathbf{K} \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{J}^T\mathbf{N}. \quad (2.11)$$

The significance of this will be discussed shortly. Note that the ordinary Moore-Penrose inverse \mathbf{W}^\dagger is recovered when \mathbf{N} and Γ are both equal to the identity matrix.

Unfortunately $\hat{\mathbf{p}}_0$ is ‘ill-conditioned’, i.e. excessively sensitive to small perturbations of the data. This is because convolving the PPS with an integral kernel acts as a smoothing operation, due to the finite width of the kernel. High frequency components of the PPS are damped more than low frequency ones. Conversely, high frequency noise is amplified to produce wild excursions in $\hat{\mathbf{p}}_0$, rendering the estimate useless. Roughly speaking, the condition number of a system measures how it amplifies errors. The condition number of the weighted Moore-Penrose is σ_1/σ_ρ , which indicates that the ill-conditioning of $\hat{\mathbf{p}}_0$ is associated with \mathbf{W} possessing small weighted singular values, as can be seen from eq.(2.11).

So-called regularisation schemes are designed to produce a well-behaved approximate solution to an ill-posed inverse problem. They effectively filter out the troublesome small singular values of \mathbf{W} . Regularisation reduces the propagation of error from the data to the solution at the cost of inducing a bias; the former is intended to offset the latter. Since the optimum solution depends on the level of noise in the data, a regularisation scheme actually produces a family of solutions characterised by a ‘regularisation parameter’. Obtaining the optimum solution then amounts to choosing the appropriate value of the regularisation parameter according to the noise level. Common regularisation schemes include spectral cut-off methods such as truncated singular value decomposition, iterative methods such as Landweber iteration, and the Tikhonov regularisation scheme — which we use in this work.

Tikhonov regularisation is based upon the notion of a ‘correctness set’ [155]. For an inverse problem this is a subset of solution space (i.e. the space of all possible solutions) known *a priori* to contain the true solution. Moreover, the solution is continuous for data such that the solution remains in the correctness set.⁶ Since the PPS is believed to be smooth, there is effectively an upper bound R_0 on its roughness: spectra rougher than R_0 may be regarded as physically implausible. The compact set of spectra which satisfy $R(\mathbf{p}) \leq R_0$ form a natural correctness set. The estimated PPS is chosen as the member of the correctness set which *maximises* the likelihood function. In practice the roughest elements of the correctness set give the best fit to the data and so the estimate satisfies $R(\hat{\mathbf{p}}) = R_0$. Therefore to obtain $\hat{\mathbf{p}}$ we maximise the likelihood subject to this constraint. This is equivalent to the more convenient unconstrained minimisation of the quantity

$$Q(\mathbf{p}, \mathbf{d}, \hat{\boldsymbol{\theta}}, \lambda) \equiv L(\mathbf{p}, \hat{\boldsymbol{\theta}}, \mathbf{d}) + \lambda R(\mathbf{p}), \quad (2.12)$$

where $L(\mathbf{p}, \boldsymbol{\theta}, \mathbf{d}) \equiv -2 \ln \mathcal{L}(\mathbf{p}, \boldsymbol{\theta} | \mathbf{d})$ and the regularisation parameter λ plays the role of a Lagrange multiplier. Thus the estimated PPS is given by

$$\hat{\mathbf{p}}(\mathbf{d}, \hat{\boldsymbol{\theta}}, \lambda) = \arg \min_{\mathbf{p}} Q(\mathbf{p}, \mathbf{d}, \hat{\boldsymbol{\theta}}, \lambda). \quad (2.13)$$

For a quadratic roughness function, the estimate $\hat{\mathbf{p}}$ is a unique continuous function of the data. It is defined as the best-fitting spectrum for a particular roughness value, but it can equally well be thought of as the *smoothest* spectrum for a given value of the likelihood.

⁶Indeed regularisation schemes need to incorporate *some* prior expectation concerning the solution in order to overcome the generally ‘ill-posed’ nature of the problem (as is obvious in e.g. applications to remote geophysical sensing or medical imaging where the physically plausible answer is easily recognised) [147, 148].

The regularisation parameter thus controls the trade-off between the smoothness of the estimated PPS and its fidelity to the data. Increasing λ reduces the correctness set so that $\hat{\mathbf{p}}$ becomes smoother and less spiky but fits the data less well. As λ tends to zero, $\hat{\mathbf{p}}$ tends to the smoothest spectrum which maximises the likelihood, viz. the ill-conditioned $\hat{\mathbf{p}}_0$. If the noise level of the data and the error in the estimated background parameters also tend to zero, then $\hat{\mathbf{p}}$ approaches the true PPS. In the opposite limit $\hat{\mathbf{p}}$ approaches the best-fitting spectrum which minimises R as λ tends to infinity. Thus we attempt to find a smooth PPS which provides a reasonable fit to the data, in a similar manner to nonparametric regression.

While the regularisation parameter determines the *amount* by which $\hat{\mathbf{p}}$ is smoothed, the roughness function governs the *manner* in which $\hat{\mathbf{p}}$ is smoothed. In this work we employ the roughness function $R(\mathbf{p}) = \mathbf{p}^T \Gamma \mathbf{p}$, where

$$\Gamma \equiv \mathbf{L}^T \mathbf{L} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}, \quad (2.14)$$

and \mathbf{L} is a discrete version of the first-order derivative operator. For a logarithmically spaced wavenumber grid, R represents a discretisation of the integral of $(d\mathcal{P}_{\mathcal{R}}/d \ln k)^2$ over $\ln k$. This choice of roughness function is known as first-order Tikhonov regularisation [156]. Similar roughness functions involving first-order derivatives were used previously in [141, 142]. Any vector with equal components minimises R , corresponding to a Harrison-Zeldovich (H-Z) PPS with $n_s = 1$. Note that for wavenumbers where the corresponding data is incomplete or does not constrain the PPS, $\hat{\mathbf{p}}$ is determined mainly by the roughness function. Hence with the above choice of the roughness function we ensure that the recovered PPS is generally positive. This would not be the case if we chose instead a second-order derivative operator corresponding to a ‘tilted’ spectrum with $n_s \lesssim 1$ (even though this is in fact the best-fit to data for a Λ CDM cosmology). Note that the procedure used in [140] is equivalent to setting Γ equal to the identity matrix, which corresponds to zeroth-order Tikhonov regularisation.

We minimise Q using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm — a quasi-Newton method [157]. In the usual Newton-Raphson method the function to be minimised is assumed to be quadratic near its minimum and at each iteration a better approximation for the minimum is obtained by solving a linear system of equations involving the gradient vector and the Hessian matrix of second partial derivatives of Q . The BFGS algorithm avoids the computationally expensive evaluation of the Hessian; instead an accurate approximation to the inverse Hessian is built up using changes in the gradient vector between successive iterations, in a multidimensional generalisation of the secant method. Expressions for the gradients of the various likelihood functions are listed in Appendix B.2, together with the initial guesses for the Hessian.

3 Results

In Appendix C we validate our inversion method on simulated data. Below we apply it to real cosmological data.

3.1 Recovered spectra

We refer to 4 data set combinations of CMB and LSS data as follows:

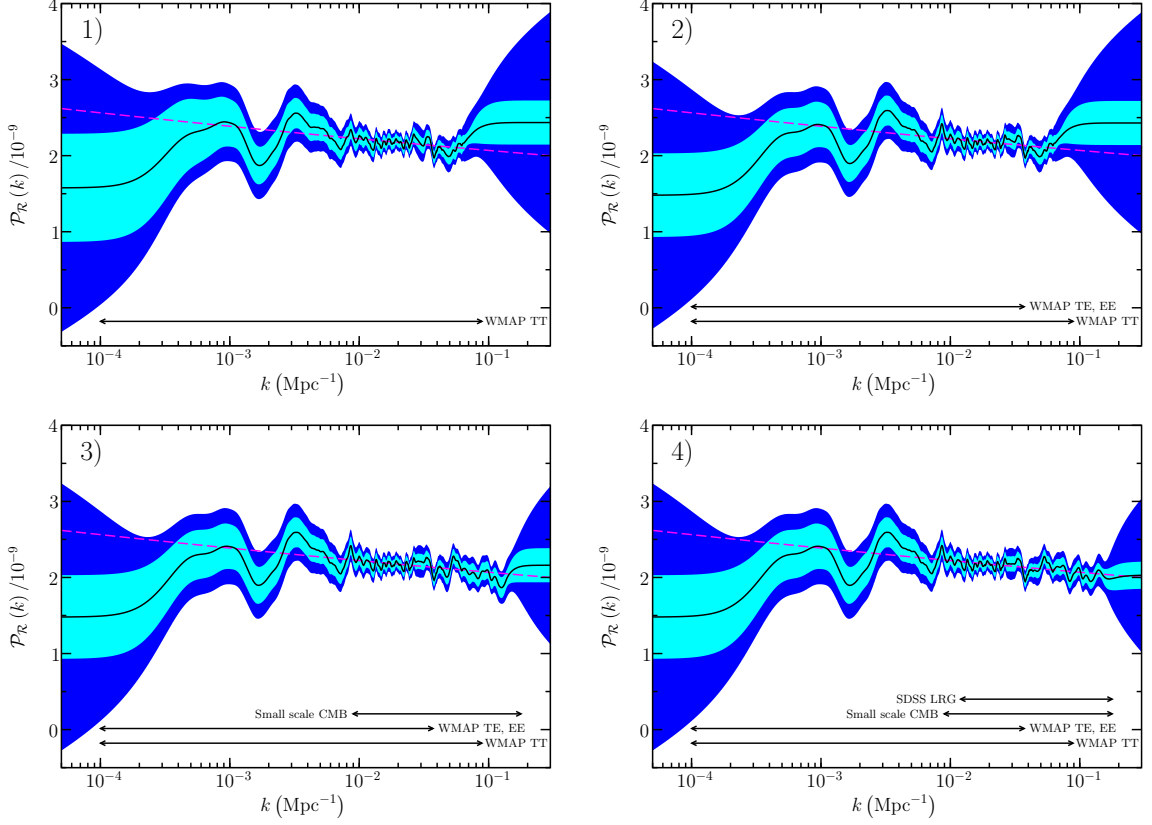


Figure 1. The PPS recovered with $\lambda = 100$ from the 1) WMAP-9 temperature data alone, 2) WMAP-9 temperature and polarisation data, 3) WMAP-9 plus small angular scale CMB data, 4) WMAP-9, small-scale CMB and SDSS-4 LRG data. In each plot the central black line is the reconstructed PPS and the dark band is the 1σ error given by the square root of the diagonal elements of the Bayesian covariance matrix Σ_B (eq. A.28), while the overlaying light band is similarly obtained from the frequentist covariance matrix Σ_F (eq. A.10). The dashed line is our ‘best-fit’ power-law spectrum with slope $n_s = 0.969$ assuming the standard Λ CDM model parameters. Also indicated are the approximate wavenumber ranges over which the different datasets have the most impact.

- Data combination 1: WMAP TT
- Data combination 2: WMAP all (TT + TE + EE)
- Data combination 3: All CMB (WMAP all + ground-based: ACBAR + BOOMERanG (TT + TE + EE) + CBI + QUAD (TT + TE + EE) + VSA)
- Data combination 4: All CMB + SDSS-4 LRG

Obviously these can be updated with data from other experiments as and when necessary.

We adopt typical parameter values for the standard Λ CDM model: $\omega_b \equiv \Omega_b h^2 = 0.0224$, $\omega_c \equiv \Omega_c h^2 = 0.102$, $h = 0.73$, optical depth to last scattering $\tau = 0.095$, and bias of luminous red galaxies $b_{\text{LRG}} = 1.9$ [179]. This model is spatially flat, hence $\Omega_\Lambda \equiv 1 - \Omega_b - \Omega_c$.

As shown in Appendix D.2, if the PPS is an exact power-law then deconvoluting with $\lambda \simeq 5000$ would be most likely to minimise the squared-error (Panel A, Fig. 22). However if the PPS has features then a lower value like $\lambda \simeq 100$ would be more appropriate (Panels

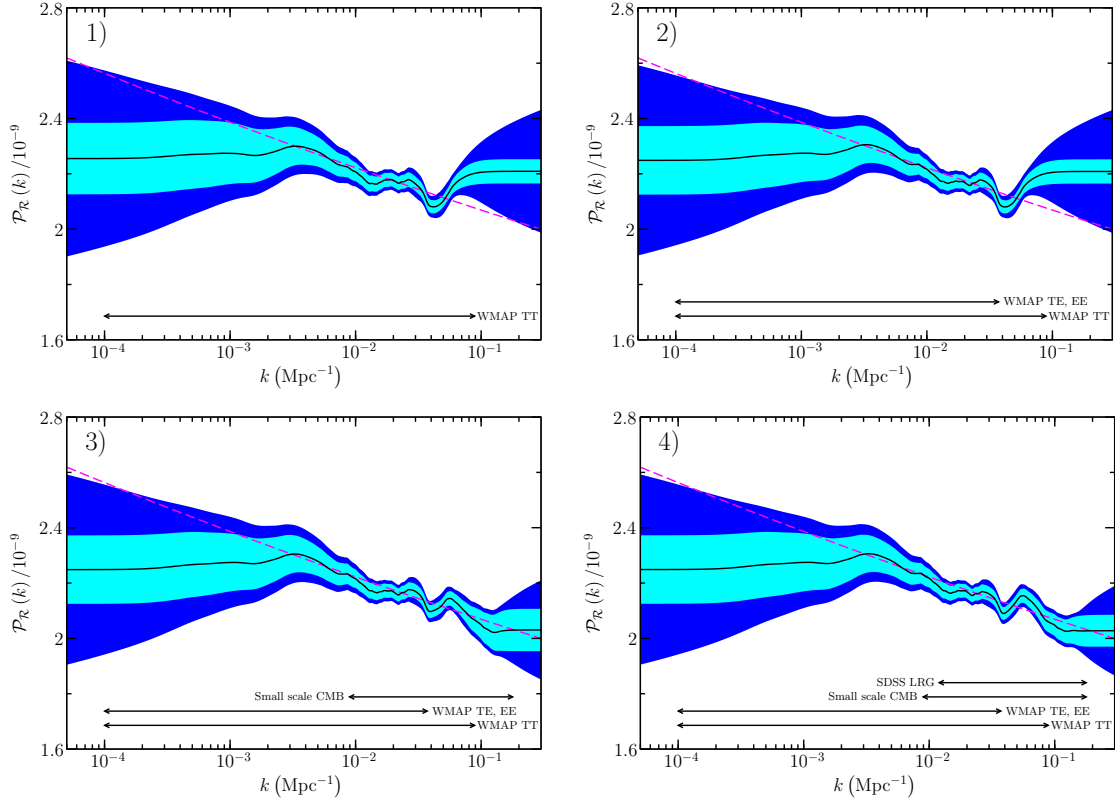


Figure 2. Same as Fig. 1, but for inversions with $\lambda = 5000$.

B and C, Fig. 22). Therefore we perform reconstructions with both $\lambda = 100$ (Fig. 1) and $\lambda = 5000$ (Fig. 2) but are not able to advocate any particular value of λ in this range.

The recovered spectra oscillate about a power-law with $n_s = 0.969$, consistent with the spectral index found by the WMAP team. However, there are several interesting features. The $\lambda = 100$ spectra have the well-known infrared cutoff from the low WMAP-9 TT quadrupole. There are also ‘bumps’ at $k \simeq 0.0032, 0.0086$ and 0.033 Mpc^{-1} due to the excess power in the WMAP-9 temperature angular power spectrum around the $\ell \simeq 40, 117$ and 470 multipoles. Similarly ‘dips’ at $k \simeq 0.0017, 0.0072, 0.013, 0.015$ and 0.038 Mpc^{-1} are caused by the lack of power around the $\ell \simeq 24, 95, 181, 209$ and 540 multipoles. Adding the WMAP-9 polarisation data slightly increases the infrared suppression of the reconstructed PPS, boosts the bump at $k \simeq 0.0086 \text{ Mpc}^{-1}$ and reduces the error bands on large scales. Including the small-scale CMB data improves the reconstruction for $k \gtrsim 0.05 \text{ Mpc}^{-1}$. The small scale observations at $\ell \simeq 750$ create an additional bump at $k \simeq 0.053 \text{ Mpc}^{-1}$. Adding the SDSS-4 LRG data further sharpens the estimated PPS on small scales and introduces further dips at $k \simeq 0.084 \text{ Mpc}^{-1}$ and $k \simeq 0.12 \text{ Mpc}^{-1}$. For $\lambda = 5000$ however much of the structure in the $\lambda = 100$ reconstruction is smoothed away, particularly on large scales where the infrared cutoff disappears — this is due to our prior of a H-Z spectrum combined with the increased cosmic variance on large scales. Prominent features remaining in the PPS estimated from the WMAP-9, small-scale CMB and SDSS-4 LRG data include a peak at $k \simeq 0.055 \text{ Mpc}^{-1}$ and dips at $k \simeq 0.014 \text{ Mpc}^{-1}$ and $k \simeq 0.038 \text{ Mpc}^{-1}$.

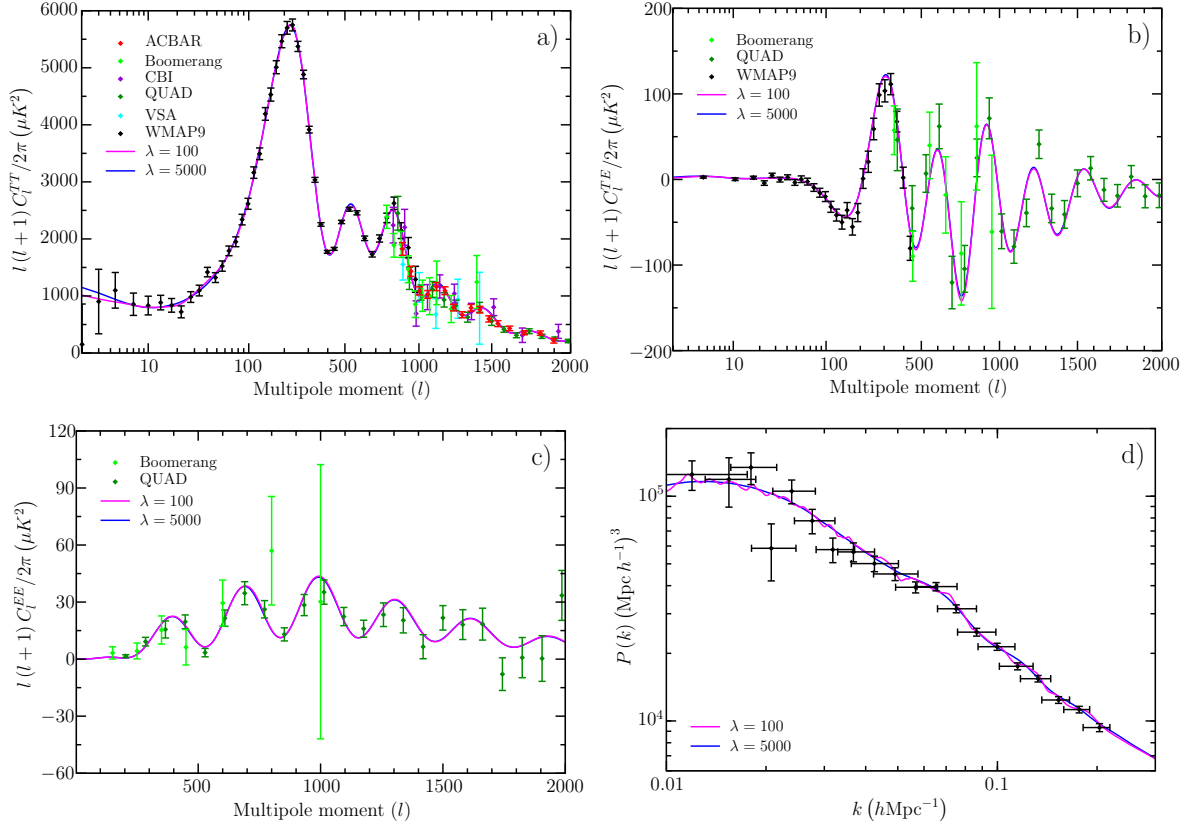


Figure 3. The fit to the data of the a) TT power spectrum b) TE power spectrum c) EE power spectrum d) LRG power spectrum corresponding to the PPS recovered from the WMAP-9, small scale CMB and SDSS-4 LRG data. In each case results are shown for both $\lambda = 100$ and $\lambda = 5000$.

The features in the recovered spectra can be understood with reference to Fig. 3, which shows a comparison to the data of the angular power spectra and the galaxy power spectrum derived from the estimated PPS. Since our method maximises the likelihood for a given roughness of the PPS, the fit to the data is correctly weighted by the inverse covariance matrices. The recovered spectra all provide good fits to the observations. The $\lambda = 100$ spectrum found from the WMAP-9, small-scale CMB and SDSS-4 LRG data (1582 data points) has $\chi^2 = 1629$ while the $\lambda = 5000$ spectrum has $\chi^2 = 1650$. These should be compared to $\chi^2 = 1664$ for the best-fit power-law model with $n_s = 0.969$.

3.2 Uncorrelated bandpowers

Since the reconstructed PPS is smooth by design, neighbouring elements of it are highly *correlated*, particularly on large and small scales. This can be seen using the correlation function calculated using the elements of the frequentist covariance matrix (eq. A.10):

$$C(k_0; k) \equiv \frac{\sum_{i,j} \Sigma_{F|ij} \phi_i(k_0) \phi_j(k)}{\left[\sum_{i,j} \Sigma_{F|ij} \phi_i(k_0) \phi_j(k_0) \right]^{1/2} \left[\sum_{i,j} \Sigma_{F|ij} \phi_i(k) \phi_j(k) \right]^{1/2}}, \quad (3.1)$$

which is displayed in Fig. 4. The correlated errors hinder the interpretation of the significance of any features in the reconstructed PPS. To display the true information content of the

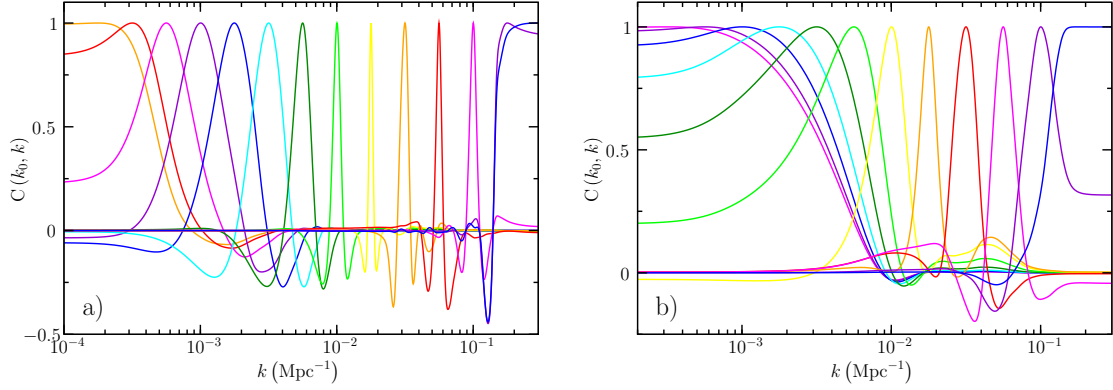


Figure 4. The correlation function (eq. 3.1) of the frequentist covariance error matrix Σ_F (eq. A.10) for (a) $\lambda = 100$ and (b) $\lambda = 5000$, for the reconstructions using the WMAP-9, small-scale CMB and SDSS-4 LRG data. The lines correspond to logarithmically equally spaced values from $k_0 = 2 \times 10^{-4} \text{ Mpc}^{-1}$ to $k_0 = 0.3 \text{ Mpc}^{-1}$ in panel (a) and from $k_0 = 3 \times 10^{-4} \text{ Mpc}^{-1}$ to $k_0 = 0.2 \text{ Mpc}^{-1}$ in panel (b). Note that $C(k_0; k_0) = 1$ and that $C(k_0; k) = C(k; k_0)$.

recovered PPS we follow [129, 143] and construct statistically *independent* band powers.

Correlated bandpowers $\mathbf{q} = \mathbf{T}\hat{\mathbf{p}}$ are defined as the mean of the PPS over separate neighbouring wavenumber ranges. Thus the elements of each row of \mathbf{T} are identical when not equal to zero and sum to unity. The columns only have one nonzero entry. The frequentist covariance matrix of the bandpowers is $\Sigma_N = \mathbf{T}\Sigma_F\mathbf{T}^T$. For $\lambda = 100$, the various estimates of the number of effective parameters (as defined in Appendix D) are $\nu_1 = 33.5$, $\nu_2 = 38.6$ and $\nu_3 = 28.4$, while for $\lambda = 5000$ they are $\nu_1 = 10.5$, $\nu_2 = 14.5$ and $\nu_3 = 6.6$. Hence we choose 35 bandpowers for $\lambda = 100$ and 10 bandpowers for $\lambda = 5000$.

Any transformation of the form $\tilde{\mathbf{q}} = \mathbf{G}\mathbf{q}$ where $\mathbf{G} \equiv \mathbf{D}\mathbf{O}\tilde{\mathbf{G}}$ and \mathbf{D} is diagonal, \mathbf{O} is orthogonal and $\tilde{\mathbf{G}}$ satisfies $\Sigma_N^{-1} = \tilde{\mathbf{G}}^T\tilde{\mathbf{G}}$ will produce uncorrelated bandpowers $\tilde{\mathbf{q}}$ with the covariance matrix \mathbf{D}^2 . As recommended in [209] we obtain $\tilde{\mathbf{G}}$ by first performing the diagonalisation $\Sigma_N = \mathbf{E}^T\mathbf{\Pi}\mathbf{E}$, where \mathbf{E} is the eigenvector matrix and $\mathbf{\Pi}$ is the diagonal matrix of eigenvalues, and then setting $\tilde{\mathbf{G}} = \mathbf{E}^T\mathbf{\Pi}^{-1/2}\mathbf{E}$. After setting $\mathbf{O} = \mathbf{I}$ and $D_{II}^{-1} = \sum_J \tilde{G}_{IJ}$ the rows of \mathbf{G} form bandpower window functions which are normalised to unity, $\sum_J G_{IJ} = 1$. The bandpowers \mathbf{q} are chosen by an iterative algorithm designed to ensure that the window functions are as well-behaved and non-negative as possible.⁷ The window functions and the uncorrelated bandpowers are shown in Figs. 5 and 6 respectively. The connection between the window functions and the resolution kernels of Fig. 12 is clearly apparent; the window functions are sharper and more densely clustered where the resolution is higher. Thus the window functions at $k = 0.03 \text{ Mpc}^{-1}$ and $k = 0.05 \text{ Mpc}^{-1}$, which correspond to the first and second troughs in the CMB TT power spectrum where the resolution is poorer, are lower and broader than their neighbours. This is particularly interesting given that no explicit information about the resolution was included in the algorithm that chose the bandpowers.⁸

⁷Note this is different from selecting the bandpowers according to the signal-to-noise ratio [210].

⁸The importance of resolution was previously noted in [209] where it was reported that the window functions become jittery when the natural resolution of the galaxy surveys is exceeded.

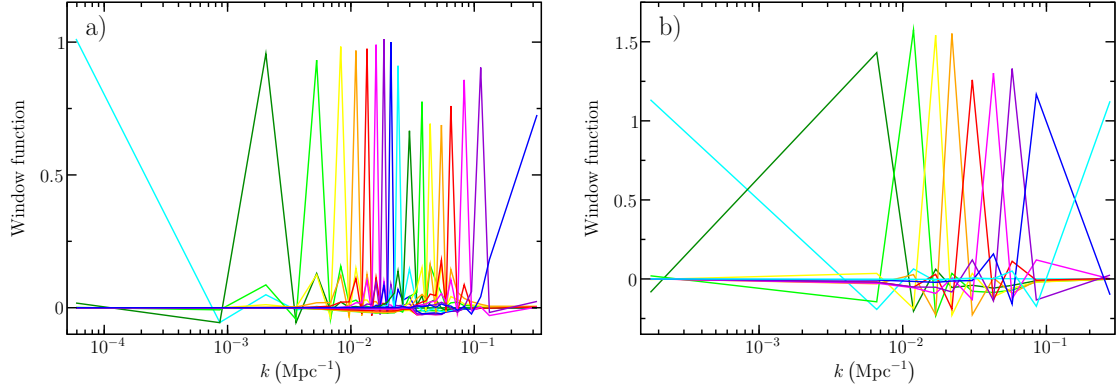


Figure 5. The bandpower window functions for (a) $\lambda = 100$ and (b) $\lambda = 5000$ (right panel) of the reconstructions using the WMAP-9, small-scale CMB and SDSS-4 LRG data. Only even number window functions are displayed in panel (a) for clarity.

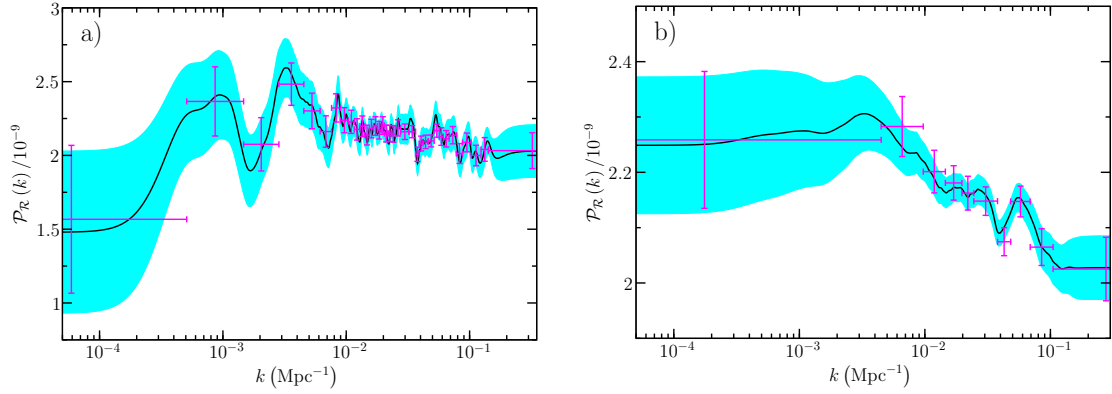


Figure 6. The decorrelated bandpowers for (a) $\lambda = 100$ and (b) $\lambda = 5000$. The black line is the PPS recovered from the WMAP-9, small-scale CMB and SDSS-4 LRG data. The light band is the 1σ error obtained from the square root of the diagonal elements of the frequentist covariance matrix Σ_F (eq. A.10). The vertical error bars are the 1σ errors D_{II} Sec. 3.2). The horizontal error bars indicate the wavenumber range to which the bandpowers are most sensitive.

3.3 Background parameter errors

As an illustration of how errors in the background parameters affect the recovered PPS, we calculate the associated covariance matrix Σ_P (eq. A.11) using the error matrix

$$U = \text{diag} \left((0.025 \omega_b)^2, (0.050 \omega_c)^2, (0.035 h)^2, (0.17 \tau)^2, (0.037 b_{\text{LRG}})^2 \right), \quad (3.2)$$

which corresponds to the uncertainties in the parameters from the WMAP and SDSS-4 LRG data.⁹ Fig. 7 shows how these uncertainties contribute to the diagonal elements of Σ_P for $\lambda = 100$. The results were obtained by error propagation using the derivatives $\partial W_{\mathbb{Z}}/\partial \theta$

⁹These uncertainties were obtained under the assumption of a power-law PPS. In principle *independent* data sets ought to be used to determine the background parameters and the PPS separately, but we are presenting an illustrative example here rather than a full analysis.

evaluated numerically by a modified version of CAMB [165].¹⁰ Varying the background parameters causes the reconstructed PPS to alter in such a way that the predicted data remain almost constant. Hence $\hat{\mathbf{p}}$ changes in the opposite direction to the change in the predicted data when the PPS is held fixed. A parameter that, for instance, increases the height of the integral kernels $\mathcal{K}_a^{(\mathbb{Z})}(\boldsymbol{\theta}, k)$ will reduce the amplitude of the estimated PPS.

For the reconstruction from the WMAP TT data with $\lambda = 100$ the error is largest on the scales corresponding to the CMB acoustic peaks where the PPS is most sensitive to the background parameters. Increasing ω_b raises the first acoustic peak and lowers the second due to greater baryon loading, while increasing ω_c reduces the early ISW effect and lowers both peaks. The angular diameter distance of the last scattering surface falls with increasing h , which shifts the peaks laterally towards larger scales. Incorrect values of these three parameters produce different oscillatory patterns in the recovered PPS. This leads to the series of peaks seen in the figures. Most of the error on large scales originates from ω_c and h . In a flat universe increasing h or decreasing ω_c with the other parameters held constant is equivalent to raising Ω_Λ . This enhances the late ISW effect, boosting the predicted TT angular power spectrum for low multipoles and lowering $\hat{\mathbf{p}}$ on large scales. Increasing τ , the optical depth to last scattering, suppresses the TT power spectrum for $\ell \gtrsim 10$ and raises the reionisation bump at $\ell \simeq 4$ in the TE and EE spectra. Hence increasing τ raises the reconstructed PPS for $k \gtrsim 10^{-3} \text{ Mpc}^{-1}$ but also lowers it on large scales when the WMAP EE and TE polarisation data is added. Consequently the contribution to the error on large scales from the uncertainty in τ is negligible for the WMAP TT data alone but greatly increased by the inclusion of the TE data. Adding the small-scale CMB data introduces features into the error associated with the higher acoustic peaks. When the SDSS-4 LRG data is used in the deconvolution, increasing ω_c and b_{LRG} suppresses $\hat{\mathbf{p}}$ on small scales. This is because increasing ω_c delays the epoch of matter-radiation equality, which moves the matter power spectrum turnover towards smaller scales and raises the predicted SDSS-4 LRG data. Increasing b_{LRG} boosts the overall normalisation of the galaxy power spectrum. Thus the ω_c and b_{LRG} contributions dominate the error on small scales. These uncertainties are *not* included in the usual parameter estimates obtained assuming a power-law PPS (e.g. [167]).

As shown in Fig. 8, the error in the reconstruction is lower for $\lambda = 5000$. The lower resolution means that the reconstructions are more stable and less sensitive to uncertainties in the background parameters as well as to noise in the data. However the reconstruction on large scales is particularly sensitive to τ as it depends on the TT power spectrum at high multipoles. The error due to noise in the data dominates the error due to uncertain parameter values on large scales because of cosmic variance, but on smaller scales it is subdominant.

3.4 Statistical significance of the features

To investigate whether the features in the recovered spectra are present in the true PPS or merely noise induced artifacts, we must determine if the reconstructions are statistically consistent with a featureless power-law. Following [135] and [142] we perform a hypothesis test, with the null hypothesis being that the true spectrum is a power-law with $n_s = 0.969$.

We generate 10^5 simulated data realisations under the assumption that the null hypothesis is true, perform inversions and compare the distribution of the results with the reconstructions from the actual data in Fig. 9. From the scatter shown in the plots it is apparent that the features represent at most $\sim 2\sigma$ deviations from a power-law.

¹⁰Note that when higher accuracy is required it would be better to rely on Monte Carlo simulations rather than attempt to calculate higher order derivatives.

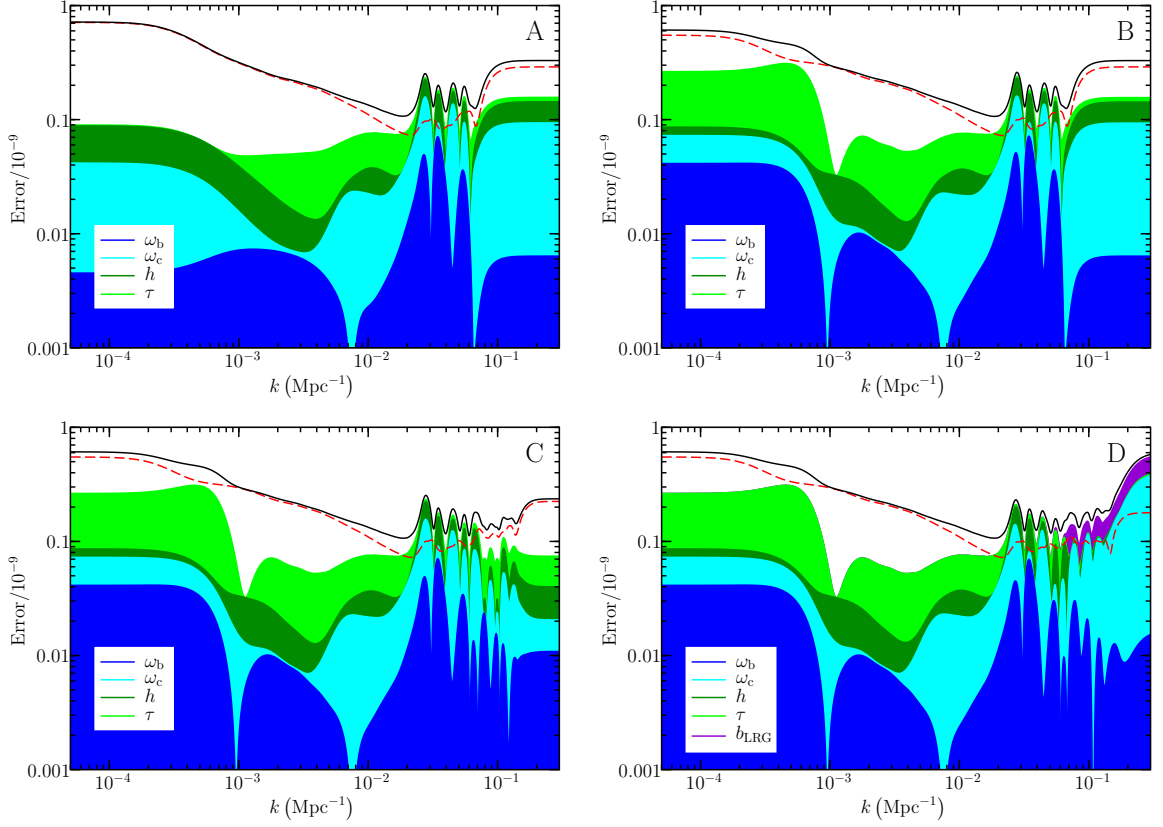


Figure 7. The contributions of the different background parameters to the square root of the diagonal elements of the matrix Σ_P (eq. A.11) for the 4 data combinations. The error contributions are added in quadrature and $\lambda = 100$ throughout. In each panel the dashed line is the square root of the diagonal elements of the matrix Σ_F (eq. A.10) and is included for comparison. The solid line is the square root of the diagonal elements of the matrix Σ_T (eq. A.9).

To quantify the significance of the individual features more precisely we employ two test statistics. The first is a two-tailed test statistic [164]

$$T_1 \equiv \int_0^\infty \xi(k) \hat{\mathcal{P}}_{\mathcal{R}}(k) dk = \sum_i r_i \hat{p}_i, \quad (3.3)$$

where the kernel $\xi(k)$ is non-zero in the region of the feature under study, and

$$r_i \equiv \int_{k_i}^{k_{i+1}} \xi(k) dk. \quad (3.4)$$

The second is a one-tailed statistic [126]

$$T_2 \equiv \int_0^\infty \xi(k) \left[\hat{\mathcal{P}}_{\mathcal{R}}(k) - \mathcal{P}_{\mathcal{R}}^{\text{PL}}(k) \right]^2 dk = \sum_i r_i (\hat{p}_i - \hat{p}_i^{\text{PL}})^2. \quad (3.5)$$

Here $\mathcal{P}_{\mathcal{R}}^{\text{PL}}(k) \equiv \sum_i p_i^{\text{PL}} \phi_i(k)$ is the power-law spectrum associated with the null hypothesis.

Recall that the p -value of a test statistic is the probability under the null hypothesis that the statistic has a value at least as extreme as the one actually observed. For each feature

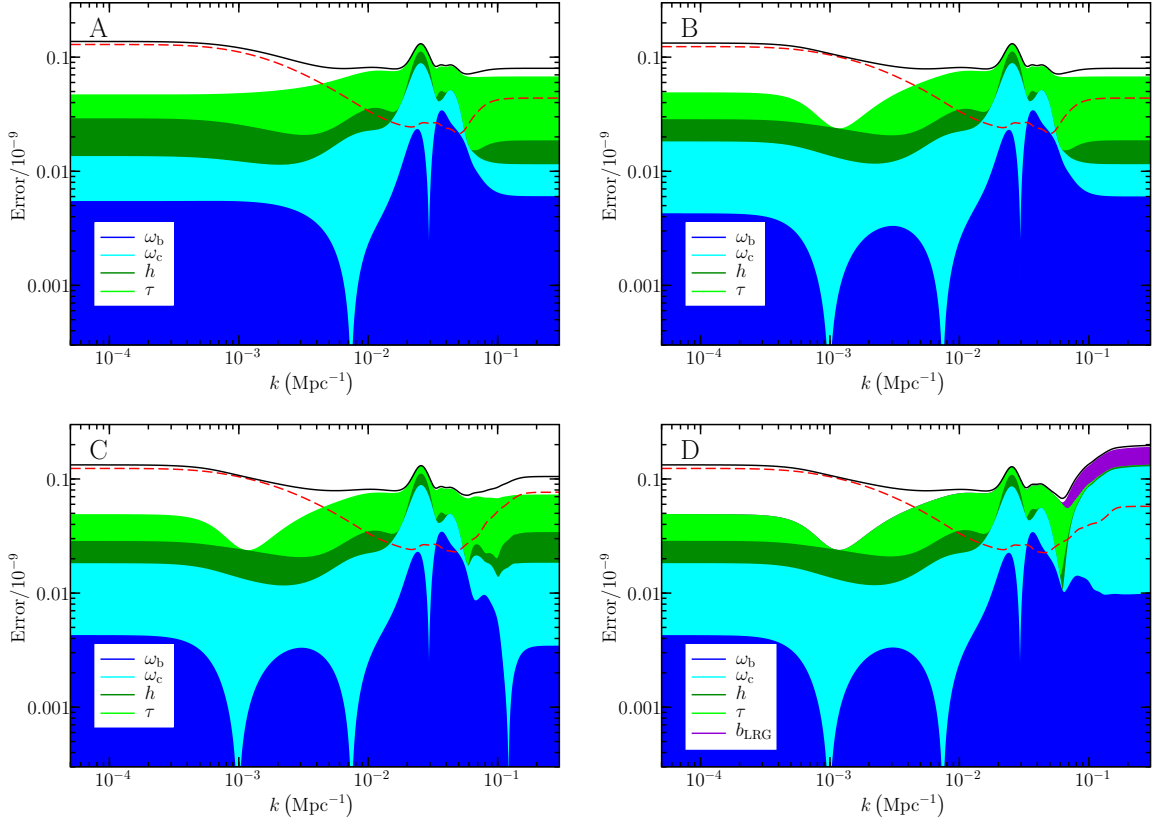


Figure 8. Same as Fig. 7, but for $\lambda = 5000$.

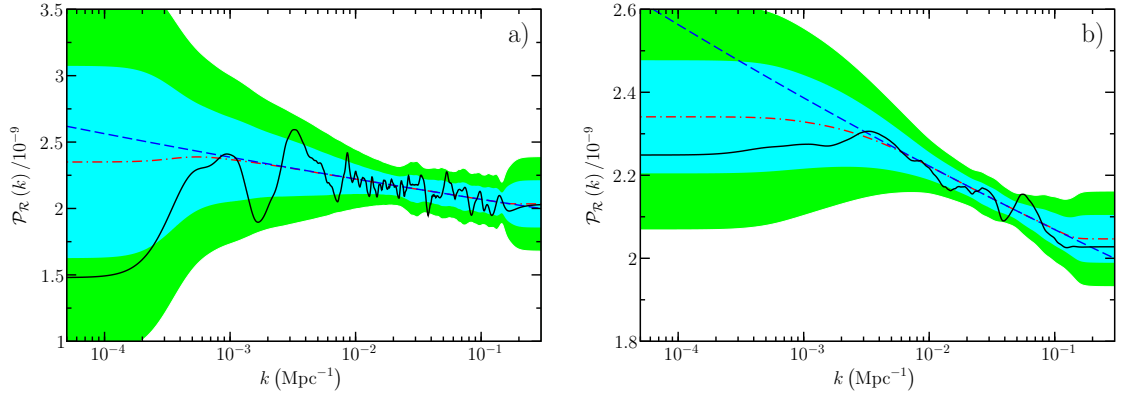


Figure 9. Comparison of the PPS recovered from the WMAP-9, small-scale CMB and SDSS-4 LRG data (full black line) with the results of 10^5 simulated reconstructions for a) $\lambda = 100$, b) $\lambda = 5000$. The simulations were generated from a power-law PPS with $n_s = 0.969$ (dashed blue line) and the shaded bands indicate the 1σ and 2σ error estimate from Monte Carlo simulations. The red dot-dashed line is the mean of the reconstructions.

we use a top-hat kernel $\xi(k)$ and estimate the p -values of T_1 and T_2 from the distributions of the test statistics in the simulated PPS reconstructions for two values of the regularisation parameter λ . The results are listed in Tables 1 and 2.

$k\text{-range / Mpc}^{-1}$	$5 \times 10^{-5} - 3 \times 10^{-4}$	0.0013 – 0.0023	0.0030 – 0.0040	0.0067 – 0.0075
T_1 p -value	0.13	0.077	0.088	0.10
T_2 p -value	0.17	0.16	0.19	0.22

$k\text{-range/Mpc}^{-1}$	0.0084 – 0.0088	0.0125 – 0.0131	0.0145 – 0.0150	0.0215 – 0.0223
T_1 p -value	0.089	0.14	0.13	0.12
T_2 p -value	0.18	0.29	0.27	0.27

$k\text{-range/Mpc}^{-1}$	0.0326 – 0.0344	0.0362 – 0.0402	0.0510 – 0.0560	0.0810 – 0.0865
T_1 p -value	0.13	0.026	0.055	0.097
T_2 p -value	0.28	0.089	0.14	0.22

Table 1. The p -values of selected features in the PPS recovered from the WMAP-9, small-scale CMB and SDSS-4 LRG data with $\lambda=100$, for two-sided (eq. 3.3) and one-sided (eq. 3.5) statistics.

$k\text{-range/Mpc}^{-1}$	0.0125 – 0.0155	0.036 – 0.044	0.049 – 0.070
T_1 p -value	0.12	0.057	0.041
T_2 p -value	0.26	0.13	0.11

Table 2. Same as Table 1, but for $\lambda = 5000$.

A high T_2 value can arise from a peak, dip or an oscillation about the null hypothesis power-law. Only the first two cases will result in an extreme T_1 value. Hence the T_1 p -value of a feature is normally less than half that of the T_2 p -value.¹¹ Unlike T_2 , the test statistic T_1 is found to closely obey a Gaussian distribution under the null hypothesis. The dips in the $\lambda = 100$ reconstruction at $k \simeq 0.0017$ and 0.038 Mpc^{-1} and the bump at $k \simeq 0.053 \text{ Mpc}^{-1}$ therefore represent 1.4σ , 1.9σ and 1.6σ deviations, according to their T_1 p -values. Similarly, the dip at $k \simeq 0.038 \text{ Mpc}^{-1}$ and the bump at $k \simeq 0.055 \text{ Mpc}^{-1}$ in the $\lambda = 5000$ reconstruction have 1.6σ and 1.7σ statistical significance respectively.

Turning to the bandpower analysis, Fig. 10 displays the distribution of the bandpowers obtained from the 10^5 simulated reconstructions together with the bandpowers derived from the observed data. For $\lambda = 100$ the bandpowers at $k = 0.0020$, 0.038 and 0.054 Mpc^{-1} correspond to 1.3σ , 2.1σ and 1.8σ fluctuations, while for $\lambda = 5000$ the bandpowers at $k = 0.043$ and 0.058 Mpc^{-1} constitute 1.3σ and 2.1σ fluctuations.

In addition to individual features we ought to consider statistics associated with the entire PPS in order to avoid *a posteriori* selection effects. For $\lambda = 100$ the p -value of the (χ^2 of the) fit of all 35 power-law bandpowers to those recovered from the WMAP-9, small-scale CMB and SDSS-4 LRG data is 0.93 (equivalent to a -1.5σ deviation). The p -value drops to 0.45 for the 10 bandpowers of the $\lambda = 5000$ reconstruction (a 0.1σ deviation). Thus the reconstructions with $\lambda = 100$ and $\lambda = 5000$ are both statistically *consistent* with a power-law.

This may seem surprising, given the poor χ^2 of the fit of the power-law model to the WMAP-9 data. Indeed, of the simulated TT power spectra generated under the null hypothesis, only 6.6% have as high a χ^2 value as that of the actual data. However, the reconstructed PPS is sensitive to the *running average* of the data. This is illustrated in Fig. 11, which shows how the TT spectra corresponding to the reconstructions with $\lambda = 100$ and $\lambda = 5000$ trace the $\ell = 25$ and $\ell = 97$ running averages. From the simulations we find

¹¹The exception occurs for the infrared cutoff due to the bias in the reconstruction on large scales.

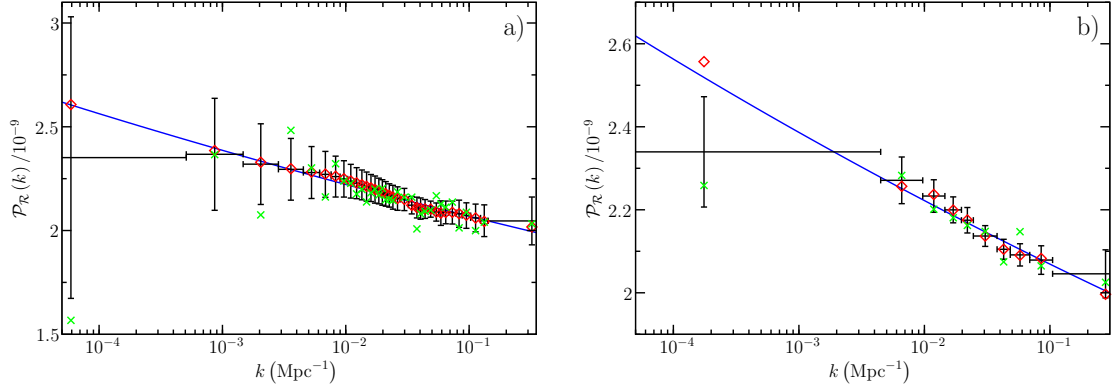


Figure 10. The solid line is the power-law spectrum corresponding to the null hypothesis. The red diamonds are the bandpowers calculated from the power-law spectrum — if the bandpower window functions were perfectly well-behaved, the diamonds would lie on the solid line. The vertical error bars are the 1σ scatter of the bandpowers from 10^5 simulated reconstructions while the horizontal error bars indicate the wavenumber range to which the bandpowers are most sensitive. Note that these are not centred on the diamonds due to the bias of the deconvolution method on very large and very small scales. The green crosses are the bandpowers recovered from the actual WMAP-9, small scale CMB and SDSS-4 LRG data. Panel a) is for $\lambda = 100$ and panel b) is for $\lambda = 5000$.

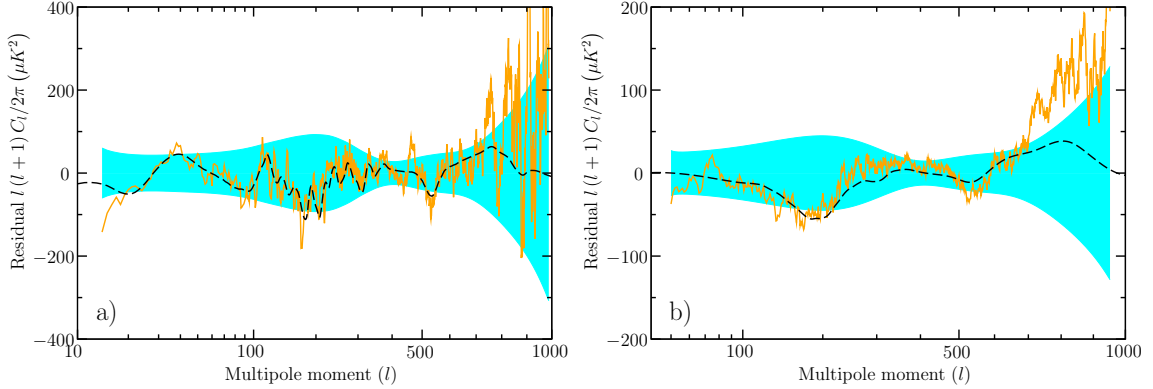


Figure 11. a) Comparison of residuals for the $\ell = 25$ running average of the WMAP-9 TT data (solid orange line) with the residuals corresponding to the $\lambda = 100$ reconstruction from the WMAP-9, small-scale CMB and SDSS-4 LRG data (dashed black line). The residuals are obtained by subtracting off the TT spectrum of the null hypothesis power-law model. The band indicates the 1σ scatter of the $\ell = 25$ running average data, calculated from the WMAP-9 TT covariance matrix. b) Same as panel a) but for the $\ell = 97$ running average and the $\lambda = 5000$ reconstruction.

the correlation coefficient between the bandpower χ^2 (for $\lambda = 100$) and the χ^2 of the $\ell = 25$ running average data to be 0.55, compared to a correlation coefficient of 0.12 between the bandpower χ^2 and the χ^2 of the unaveraged data. For $\lambda = 5000$ the equivalent correlation coefficients are 0.40 for the $\ell = 97$ running average and 0.077 for the unaveraged data. The p -value of the χ^2 of the $\ell = 25$ running average data is 0.96 (a -1.7σ deviation) while that of the $\ell = 97$ running average data is 0.51 (a -0.02σ deviation). Since fluctuations in the running average are responsible for most of those in the reconstructions, both reconstructions can be said to be presently consistent with a power-law for an assumed Λ CDM cosmology.

4 Conclusions

The task of inverting noisy data with limited resolution and an uncertain background cosmology to extract the PPS of curvature perturbations is a challenging one, with *no* unique solution. We have studied and validated a robust regularisation procedure for obtaining a stable estimate of the PPS from multiple cosmological data sets. A perturbative analysis of the relationship of the estimate to the true PPS (which can be applied to any inversion method) has allowed us (see Appendix C.3) to identify the different components of the bias and variance that characterise the performance of the reconstruction. The bias is found to depend on first- and second-order resolution kernels which indicate the resolving power of the inversion. Moreover, we find that data with a non-Gaussian likelihood function can increase the bias, even if the data itself is unbiased. The variance arises from both noise in the data and uncertainties in the background cosmological parameters.

We have seen how the regularisation parameter governs the trade-off between the bias and variance, and tested (in Appendix D) several methods for choosing the optimum value of the parameter, with mixed results. The recovered spectra exhibit interesting deviations from the usually assumed scale-free power-law, with a statistical significance of around 2σ as determined from Monte Carlo simulations. However, using a decorrelated bandpower analysis the spectra are found to be statistically consistent with a power-law. We emphasise that these conclusions are dependent on the *assumed* background cosmology — here taken to be the standard Λ CDM model. We do not expect the statistical significance of the features in the deconvolved spectrum to change much when the Λ CDM parameters are varied within their quoted uncertainties. However if the assumed cosmology is radically different then the recovered spectrum may indeed look rather different too (e.g. Spectrum D in Sec. B.3 would be recovered for a flat CHDM model with no dark energy [14, 15, 17]).

Our results are in agreement with previous work by other authors (e.g. [134, 135, 142]) who also reported a cutoff in the PPS on large scales and features at $k \simeq 0.0017 \text{ Mpc}^{-1}$ and $k \simeq 0.0032 \text{ Mpc}^{-1}$ due to the WMAP glitches. In particular, [129, 143] found the χ^2 for 34 and 46 bandpowers in the ranges $9 \times 10^{-3} \lesssim k \lesssim 0.03 \text{ Mpc}^{-1}$ and $2 \times 10^{-3} \lesssim k \lesssim 0.03 \text{ Mpc}^{-1}$ recovered from the WMAP-5 data to have -1.5σ and -1.1σ significance respectively, using the Tikhonov regularisation and cosmic inversion methods. This is quite consistent with the -1.5σ deviation of our 35 bandpowers. In summary, the statistical significance of these deviations is not yet compelling, however the potential here for probing the physical process that generated the primordial curvature perturbations certainly motivates further investigation.

Many studies have compared regularisation parameter selection methods, though not in this context (see [211–213] and references therein). The consensus in the literature is that the results are strongly dependent on the application at hand [196, 214]. Here the selection of the regularisation parameter is complicated by the fact that the width of the CMB integral kernels and the noise level of the data vary strongly with the multipole moment (see Appendix D).¹² Until a suitable regularisation parameter selection method is found, we recommend performing inversions using a range of λ values informed by simulations.

There are some subtleties associated with adding data to improve the inversion. The bias is reduced only if the PPS departs from the flat H-Z form on the scales covered by the new data. Including a non-Gaussian distributed data set can increase the bias, particularly for small λ . The variance will increase if the additional data is strongly dependent on background cosmological parameters with large uncertainties.

¹²By contrast, other inversion applications (e.g. image restoration) do not suffer from an equivalent problem.

The next step would naturally be to apply our method to the recent data release from the Planck satellite [215], as well as small angular experiments such as ACT [216] and SPT [217]. Planck has much greater resolution than WMAP on small scales but the Planck team have applied [218] only a parametric method using a smoothing spline (e.g. [107]), which we consider inadequate for recovering sharp features in the spectrum, such as we have uncovered in the WMAP data. Should these turn out to exist in the Planck data too, it would spell the death of single-field inflationary models! The “precision” estimation of cosmological parameters *assuming* a power-law spectrum [219] would also then need to be revisited.

5 Acknowledgments

We acknowledge use of the `CAMB` and `cosmoMC` codes and thank the WMAP and Planck teams for making their data and analysis tools publicly available. This work was supported by the EU Marie Curie Network ‘UNILHC’ (PITN-GA-2009-237920) and a Niels Bohr Professorship awarded by the Danish National Research Foundation.

A Error analysis

We analyse Tikhonov regularisation within the frequentist statistical framework. This approach involves a hypothetical ensemble of PPS estimates obtained from repeated independent identical measurements of the data and the background parameters. There exists an unknown true PPS \mathbf{p}_t and a true set of background parameters $\boldsymbol{\theta}_t$. The estimates constitute samples from the distribution $P(\hat{\mathbf{p}}|\mathbf{p}_t)$ which must be appropriately characterised.

Since $\mathbf{d} = \mathbf{W}(\boldsymbol{\theta}_t)\mathbf{p}_t + \mathbf{n}$ the data depend on the true background parameters $\boldsymbol{\theta}_t$ but the reconstruction is performed using the estimate $\hat{\boldsymbol{\theta}}$. Thus the recovered PPS can be written as

$$\hat{\mathbf{p}}(\mathbf{d}, \hat{\boldsymbol{\theta}}) = \mathcal{T}(\mathbf{p}_t, \boldsymbol{\theta}_t, \hat{\boldsymbol{\theta}}, \mathbf{n}), \quad (\text{A.1})$$

where the transfer function \mathcal{T} gives the relationship of $\hat{\mathbf{p}}$ to \mathbf{p}_t , which is nonlinear in general. Performing a Taylor expansion of the transfer function about some PPS \mathbf{p}_1 close to \mathbf{p}_t gives

$$\begin{aligned} \hat{p}_i(\mathbf{d}, \hat{\boldsymbol{\theta}}) &= \mathcal{T}_i(\mathbf{p}_1, \boldsymbol{\theta}_t, \boldsymbol{\theta}_t, \mathbf{0}) + \sum_j R_{ij} \Delta p_j + \frac{1}{2} \sum_{j,k} Y_{ijk} \Delta p_j \Delta p_k \\ &+ \sum_{\mathbb{Z}, a} M_{ia}^{(\mathbb{Z})} n_a^{(\mathbb{Z})} + \sum_{\alpha} M_{i\alpha} u_{\alpha} + \sum_{\mathbb{Z}, j, a} Z_{ija}^{(\mathbb{Z})} \Delta p_j n_a^{(\mathbb{Z})} + \sum_{j, \alpha} Z_{ij\alpha} \Delta p_j u_{\alpha} \\ &+ \frac{1}{2} \sum_{\mathbb{Z}, \mathbb{Z}', a, b} X_{iab}^{(\mathbb{Z}\mathbb{Z}')} n_a^{(\mathbb{Z})} n_b^{(\mathbb{Z}')} + \sum_{\mathbb{Z}, a, \alpha} X_{ia\alpha}^{(\mathbb{Z})} n_a^{(\mathbb{Z})} u_{\alpha} + \frac{1}{2} \sum_{\alpha, \beta} X_{i\alpha\beta} u_{\alpha} u_{\beta} + \dots \end{aligned} \quad (\text{A.2})$$

Here $\Delta p_i \equiv p_{ti} - p_{1i}$, and

$$\begin{aligned}
M_{ia}^{(\mathbb{Z})} &\equiv \left. \frac{\partial \hat{p}_i}{\partial d_a^{(\mathbb{Z})}} \right|_{\hat{\mathbf{d}}_1, \boldsymbol{\theta}_t}, & M_{i\alpha} &\equiv \left. \frac{\partial \hat{p}_i}{\partial \theta_\alpha} \right|_{\hat{\mathbf{d}}_1, \boldsymbol{\theta}_t}, \\
X_{iab}^{(\mathbb{Z}\mathbb{Z}')} &\equiv \left. \frac{\partial^2 \hat{p}_i}{\partial d_a^{(\mathbb{Z})} \partial d_b^{(\mathbb{Z}')}} \right|_{\hat{\mathbf{d}}_1, \boldsymbol{\theta}_t}, & X_{ia\alpha}^{(\mathbb{Z})} &\equiv \left. \frac{\partial^2 \hat{p}_i}{\partial d_a^{(\mathbb{Z})} \partial \theta_\alpha} \right|_{\hat{\mathbf{d}}_1, \boldsymbol{\theta}_t}, & X_{i\alpha\beta} &\equiv \left. \frac{\partial^2 \hat{p}_i}{\partial \theta_\alpha \partial \theta_\beta} \right|_{\hat{\mathbf{d}}_1, \boldsymbol{\theta}_t}, \\
R_{ij} &\equiv \sum_{\mathbb{Z}, a} M_{ia}^{(\mathbb{Z})} W_{aj}^{(\mathbb{Z})}(\boldsymbol{\theta}_t), & Y_{ijk} &\equiv \sum_{\mathbb{Z}, \mathbb{Z}', a, b} X_{iab}^{(\mathbb{Z}\mathbb{Z}')} W_{aj}^{(\mathbb{Z})}(\boldsymbol{\theta}_t) W_{bk}^{(\mathbb{Z}')}(\boldsymbol{\theta}_t), \\
Z_{ija}^{(\mathbb{Z})} &\equiv \sum_{\mathbb{Z}', b} X_{iab}^{(\mathbb{Z}\mathbb{Z}')} W_{bj}^{(\mathbb{Z}')}(\boldsymbol{\theta}_t), & Z_{ij\alpha} &\equiv \sum_{\mathbb{Z}, a} X_{ia\alpha}^{(\mathbb{Z})} W_{aj}^{(\mathbb{Z})}(\boldsymbol{\theta}_t), \tag{A.3}
\end{aligned}$$

where $\hat{\mathbf{d}}_1 \equiv \mathbf{W}(\boldsymbol{\theta}_t) \mathbf{p}_1$ is the data estimated from \mathbf{p}_1 . In eq.(A.2), $\mathcal{T}(\mathbf{p}_1, \boldsymbol{\theta}_t, \boldsymbol{\theta}_t, \mathbf{0})$ is the reconstruction resulting from the noise-free data $\hat{\mathbf{d}}_1$ using the true background parameters and should ideally equal \mathbf{p}_1 , but does not do so in general. The first-order resolution matrix \mathbf{R} characterises the linear mapping of \mathbf{p}_t to $\hat{\mathbf{p}}$. Perfect resolution is attained when \mathbf{R} equals the identity matrix \mathbf{I} . However, multiplication by \mathbf{R} usually acts as a smoothing operation so that $\hat{\mathbf{p}}$ is a smoothed version of \mathbf{p}_t , even for noise-free data. Components of \mathbf{p}_t belonging to the null space of \mathbf{R} are not reproduced in $\hat{\mathbf{p}}$. The second-order resolution matrix \mathbf{Y} describes an undesirable quadratic mapping from \mathbf{p}_t to $\hat{\mathbf{p}}$ and should ideally vanish. The remaining contributions to $\hat{\mathbf{p}}$ represent artifacts caused by noise in the data and incorrect estimates of the background parameters.

Using eq.(A.2) the leading sources of bias in the reconstruction can be identified:

$$\begin{aligned}
\text{Bias}(\hat{p}_i) &\equiv \langle \hat{p}_i - p_{ti} \rangle = \mathcal{T}_i(\mathbf{p}_1, \boldsymbol{\theta}_t, \boldsymbol{\theta}_t, \mathbf{0}) - p_{1i} + \sum_j (R_{ij} - I_{ij}) \Delta p_j \\
&\quad + \frac{1}{2} \sum_{j,k} Y_{ijk} \Delta p_j \Delta p_k + \frac{1}{2} \sum_{\mathbb{Z}, a, b} X_{iab}^{(\mathbb{Z}\mathbb{Z})} N_{ab}^{(\mathbb{Z})} + \frac{1}{2} \sum_{\alpha, \beta} X_{i\alpha\beta} N_{\alpha\beta} + \dots \tag{A.4}
\end{aligned}$$

Here $\mathcal{T}(\mathbf{p}_1, \boldsymbol{\theta}_t, \boldsymbol{\theta}_t, \mathbf{0}) - \mathbf{p}_1$ is the error in recovering \mathbf{p}_1 . The quantity $(\mathbf{R} - \mathbf{I}) \Delta \mathbf{p}$ represents the error due to the limited resolution of the reconstruction and is labelled the ‘null space error’ in [158]. The quadratic mapping associated with the second-order resolution matrix contributes directly to the bias. It is apparent that a quadratic (or higher order) dependence of the estimated PPS on the data or the background parameters will also increase the bias.

Rather than work with the resolution matrices it is more convenient to use their continuous counterparts, the resolution kernels. Resolution kernels enable the resolving power of different inversion methods to be compared directly (see [159] for an example). First-order resolution kernels were originally introduced as part of the Backus-Gilbert method [160] and were generalised subsequently to higher order in [161]. The continuous analogue of the first-order contribution $\delta \hat{p}_i^{(1)} \equiv \sum_j R_{ij} p_{tj}$ of \mathbf{p}_t to $\hat{\mathbf{p}}$ is the contribution

$$\delta \hat{\mathcal{P}}_{\mathcal{R}}^{(1)}(k_0) \equiv \int_0^\infty R(k_0; k) \mathcal{P}_{\mathcal{R}}(k) dk \tag{A.5}$$

to $\hat{\mathcal{P}}_{\mathcal{R}}(k_0) \equiv \sum_i \hat{p}_i \phi_i(k_0)$. Here the first-order resolution kernel $R(k_0; k)$ is given by

$$R(k_0; k) \equiv \sum_{\mathbb{Z}, i, a} \phi_i(k_0) M_{ia}^{(\mathbb{Z})} \mathcal{K}_a^{(\mathbb{Z})}(k). \tag{A.6}$$

Similarly, the continuous analogue of the second-order contribution $\delta \hat{\mathbf{p}}_i^{(2)} \equiv \frac{1}{2} \sum_{j,k} Y_{ijk} \mathbf{p}_{tj} \mathbf{p}_{tk}$ is

$$\delta \hat{\mathcal{P}}_{\mathcal{R}}^{(2)}(k_0) \equiv \frac{1}{2} \int_0^\infty \int_0^\infty Y(k_0; k_1, k_2) \mathcal{P}_{\mathcal{R}}(k_1) \mathcal{P}_{\mathcal{R}}(k_2) dk_1 dk_2 \quad (\text{A.7})$$

where the second-order resolution kernel $Y(k_0; k_1, k_2)$ is given by

$$Y(k_0; k_1, k_2) \equiv \sum_{\mathbb{Z}, \mathbb{Z}', i, a, b} \phi_i(k_0) X_{iab}^{(\mathbb{Z}\mathbb{Z}')} \mathcal{K}_a^{(\mathbb{Z})}(k_1) K_b^{(\mathbb{Z}')} (k_2). \quad (\text{A.8})$$

With k_0 kept constant the first-order resolution kernel is a sharply peaked function of k_1 that indicates the wavenumber range of the true PPS to which the estimate $\hat{\mathcal{P}}_{\mathcal{R}}(k_0)$ is sensitive. For the reconstructed PPS to be correctly scaled, necessary properties of the first-order kernel are that for each k_0 value the kernel must have a single peak at $k = k_0$ and that the kernel must integrate to unity, $\int_0^\infty R(k_0; k) dk = 1$. Clearly the closer $R(k_0; k)$ is to the Dirac delta function $\delta(k_0 - k)$ and the closer $Y(k_0; k_1, k_2)$ is to zero, the better the resolution of the recovered PPS.

It follows from eq.(A.2) that the total frequentist covariance matrix $\Sigma_{\mathbf{T}}$ which characterises variations in $\hat{\mathbf{p}}$ between members of the ensemble due to noise and errors in the estimated background parameters is given by

$$\Sigma_{\mathbf{T}} \equiv \langle (\hat{\mathbf{p}} - \langle \hat{\mathbf{p}} \rangle) (\hat{\mathbf{p}} - \langle \hat{\mathbf{p}} \rangle)^{\text{T}} \rangle = \Sigma_{\mathbf{F}} + \Sigma_{\mathbf{P}} + \dots, \quad (\text{A.9})$$

$$\Sigma_{\mathbf{F}} \equiv \sum_{\mathbb{Z}} \mathbf{M}_{\mathbb{Z}} \mathbf{N}_{\mathbb{Z}} \mathbf{M}_{\mathbb{Z}}^{\text{T}}, \quad (\text{A.10})$$

$$\Sigma_{\mathbf{P}} \equiv \mathbf{M} \mathbf{U} \mathbf{M}^{\text{T}}, \quad (\text{A.11})$$

where $\Sigma_{\mathbf{F}}$ results from the data noise, $\Sigma_{\mathbf{P}}$ results from errors in the background parameters, and higher-order terms have been omitted.

Thus far the analysis applies to *any* reconstruction method in which the PPS is parameterised using basis functions as in eq.(2.4). In the case of Tikhonov regularisation, analytic expressions for the matrices of eq.(A.3) can be derived. To this end, consider the estimates $\hat{\mathbf{p}}_1 = \hat{\mathbf{p}}(\hat{\mathbf{d}}_1, \theta_t)$ and $\hat{\mathbf{p}}_2 = \hat{\mathbf{p}}(\mathbf{d}_2, \hat{\theta}_2)$ where \mathbf{d}_2 is a realisation of the data and $\hat{\theta}_2$ is a realisation of the estimated background parameters. Performing a Taylor expansion of $\partial Q / \partial \mathbf{p}_i|_{\hat{\mathbf{p}}_2, \mathbf{d}_2, \hat{\theta}_2}$ about $\partial Q / \partial \mathbf{p}_i|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}$, we get:

$$\begin{aligned} \left. \frac{\partial Q}{\partial \mathbf{p}_i} \right|_{\hat{\mathbf{p}}_2, \mathbf{d}_2, \hat{\theta}_2} &= \left. \frac{\partial Q}{\partial \mathbf{p}_i} \right|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t} + \sum_j A_{ij} \Delta \hat{\mathbf{p}}_j + \sum_{\mathbb{Z}, a} B_{ia}^{(\mathbb{Z})} \Delta \mathbf{d}_a^{(\mathbb{Z})} + \sum_j B_{i\alpha} \Delta \theta_\alpha \\ &+ \frac{1}{2} \sum_{j,k} C_{ijk} \Delta \hat{\mathbf{p}}_j \Delta \hat{\mathbf{p}}_k + \frac{1}{2} \sum_{\mathbb{Z}, \mathbb{Z}', a, b} D_{iab}^{(\mathbb{Z}\mathbb{Z}')} \Delta \mathbf{d}_a^{(\mathbb{Z})} \Delta \mathbf{d}_b^{(\mathbb{Z}')} + \sum_{\mathbb{Z}, a, \alpha} D_{ia\alpha}^{(\mathbb{Z})} \Delta \mathbf{d}_a^{(\mathbb{Z})} \Delta \theta_\alpha \\ &+ \frac{1}{2} \sum_{\alpha, \beta} D_{i\alpha\beta} \Delta \theta_\alpha \Delta \theta_\beta + \sum_{\mathbb{Z}, j, a} E_{ija}^{(\mathbb{Z})} \Delta \hat{\mathbf{p}}_j \Delta \mathbf{d}_a^{(\mathbb{Z})} + \sum_{j, \alpha} E_{ij\alpha} \Delta \hat{\mathbf{p}}_j \Delta \theta_\alpha \dots \quad (\text{A.12}) \end{aligned}$$

Here $\Delta \hat{\mathbf{p}}_i \equiv \hat{\mathbf{p}}_{2i} - \hat{\mathbf{p}}_{1i}$, $\Delta d_a^{(\mathbb{Z})} \equiv d_{2a}^{(\mathbb{Z})} - \hat{d}_{1a}^{(\mathbb{Z})}$, $\Delta \theta_\alpha \equiv \hat{\theta}_{2\alpha} - \theta_{t\alpha}$ and

$$\begin{aligned} A_{ij} &\equiv \frac{\partial^2 Q}{\partial p_i \partial p_j} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}, & B_{ia}^{(\mathbb{Z})} &\equiv \frac{\partial^2 Q}{\partial p_i \partial d_a^{(\mathbb{Z})}} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}, & B_{i\alpha} &\equiv \frac{\partial^2 Q}{\partial p_i \partial \theta_\alpha} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t} \\ C_{ijk} &\equiv \frac{\partial^3 Q}{\partial p_i \partial p_j \partial p_k} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}, & D_{iab}^{(\mathbb{Z}\mathbb{Z}')} &\equiv \frac{\partial^3 Q}{\partial p_i \partial d_a^{(\mathbb{Z})} \partial d_b^{(\mathbb{Z}')}} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}, & D_{ia\alpha}^{(\mathbb{Z})} &\equiv \frac{\partial^3 Q}{\partial p_i \partial d_a^{(\mathbb{Z})} \partial \theta_\alpha} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t} \\ D_{i\alpha\beta} &\equiv \frac{\partial^3 Q}{\partial p_i \partial \theta_\alpha \partial \theta_\beta} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}, & E_{ija}^{(\mathbb{Z})} &\equiv \frac{\partial^3 Q}{\partial p_i \partial p_j \partial d_a^{(\mathbb{Z})}} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}, & E_{ij\alpha} &\equiv \frac{\partial^3 Q}{\partial p_i \partial p_j \partial \theta_\alpha} \Big|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}. \end{aligned} \quad (\text{A.13})$$

We also have

$$\begin{aligned} \hat{\mathbf{p}}_{2i} &= \hat{\mathbf{p}}_{1i} + \sum_{\mathbb{Z}, a} M_{ia}^{(\mathbb{Z})} \Delta d_a^{(\mathbb{Z})} + \sum_{\alpha} M_{i\alpha} \Delta \theta_\alpha + \frac{1}{2} \sum_{\mathbb{Z}, \mathbb{Z}', a, b} X_{iab}^{(\mathbb{Z}\mathbb{Z}')} \Delta d_a^{(\mathbb{Z})} \Delta d_b^{(\mathbb{Z}')} \\ &+ \sum_{\mathbb{Z}, a, \alpha} X_{ia\alpha}^{(\mathbb{Z})} \Delta d_a^{(\mathbb{Z})} \Delta \theta_\alpha + \frac{1}{2} \sum_{\alpha, \beta} X_{i\alpha\beta} \Delta \theta_\alpha \Delta \theta_\beta + \dots \end{aligned} \quad (\text{A.14})$$

Both $\partial Q / \partial p_i|_{\hat{\mathbf{p}}_1, \hat{\mathbf{d}}_1, \theta_t}$ and $\partial Q / \partial p_i|_{\hat{\mathbf{p}}_2, \hat{\mathbf{d}}_2, \hat{\theta}_2}$ vanish since $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$ minimise $Q(\mathbf{p}, \hat{\mathbf{d}}_1, \theta_t)$ and $Q(\mathbf{p}, \mathbf{d}_2, \hat{\theta}_2)$ respectively. Therefore comparing eqs.(A.12) and (A.14) gives

$$M_{ia}^{(\mathbb{Z})} = - \sum_j A_{ij}^{-1} B_{ja}^{(\mathbb{Z})}, \quad M_{i\alpha} = - \sum_j A_{ij}^{-1} B_{j\alpha}, \quad (\text{A.15})$$

$$X_{iab}^{(\mathbb{Z}\mathbb{Z}')} = - \sum_{j, k, l} A_{ij}^{-1} C_{jkl} M_{ka}^{(\mathbb{Z})} M_{lb}^{(\mathbb{Z}')} - \sum_{j, k} A_{ij}^{-1} E_{jka}^{(\mathbb{Z})} M_{kb}^{(\mathbb{Z}')} - \sum_{j, k} A_{ij}^{-1} E_{jkb}^{(\mathbb{Z}')} M_{ka}^{(\mathbb{Z})} - \sum_j A_{ij}^{-1} D_{jab}^{(\mathbb{Z}\mathbb{Z}')}, \quad (\text{A.16})$$

$$X_{ia\alpha}^{(\mathbb{Z})} = - \sum_{j, k, l} A_{ij}^{-1} C_{jkl} M_{ka}^{(\mathbb{Z})} M_{l\alpha} - \sum_{j, k} A_{ij}^{-1} E_{jka}^{(\mathbb{Z})} M_{k\alpha} - \sum_{j, k} A_{ij}^{-1} E_{jk\alpha} M_{ka}^{(\mathbb{Z})} - \sum_j A_{ij}^{-1} D_{ja\alpha}^{(\mathbb{Z})}. \quad (\text{A.17})$$

$$X_{i\alpha\beta} = - \sum_{j, k, l} A_{ij}^{-1} C_{jkl} M_{k\alpha} M_{l\beta} - \sum_{j, k} A_{ij}^{-1} E_{jk\alpha} M_{k\beta} - \sum_{j, k} A_{ij}^{-1} E_{jk\beta} M_{k\alpha} - \sum_j A_{ij}^{-1} D_{j\alpha\beta}, \quad (\text{A.18})$$

If the likelihood function is Gaussian as in eq.(2.9) it can be shown analytically that $\hat{\mathbf{p}}(\mathbf{d}, \hat{\theta}) = \sum_{\mathbb{Z}} \tilde{\mathbf{M}}_{\mathbb{Z}}(\hat{\theta}) \mathbf{d}_{\mathbb{Z}}$ where $\tilde{\mathbf{M}}_{\mathbb{Z}}(\theta) \equiv \mathbf{S}(\theta) \mathbf{W}_{\mathbb{Z}}^T(\theta) \mathbf{N}_{\mathbb{Z}}^{-1}$ and

$$\mathbf{S}(\theta) \equiv \left[\lambda \Gamma + \sum_{\mathbb{Z}} \mathbf{W}_{\mathbb{Z}}^T(\theta) \mathbf{N}_{\mathbb{Z}}^{-1} \mathbf{W}_{\mathbb{Z}}(\theta) \right]^{-1}. \quad (\text{A.19})$$

The expression for $\hat{\mathbf{p}}$ has the form of a sum over the data sets weighted by the inverse covariance matrices, so that $\hat{\mathbf{p}}$ preferentially fits more precise data sets. It follows that in this case the estimated PPS is a linear function of the true PPS, with

$$\mathcal{T}(\mathbf{p}_t, \theta_t, \hat{\theta}, \mathbf{n}) = \sum_{\mathbb{Z}} \tilde{\mathbf{M}}_{\mathbb{Z}}(\hat{\theta}) \mathbf{W}_{\mathbb{Z}}(\theta_t) \mathbf{p}_t + \sum_{\mathbb{Z}} \tilde{\mathbf{M}}_{\mathbb{Z}}(\hat{\theta}) \mathbf{n}_{\mathbb{Z}}. \quad (\text{A.20})$$

Eq.(A.2) then becomes

$$\begin{aligned} \hat{p}_i(\mathbf{d}, \hat{\theta}) &= \sum_j R_{ij} p_{tj} + \sum_{\mathbb{Z}, a} M_{ia}^{(\mathbb{Z})} n_a^{(\mathbb{Z})} + \sum_{\alpha} M_{i\alpha} u_{\alpha} \\ &+ \sum_{j, \alpha} Z_{ij\alpha} \Delta p_j u_{\alpha} + \sum_{\mathbb{Z}, a, \alpha} X_{ia\alpha}^{(\mathbb{Z})} n_a^{(\mathbb{Z})} u_{\alpha} + \frac{1}{2} \sum_{\alpha, \beta} X_{i\alpha\beta} u_{\alpha} u_{\beta} + \dots, \end{aligned} \quad (\text{A.21})$$

where $\mathbf{M}_{\mathbb{Z}} = \tilde{\mathbf{M}}_{\mathbb{Z}}(\boldsymbol{\theta}_{\text{t}})$. It was shown [162] that the row vectors of the first-order resolution matrix each sum to unity for Gaussian distributed data with our choice of roughness function, so the amplitude of the recovered PPS is correctly scaled. This is a significant advantage of Tikhonov regularisation over most other inversion methods. The second-order resolution matrix vanishes for Gaussian data. The bias is given by

$$\text{Bias}(\hat{\mathbf{p}}_i) = \sum_j (R_{ij} - I_{ij}) p_{tj} + \frac{1}{2} \sum_{\alpha, \beta} X_{i\alpha\beta} N_{\alpha\beta} + \dots \quad (\text{A.22})$$

Thus good resolution and low bias are complementary. The contribution of the data noise to the frequentist covariance matrix is

$$\Sigma_{\text{F}} = \mathbf{S}(\boldsymbol{\theta}_{\text{t}}) \left[\sum_{\mathbb{Z}} \mathbf{W}_{\mathbb{Z}}^{\text{T}}(\boldsymbol{\theta}_{\text{t}}) \mathbf{N}_{\mathbb{Z}}^{-1} \mathbf{W}_{\mathbb{Z}}(\boldsymbol{\theta}_{\text{t}}) \right] \mathbf{S}^{\text{T}}(\boldsymbol{\theta}_{\text{t}}). \quad (\text{A.23})$$

Note that including additional data sets does not necessarily reduce Σ_{F} .

A.1 Bayesian inference

A different approach to inversion uses Bayesian inference, in which all unknown quantities are treated as random variables with probability distributions that represent the uncertainty about their values [163]. It is informative to compare Tikhonov regularisation with a Bayesian inversion method that incorporates our *a priori* knowledge (or bias) about the smoothness of the PPS. Consider a two-stage hierarchical Bayes model in which the prior distribution for the PPS $P(\mathbf{p}|\tilde{\lambda})$ is conditional on a hyperparameter $\tilde{\lambda}$ that itself has a hyperprior $P(\tilde{\lambda})$. According to Bayes' theorem, the joint posterior distribution $P(\mathbf{p}, \boldsymbol{\theta}, \tilde{\lambda}|\mathbf{d})$ of the PPS, background cosmological parameters, and the hyperparameter, given the data, is:

$$P(\mathbf{p}, \boldsymbol{\theta}, \tilde{\lambda}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{p}, \boldsymbol{\theta}|\mathbf{d}) P(\boldsymbol{\theta}) P(\mathbf{p}|\tilde{\lambda}) P(\tilde{\lambda})}{P(\mathbf{d})}. \quad (\text{A.24})$$

Here $P(\mathbf{d})$ is the prior distribution of the data and

$$P(\boldsymbol{\theta}) \propto \exp \left[-\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\text{T}} \mathbf{U}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \quad (\text{A.25})$$

is the prior distribution of the background parameters which is assumed to be Gaussian. A prior distribution which gives smoother spectra a higher prior probability is $P(\mathbf{p}|\tilde{\lambda}) \propto \exp[-\tilde{\lambda} \mathbf{R}(\mathbf{p})/2]$, in which case larger values of $\tilde{\lambda}$ penalise roughness more strongly. The marginalised posterior distribution of the PPS is obtained by integrating over the background parameters and the hyperparameter,

$$P(\mathbf{p}|\mathbf{d}) = \int P(\mathbf{p}, \boldsymbol{\theta}, \tilde{\lambda}|\mathbf{d}) d\boldsymbol{\theta} d\tilde{\lambda}. \quad (\text{A.26})$$

The maximum *a posteriori* estimate of the PPS, $\hat{\mathbf{p}}_{\text{MAP}}$, is defined as the mode of $P(\mathbf{p}|\mathbf{d})$. Assuming that the posterior can be approximated by a Gaussian, the mean of the distribution is $\langle \mathbf{p} \rangle = \hat{\mathbf{p}}_{\text{MAP}}$ and the Bayesian covariance matrix is

$$\langle (\mathbf{p} - \langle \mathbf{p} \rangle) (\mathbf{p} - \langle \mathbf{p} \rangle)^{\text{T}} \rangle = \mathbf{H}^{-1} \big|_{\hat{\mathbf{p}}_{\text{MAP}}}, \quad (\text{A.27})$$

where $H_{ij} \equiv -\partial^2 \ln P(\mathbf{p}|\mathbf{d}) / \partial p_i \partial p_j$.

Now, suppose that the value of $\tilde{\lambda}$ is known to be $\hat{\lambda}$ *a priori*, and that the uncertainty in the background parameters is negligible compared to the uncertainty in the data. Then $P(\tilde{\lambda}) = \delta(\tilde{\lambda} - \hat{\lambda})$ and $P(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$. With these priors maximising $P(\mathbf{p}|\mathbf{d})$ is equivalent to minimising $Q(\mathbf{p}, \mathbf{d}, \hat{\boldsymbol{\theta}}, \hat{\lambda})$, so that $\hat{\mathbf{p}}_{\text{MAP}} = \hat{\mathbf{p}}(\mathbf{d}, \hat{\boldsymbol{\theta}}, \hat{\lambda})$. Hence Tikhonov regularisation can be thought of as maximum *a posteriori* estimation with $\hat{\lambda} = \lambda$ and negligible uncertainty in the background parameters. In this case the Bayesian covariance matrix is denoted by Σ_{B} . The elements of the inverse covariance matrix are given by the Hessian of Q evaluated at $\mathbf{p} = \hat{\mathbf{p}}(\mathbf{d}, \hat{\boldsymbol{\theta}}, \lambda)$,

$$\Sigma_{\text{B}|ij}^{-1} \equiv \frac{1}{2} \frac{\partial^2 Q(\mathbf{p}, \mathbf{d}, \hat{\boldsymbol{\theta}}, \lambda)}{\partial p_i \partial p_j} \bigg|_{\hat{\mathbf{p}}} . \quad (\text{A.28})$$

For a given data set (or collection of data sets) Σ_{B} represents a lower bound on the Bayesian covariance matrix since additional uncertainty in the hyperparameter and the background parameters will increase the covariance. In general the Bayesian and frequentist covariance matrices Σ_{B} and Σ_{T} are different due to their fundamentally different statistical motivations. The elements of Σ_{B} and Σ_{T} decrease as λ increases, reflecting the improved stability of the solution. For data with Gaussian errors, $\Sigma_{\text{B}} = \mathbf{S}(\hat{\boldsymbol{\theta}})$; in this case adding additional data sets reduces Σ_{B} , in accordance with our expectation.

A.2 Performance statistics

In estimation theory, ‘loss’ functions are used to assess the performance of an estimator: they quantify the ‘loss’ incurred by using $\hat{\mathbf{p}}$ instead of the true \mathbf{p}_{t} . The most common loss function is the squared-error (SE),

$$\text{SE}(\hat{\mathbf{p}}) \equiv (\hat{\mathbf{p}} - \mathbf{p}_{\text{t}})^{\text{T}} (\hat{\mathbf{p}} - \mathbf{p}_{\text{t}}) . \quad (\text{A.29})$$

In what follows the SE is calculated from the wavenumber $k = 10^{-4} \text{ Mpc}^{-1}$ up to $k = 0.5 \text{ Mpc}^{-1}$. Alternatively, the predictive error (PE), defined as $\text{PE}(\hat{\mathbf{p}}) \equiv L(\hat{\mathbf{p}}, \mathbf{d}_{\text{t}})$, quantifies the ability of the estimated spectrum to predict the noise-free data $\mathbf{d}_{\text{t}} \equiv \mathbf{W}\mathbf{p}_{\text{t}}$. For multiple data sets with Gaussian likelihood functions we have,

$$\text{PE}(\hat{\mathbf{p}}) = \sum_{\mathbb{Z}} (\hat{\mathbf{p}} - \mathbf{p}_{\text{t}})^{\text{T}} \mathbf{W}_{\mathbb{Z}}^{\text{T}} \mathbf{N}_{\mathbb{Z}}^{-1} \mathbf{W}_{\mathbb{Z}} (\hat{\mathbf{p}} - \mathbf{p}_{\text{t}}) . \quad (\text{A.30})$$

This loss function emphasises errors on scales where the data more tightly constrain the PPS.

Taking the ensemble average of a loss function produces the associated ‘risk’ function. Here this gives the mean squared error (MSE):

$$\text{MSE}(\hat{\mathbf{p}}) \equiv \langle (\hat{\mathbf{p}} - \mathbf{p}_{\text{t}})^{\text{T}} (\hat{\mathbf{p}} - \mathbf{p}_{\text{t}}) \rangle = \langle \hat{\mathbf{p}} - \mathbf{p}_{\text{t}} \rangle^{\text{T}} \langle \hat{\mathbf{p}} - \mathbf{p}_{\text{t}} \rangle + \langle (\hat{\mathbf{p}} - \langle \hat{\mathbf{p}} \rangle)^{\text{T}} (\hat{\mathbf{p}} - \langle \hat{\mathbf{p}} \rangle) \rangle, \quad (\text{A.31})$$

which combines the bias squared and the variance of $\hat{\mathbf{p}}$. The mean predictive error (MPE) is defined as:

$$\text{MPE}(\hat{\mathbf{p}}) \equiv \langle L(\hat{\mathbf{p}}, \mathbf{d}_{\text{t}}) \rangle . \quad (\text{A.32})$$

B Data

The primordial curvature perturbations generate observable CMB temperature and electric (E-mode) polarisation anisotropies, characterised by the TT, TE and EE angular power spectra $\mathcal{C}_\ell^{\text{TT}}$, $\mathcal{C}_\ell^{\text{TE}}$ and $\mathcal{C}_\ell^{\text{EE}}$ respectively. These are written in terms of the spherical harmonic coefficients $a_{\ell m}^x$ as

$$\mathcal{C}_\ell^X \equiv \frac{\ell(\ell+1)}{2\pi} \langle a_{\ell m}^{x*} a_{\ell m}^{x'} \rangle, \quad (\text{B.1})$$

where $X = (xx') = \text{TT}, \text{TE}$ and EE . Throughout we consider only scalar perturbations so that magnetic (B -mode) polarisation is absent (as is expected for generic small-field non-fine-tuned inflationary potentials [9, 10]). Neglecting nonlinear secondary effects, each angular power spectrum is related to the PPS by the integral equation

$$\mathcal{C}_\ell^X = 4\pi \int_0^\infty \frac{dk}{k} T_\ell^{x*}(k) T_\ell^{x'}(k) \mathcal{P}_\mathcal{R}(k). \quad (\text{B.2})$$

Here the $T_\ell^x(k)$ are angular transfer functions dependent on the background parameters. It follows that $\mathbf{d}_\ell^X = \mathbf{s}_\ell^X + \mathbf{n}_\ell^X$, where \mathbf{d}_ℓ^X are the measured angular power spectra, $\mathbf{s}_\ell^X = \sum_i W_{\ell i}^X \mathbf{p}_i$ are the theoretical power spectra and \mathbf{n}_ℓ^X are the noise vectors for the TT, TE and EE spectra. The matrices \mathbf{W}^X which derive from the discretisation of eq.(B.2) are obtained using a modified version of the CAMB cosmological Boltzmann code [165, 166].

B.1 Experimental data sets

We present the data sets used in this paper. Throughout we treat the data in the manner recommended in the cited papers.

B.1.1 WMAP

The WMAP team provide a software package which returns the likelihood of a set of angular power spectra. The temperature likelihood function consists of three components, $L_{\text{TT}} = L_{\text{Gibbs}} + L_{\text{pTT}} + L_{\text{bps}}$, as explained below. For multipoles $\ell \leq 32$ the likelihood L_{Gibbs} is evaluated using a Blackwell-Rao estimator on samples drawn by Gibbs sampling from the joint posterior distribution of the power spectrum and the true sky signal [167]. A hybrid Gaussian and offset log-Gaussian fitting formula is used for $33 \leq \ell \leq 1000$ [168],

$$\begin{aligned} L_{\text{pTT}} = & \sum_{\ell\ell'} \frac{1}{3} (\mathbf{s}_\ell^{\text{TT}} - \mathbf{d}_\ell^{\text{TT}}) (\mathbf{N}^{\text{TT}})^{-1}_{\ell\ell'} (\mathbf{s}_\ell^{\text{TT}} - \mathbf{d}_\ell^{\text{TT}}) \\ & + \sum_{\ell\ell'} \frac{2}{3} \ln \left(\frac{\mathcal{S}_\ell^{\text{TT}}}{\mathcal{D}_\ell^{\text{TT}}} \right) \mathcal{S}_\ell^{\text{TT}} (\mathbf{N}^{\text{TT}})^{-1}_{\ell\ell'} \mathcal{S}_{\ell'}^{\text{TT}} \ln \left(\frac{\mathcal{S}_{\ell'}^{\text{TT}}}{\mathcal{D}_{\ell'}^{\text{TT}}} \right). \end{aligned} \quad (\text{B.3})$$

Here $\mathcal{S}_\ell^{\text{TT}} \equiv \mathbf{s}_\ell^{\text{TT}} + \mathcal{N}_\ell^{\text{TT}}$ and $\mathcal{D}_\ell^{\text{TT}} \equiv \mathbf{d}_\ell^{\text{TT}} + \mathcal{N}_\ell^{\text{TT}}$ where $\mathcal{N}_\ell^{\text{TT}}$ is the noise spectrum. The measurements $\mathbf{d}_\ell^{\text{TT}}$ were obtained using a pseudo- \mathcal{C}_ℓ estimator (given that only a fraction f_ℓ of the sky is unmasked). The diagonal elements of the inverse TT covariance matrix are

$$(\mathbf{N}^{\text{TT}})^{-1}_{\ell\ell} = \frac{(2\ell+1) (f_\ell^{\text{TT}})^2}{2 (\mathcal{S}_\ell^{\text{TT}})^2}. \quad (\text{B.4})$$

The WMAP results are also affected by uncertainties in both instrumental beam reconstruction and extragalactic point source subtraction [169]. This leads to the final component L_{bps} which we approximate by

$$L_{\text{bps}} = - \sum_{\ell\ell'} (s_{\ell}^{\text{TT}} - d_{\ell}^{\text{TT}}) \left(N^{\text{bps}} \right)_{\ell\ell'}^{-1} (s_{\ell}^{\text{TT}} - d_{\ell}^{\text{TT}}). \quad (\text{B.5})$$

The covariance matrix N^{bps} is evaluated for a fiducial TT spectrum (PPS with $n_s = 0.969$).

The polarisation likelihood is given by $L_{\text{pol}} = L_{\text{pix}} + L_{\text{pTE}}$ where L_{pix} is the likelihood given the TE/EE/BB data for $\ell \leq 23$ and L_{pTE} is the TE likelihood for $24 \leq \ell \leq 450$. The low- ℓ likelihood is evaluated directly from low resolution pixelised sky maps [170]. The high- ℓ likelihood is approximated by a Gaussian [168],

$$L_{\text{pTE}} = \sum_{\ell\ell'} (s_{\ell}^{\text{TE}} - d_{\ell}^{\text{TE}}) (N^{\text{TE}})_{\ell\ell'}^{-1} (s_{\ell}^{\text{TE}} - d_{\ell}^{\text{TE}}) - \sum_{\ell} \ln (N^{\text{TE}})_{\ell\ell}^{-1}. \quad (\text{B.6})$$

The diagonal elements of the inverse TE covariance matrix are

$$(N^{\text{TE}})_{\ell\ell}^{-1} = \frac{(2\ell + 1) (f_{\ell}^{\text{TE}})^2}{\mathcal{S}_{\ell}^{\text{TT,TE}} \mathcal{S}_{\ell}^{\text{EE}} + (s_{\ell}^{\text{TE}})^2}, \quad (\text{B.7})$$

where $\mathcal{S}_{\ell}^{\text{TT,TE}} \equiv s_{\ell}^{\text{TT}} + \mathcal{N}_{\ell}^{\text{TT,TE}}$, and $\mathcal{S}_{\ell}^{\text{EE}} \equiv s_{\ell}^{\text{EE}} + \mathcal{N}_{\ell}^{\text{EE}}$.

When first-order derivatives of the likelihood function are calculated in the BFGS minimisation, the full Gibbs sampling and pixel-based functions are used, but the pseudo- C_{ℓ} approximations L_{pTT} and L_{pTE} are employed to evaluate the Hessian matrix (see Appendix B.2).

B.1.2 Small angular scale CMB experiments

We combine the WMAP results with data from a number of ground-based small angular scale CMB experiments. These include the Very Small Array (VSA) [171], the Arcminute Cosmology Bolometer Array Receiver (ACBAR) [172], the Cosmic Background Imager (CBI) [173], BOOMERanG [174–176] and QUaD [177]. In particular the ‘NA pipeline’ BOOMERanG results and the ‘pipeline 1’ QUaD results are used. All the ground-based experiments quote bandpower estimates d_b^X due to their lower sky coverage. The predicted bandpowers are given by

$$s_b^X = \sum_{\ell} W_{b\ell}^X s_{\ell}^X = \sum_i T_{bi}^X p_i, \quad (\text{B.8})$$

where the rows of W^X are bandpower window functions and $T_{bi}^X \equiv \sum_{\ell} W_{b\ell}^X T_{\ell i}^X$. To ensure that the WMAP and small angular scale data sets are statistically independent, we use only TT bandpowers with mean multipoles of $\ell > 750$ and TE bandpowers with $\ell > 350$. In addition, bandpowers with $\ell > 2000$ are discarded in order to avoid scales which are affected by secondary processes such as gravitational lensing and the Sunyaev-Zeldovich effect.

The likelihood functions of the small angular scale data sets, L_{ss} , are either Gaussian or log-Gaussian in the bandpowers. The likelihoods are modified to include beam and calibration uncertainties and have the form

$$L_{\text{ss}} = \sum_{X, X', b, b'} \mathcal{Z}_b^X \left(V^{XX'} \right)_{bb'}^{-1} \mathcal{Z}_{b'}^{X'}. \quad (\text{B.9})$$

Here $\mathcal{Z}_b^X = s_b^X - d_b^X$ for a Gaussian bandpower and $\mathcal{Z}_b^X = \ln(\mathcal{S}_b^X/\mathcal{D}_b^X)$ for a log-Gaussian bandpower, with $\mathcal{S}_b^X \equiv s_b^X + \mathcal{N}_b^X$ and $\mathcal{D}_b^X \equiv d_b^X + \mathcal{N}_b^X$. The weight matrices $V^{XX'}$ are given by $(V^{XX'})_{bb'}^{-1} = D_b^X (\tilde{N}^{XX'})_{bb'}^{-1} D_{b'}^{X'}$ where for a Gaussian bandpower $D_b^X = 1$ and for a log-Gaussian bandpower $D_b^X = \mathcal{D}_b^X$. For a Gaussian beam of width θ the matrices $\tilde{N}^{XX'}$ are related to the bandpower covariance matrices $N^{XX'}$ by

$$\tilde{N}_{bb'}^{XX'} = N_{bb'}^{XX'} + \sigma_{\text{cal}}^2 d_b^X d_{b'}^{X'} + 2\sigma_\theta \theta \ell_b^X \ell_{b'}^{X'} d_b^X d_{b'}^{X'}, \quad (\text{B.10})$$

where σ_{cal} is the calibration error, σ_θ is the error in the beam width and ℓ_b^X is the mean multipole of the bandpower labelled by X and b . Using $\tilde{N}^{XX'}$ instead of $N^{XX'}$ in eq.(B.9) approximates marginalising over nuisance parameters associated with the calibration and beam uncertainties [178].

All the bandpowers of the VSA [171], ACBAR [172] and CBI [173] data sets consist of TT measurements alone and are taken to be log-Gaussian distributed. BOOMERanG and QUaD measure both temperature and polarisation anisotropies. For BOOMERanG the TT bandpowers [174] are log-Gaussian while the TE [175] and EE [176] bandpowers are Gaussian. The QUaD TT and EE bandpowers are log-Gaussian while the TE bandpowers are Gaussian [177]. We use the calibration and beam errors reported by the experimental teams.

B.1.3 SDSS

We use the power spectrum obtained from the luminous red galaxy (LRG) sample of the Sloan Digital Sky Survey fourth data release (SDSS-4) [179]. The power spectrum measurements extend into the small-scale nonlinear regime. We use the nonlinear modelling prescription of [179] to calculate the theoretical galaxy power spectrum $\mathcal{P}_g(k)$. Baryon acoustic oscillations in the power spectrum are suppressed on small-scales by nonlinear structure formation [180]. This effect is modelled by the power spectrum

$$\mathcal{P}_{\text{suposc}}(k) \equiv \left\{ e^{-\frac{1}{2}\left(\frac{k}{k_*}\right)^2} [T_{\text{osc}}^2(k) - T_{\text{noosc}}^2(k)] + T_{\text{noosc}}^2(k) \right\} \mathcal{P}_{\mathcal{R}}(k), \quad (\text{B.11})$$

which smoothly interpolates between the linear power spectrum, $\mathcal{P}_{\text{osc}}(k) = T_{\text{osc}}^2(k) \mathcal{P}_{\mathcal{R}}(k)$, on large scales and an oscillation-free spectrum, $\mathcal{P}_{\text{noosc}}(k) = T_{\text{noosc}}^2(k) \mathcal{P}_{\mathcal{R}}(k)$, on small scales. Here $T_{\text{osc}}(k)$ is the linear matter transfer function, which we obtain from CAMB, and $T_{\text{noosc}}(k)$ is the analytic transfer function without oscillations [181, 182]. The suppression occurs for wavenumbers $k \gtrsim k_*$ where

$$k_*^{-1} \equiv cD(1+f)^{1/3} \mathcal{P}_{0.05}^{1/2}. \quad (\text{B.12})$$

Here D is the linear growth factor, f is the linear growth rate (both evaluated at the mean redshift $z = 0.35$ of the LRG sample) and c is a constant equal to $6.19 h^{-1} \text{ Mpc}$. For a featureless PPS $\mathcal{P}_{0.05}$ would be equal to $\mathcal{P}_{\mathcal{R}}$ evaluated at $k = 0.05 \text{ Mpc}^{-1}$ but instead we take it to be the mean amplitude of the PPS over the interval $0.025 < k < 0.075 \text{ Mpc}^{-1}$. The Q model [183, 184] describes scale-dependent bias and the enhancement of small-scale power due to nonlinear evolution. It is used to obtain the galaxy power spectrum from $\mathcal{P}_{\text{suposc}}(k)$:

$$\mathcal{P}_g(k) = b_{\text{LRG}}^2 \frac{1 + Q_{\text{nl}} k^2}{1 + A_{\text{nl}} k} \mathcal{P}_{\text{suposc}}(k), \quad (\text{B.13})$$

where b_{LRG} is the scale-independent LRG bias and Q_{nl} and A_{nl} are empirical fitting parameters.

The measurement of the LRG power spectrum required a cosmological model to convert redshift observations into comoving distances. This was taken to be a fiducial Λ CDM model. To determine the power spectrum for a different model the fiducial data points can be rescaled rather than repeating the entire analysis for the new model. However it is more convenient to transform instead the theoretical power spectrum before comparing it with the fiducial data. The transformed spectrum is

$$\mathcal{P}_g^T(k) \equiv \frac{1}{\gamma^3} \mathcal{P}_g\left(\frac{k}{\gamma}\right), \quad (\text{B.14})$$

where the scaling factor γ is

$$\gamma \equiv \left(\frac{D_A^2 H_{\text{fid}}}{D_{A,\text{fid}}^2 H} \right)^{1/3}. \quad (\text{B.15})$$

Here D_A is the angular diameter distance and H is the Hubble parameter, both evaluated at redshift $z = 0.35$, and the subscript ‘fid’ refers to the fiducial model quantities. This transformation of the spectrum is equivalent to transforming the window functions $W_a(k)$ of the data points into $\mathcal{W}_a(k) \equiv \gamma^{-2} W_a(\gamma k)$. Hence the data points are given by

$$d_a = \int_0^\infty \mathcal{W}_a(k) \mathcal{P}_g(k) dk + n_a. \quad (\text{B.16})$$

Discretising the above integral gives $s_a = \sum_a W_{ai}^{\text{LRG}} p_i$. The likelihood function is

$$L_{\text{LRG}} = \sum_{a,a'} (s_a - d_a) \left(\tilde{N}^{\text{LRG}} \right)_{aa'}^{-1} (s_{a'} - d_{a'}) + \ln \det \tilde{N}^{\text{LRG}}, \quad (\text{B.17})$$

where

$$\tilde{N}_{aa'}^{\text{LRG}} = N_{aa'}^{\text{LRG}} + \sigma_{\text{cal}}^2 s_a s_{a'}. \quad (\text{B.18})$$

B.2 Likelihood function derivatives

Next we present the likelihood function derivatives used in the calculation of $\hat{\mathbf{p}}$. While the first-order derivatives must be exact, only approximations to the second-order derivatives are required by the BFGS algorithm used — the ones listed below were found to be satisfactory.

B.2.1 WMAP

The first-order derivative of the Gibbs sampler component of the WMAP temperature likelihood function is

$$\frac{\partial L_{\text{Gibbs}}}{\partial p_\alpha} = \sum_\ell W_{\ell\alpha}^{\text{TT}} \frac{\partial L_{\text{Gibbs}}}{\partial s_\ell^{\text{TT}}}. \quad (\text{B.19})$$

The derivative $\partial L_{\text{Gibbs}} / \partial s_\ell^{\text{TT}}$ is calculated numerically. The first-order derivative of the high- ℓ component is

$$\begin{aligned}
\frac{\partial L_{\text{pTT}}}{\partial p_\alpha} &= \frac{2}{3} \sum_{\ell\ell'} W_{\ell\alpha}^{\text{TT}} (N^{\text{TT}})^{-1}_{\ell\ell'} (s_{\ell'}^{\text{TT}} - d_{\ell'}^{\text{TT}}) + \frac{1}{3} \sum_{\ell} \frac{\partial (N^{\text{TT}})^{-1}_{\ell\ell}}{\partial p_\alpha} (s_{\ell}^{\text{TT}} - d_{\ell}^{\text{TT}})^2 \\
&+ \frac{4}{3} \sum_{\ell\ell'} \ln \left(\frac{\mathcal{S}_{\ell}^{\text{TT}}}{\mathcal{D}_{\ell}^{\text{TT}}} \right) W_{\ell\alpha}^{\text{TT}} (N^{\text{TT}})^{-1}_{\ell\ell'} \mathcal{S}_{\ell'}^{\text{TT}} \ln \left(\frac{\mathcal{S}_{\ell'}^{\text{TT}}}{\mathcal{D}_{\ell'}^{\text{TT}}} \right) \\
&+ \frac{4}{3} \sum_{\ell\ell'} W_{\ell\alpha}^{\text{TT}} (N^{\text{TT}})^{-1}_{\ell\ell'} \mathcal{S}_{\ell'}^{\text{TT}} \ln \left(\frac{\mathcal{S}_{\ell'}^{\text{TT}}}{\mathcal{D}_{\ell'}^{\text{TT}}} \right) \\
&+ \frac{2}{3} \sum_{\ell} \frac{\partial (N^{\text{TT}})^{-1}_{\ell\ell}}{\partial p_\alpha} (\mathcal{S}_{\ell}^{\text{TT}})^2 \left[\ln \left(\frac{\mathcal{S}_{\ell}^{\text{TT}}}{\mathcal{D}_{\ell}^{\text{TT}}} \right) \right]^2,
\end{aligned} \tag{B.20}$$

where

$$\frac{\partial (N^{\text{TT}})^{-1}_{\ell\ell}}{\partial p_\alpha} = - \frac{(2\ell + 1) (f_{\ell}^{\text{TT}})^2 W_{\ell\alpha}^{\text{TT}}}{(\mathcal{S}_{\ell}^{\text{TT}})^3}. \tag{B.21}$$

The first-order derivative of the beam and point source component is

$$\frac{\partial L_{\text{bps}}}{\partial p_\alpha} = -2 \sum_{\ell\ell'} W_{\ell\alpha}^{\text{TT}} (N^{\text{bps}})^{-1}_{\ell\ell'} (s_{\ell'}^{\text{TT}} - d_{\ell'}^{\text{TT}}). \tag{B.22}$$

The second-order derivative of the WMAP TT likelihood is approximated by

$$\frac{\partial^2 L_{\text{TT}}}{\partial p_\alpha \partial p_\beta} \simeq 2 \sum_{\ell\ell'} W_{\ell\alpha}^{\text{TT}} (N^{\text{TT}})^{-1}_{\ell\ell'} W_{\ell\beta}^{\text{TT}}. \tag{B.23}$$

The first-order derivative of the low- ℓ pixel-based component of the WMAP polarisation likelihood is

$$\frac{\partial L_{\text{pix}}}{\partial p_\alpha} = \sum_{\ell X} W_{\ell\alpha}^X \frac{\partial L_{\text{pix}}}{\partial s_{\ell}^X}, \tag{B.24}$$

where $\partial L_{\text{pix}}/\partial s_{\ell}^X$ is evaluated numerically. The first-order derivative of the high- ℓ component of the polarisation likelihood is

$$\begin{aligned}
\frac{\partial L_{\text{pTE}}}{\partial p_\alpha} &= 2 \sum_{\ell\ell'} W_{\ell\alpha}^{\text{TE}} (N^{\text{TE}})^{-1}_{\ell\ell'} (s_{\ell'}^{\text{TE}} - d_{\ell'}^{\text{TE}}) + \sum_{\ell} \frac{\partial (N^{\text{TE}})^{-1}_{\ell\ell}}{\partial p_\alpha} (s_{\ell}^{\text{TE}} - d_{\ell}^{\text{TE}})^2 \\
&- \sum_{\ell} N_{\ell\ell}^{\text{TE}} \frac{\partial (N^{\text{TE}})^{-1}_{\ell\ell}}{\partial p_\alpha},
\end{aligned} \tag{B.25}$$

where

$$\frac{\partial (N^{\text{TE}})^{-1}_{\ell\ell}}{\partial p_\alpha} = - \frac{(2\ell + 1) (f_{\ell}^{\text{TT}})^2 \left(W_{\ell\alpha}^{\text{EE}} \mathcal{S}_{\ell}^{\text{TT,TE}} + W_{\ell\alpha}^{\text{TT}} \mathcal{S}_{\ell}^{\text{EE}} + 2 W_{\ell\alpha}^{\text{TE}} s_{\ell}^{\text{TE}} \right)}{2 (\mathcal{S}_{\ell}^{\text{TT}})^2 \left[\mathcal{S}_{\ell}^{\text{TT,TE}} \mathcal{S}_{\ell}^{\text{EE}} + (s_{\ell}^{\text{TE}})^2 \right]}. \tag{B.26}$$

The second-order derivative of the WMAP polarisation likelihood is approximated by

$$\frac{\partial^2 L_{\text{pol}}}{\partial p_\alpha \partial p_\beta} \simeq 2 \sum_{\ell\ell'} W_{\ell\alpha}^{\text{TE}} (N^{\text{TE}})^{-1}_{\ell\ell'} W_{\ell\beta}^{\text{TE}} + \sum_{\ell} (N_{\ell\ell}^{\text{TE}})^2 \frac{\partial (N^{\text{TE}})^{-1}_{\ell\ell}}{\partial p_\alpha} \frac{\partial (N^{\text{TE}})^{-1}_{\ell\ell}}{\partial p_\beta}. \tag{B.27}$$

The first-order derivative of the low- ℓ one-dimensional Wishart likelihood used with mock WMAP TT data alone is

$$\frac{\partial L_{W1}}{\partial p_\alpha} = 2 \sum_\ell W_{\ell\alpha}^{\text{TT}} (N^{\text{TT}})^{-1}_{\ell\ell} (s_\ell^{\text{TT}} - d_\ell^{\text{TT}}). \quad (\text{B.28})$$

The first-order derivative of the low- ℓ two-dimensional Wishart likelihood used with mock WMAP temperature and polarisation data is (see eq. B.40)

$$\begin{aligned} \frac{\partial L_{W2}}{\partial p_\alpha} = \sum_\ell n_{\text{dof}} \left\{ \frac{(\mathcal{S}_\ell^{\text{EE}} + \mathcal{D}_\ell^{\text{EE}}) W_{\ell\alpha}^{\text{TT}} + (\mathcal{S}_\ell^{\text{TT}} + \mathcal{D}_\ell^{\text{TT}}) W_{\ell\alpha}^{\text{EE}} - 2(s_\ell^{\text{TE}} + d_\ell^{\text{TE}}) W_{\ell\alpha}^{\text{TE}}}{\mathcal{S}_\ell^{\text{TT}} \mathcal{S}_\ell^{\text{EE}} - (s_\ell^{\text{TE}})^2} \right. \\ \left. + \frac{\mathcal{S}_\ell^{\text{TT}} \mathcal{D}_\ell^{\text{EE}} + \mathcal{D}_\ell^{\text{TT}} \mathcal{S}_\ell^{\text{EE}} - 2s_\ell^{\text{TE}} d_\ell^{\text{TE}}}{[\mathcal{S}_\ell^{\text{TT}} \mathcal{S}_\ell^{\text{EE}} - (s_\ell^{\text{TE}})^2]^2} (2s_\ell^{\text{TE}} W_{\ell\alpha}^{\text{TE}} - \mathcal{S}_\ell^{\text{EE}} W_{\ell\alpha}^{\text{TT}} - \mathcal{S}_\ell^{\text{TT}} W_{\ell\alpha}^{\text{EE}}) \right\} \quad (\text{B.29}) \end{aligned}$$

B.2.2 Small angular scale CMB experiments

The first-order derivative of the small angular scale CMB likelihood is

$$\frac{\partial L_{\text{ss}}}{\partial p_\alpha} = 2 \sum_{X, X', b, b'} \mathcal{Z}_b^X (V^{XX'})^{-1}_{bb'} \frac{\partial \mathcal{Z}_{b'}^{X'}}{\partial p_\alpha}, \quad (\text{B.30})$$

where

$$\frac{\partial \mathcal{Z}_b^X}{\partial p_\alpha} = \begin{cases} T_{b\alpha}^X, & \text{Gaussian bandpower,} \\ T_{b\alpha}^X / \mathcal{S}_b^X, & \text{log-Gaussian bandpower.} \end{cases} \quad (\text{B.31})$$

The second-order derivative of the small angular scale CMB likelihood is approximated by

$$\frac{\partial^2 L_{\text{ss}}}{\partial p_\alpha \partial p_\beta} \simeq 2 \sum_{X, X', b, b'} \frac{\partial \mathcal{Z}_b^X}{\partial p_\alpha} (V^{XX'})^{-1}_{bb'} \frac{\partial \mathcal{Z}_{b'}^{X'}}{\partial p_\beta}. \quad (\text{B.32})$$

B.2.3 SDSS

The SDSS-4 LRG likelihood can be rewritten as [178]

$$L_{\text{LRG}} = \sum_{a, a'} (s_a - d_a) (N^{\text{LRG}})^{-1}_{aa'} (s_{a'} - d_{a'}) - \frac{A^2}{C} + \ln(1 + \sigma_{\text{cal}}^2 B) + \ln \det N^{\text{LRG}}, \quad (\text{B.33})$$

where

$$A \equiv \sum_{a, a'} s_a (N^{\text{LRG}})^{-1}_{aa'} (s_{a'} - d_{a'}), \quad B \equiv \sum_{a, a'} s_a (N^{\text{LRG}})^{-1}_{aa'} s_{a'}, \quad C \equiv \frac{1 + \sigma_{\text{cal}}^2 B}{\sigma_{\text{cal}}^2}. \quad (\text{B.34})$$

Hence the first-order derivative of SDSS-4 LRG likelihood is

$$\begin{aligned} \frac{\partial L_{\text{LRG}}}{\partial p_\alpha} = 2 \sum_{a, a'} W_{a\alpha}^{\text{LRG}} (N^{\text{LRG}})^{-1}_{aa'} (s_{a'} - d_{a'}) - \frac{2A}{C} \sum_{a, a'} W_{a\alpha}^{\text{LRG}} (N^{\text{LRG}})^{-1}_{aa'} (2s_{a'} - d_{a'}) \\ + \frac{A^2 + C}{C^2} \sum_{a, a'} W_{a\alpha}^{\text{LRG}} (N^{\text{LRG}})^{-1}_{aa'} s_{a'}. \end{aligned} \quad (\text{B.35})$$

The second-order derivative of the SDSS-4 LRG likelihood is approximated by

$$\frac{\partial^2 L_{\text{LRG}}}{\partial p_\alpha \partial p_\beta} = 2 \sum_{a, a'} W_{a\alpha}^{\text{LRG}} (\tilde{N}^{XX'})^{-1}_{aa'} W_{a'\alpha}^{\text{LRG}}. \quad (\text{B.36})$$

B.3 Test spectra

We test our method on four examples of primordial power spectra, labelled spectra A to D, to determine how successfully they are reconstructed from mock data sets. Spectrum A is the best-fit power-law spectrum found by the WMAP team, with a spectral index of $n_s = 0.963$ for the WMAP 7-year data [167].

Spectrum B has the form of a ‘step’ located at $k = k_0$ with superimposed damped high frequency oscillations. It is generated by a toy model of inflation with a sharp, discontinuous change in the slope of the inflaton potential from a value of V_-' to V_+' ; The spectrum is obtained analytically as [30]:

$$\mathcal{P}_{\mathcal{R}}(k) = \mathcal{P}_{\mathcal{R}0}(k) \left\{ 1 - 3(r-1) \frac{1}{y} \left[\left(1 - \frac{1}{y^2}\right) \sin 2y + \frac{2}{y} \cos 2y \right] + \frac{9}{2}(r-1)^2 \frac{1}{y^2} \left(1 + \frac{1}{y^2}\right) \left[1 + \frac{1}{y^2} + \left(1 - \frac{1}{y^2}\right) \cos 2y - \frac{2}{y} \sin 2y \right] \right\}. \quad (\text{B.37})$$

Here $y \equiv k/k_0$, $r \equiv V_-'/V_+'$ and $\mathcal{P}_{\mathcal{R}0}$ is the underlying power spectrum of the model, which is taken to be a power-law with $n_s = 0.963$. We set $k_0 = 2 \times 10^{-4} \text{ Mpc}^{-1}$ and $r = 0.5$.

Spectrum C arises from an inflation model in which a scalar field coupled to the inflaton performs damped oscillations about the origin during inflation, beginning when the Hubble parameter falls below the mass of the field. This causes the inflaton mass to decrease from much greater than, to much less than, the Hubble parameter and leads to an infrared cutoff in the spectrum. On intermediate scales the spectrum has a number of irregular features due to a parametric resonance effect. The parameters of the model [61] have the values $q_{\text{in}} = 10$ and $m/H = 2\sqrt{3}$.

Spectrum D is that of the ‘CHDM bump’ model. This Einstein-de Sitter (E-deS) universe has about 80% cold, 10% baryonic and 10% hot dark matter (in the form of 3 mass-degenerate neutrinos with mass $\sim 0.5 \text{ eV}$) and fits both the CMB and large-scale structure data *without* need for a dominant dark energy component, if there a suitably located ‘bump’ in the power spectrum [14, 15]. This can result e.g. from multiple inflation in $N = 1$ supergravity wherein ‘flat direction’ scalar fields undergo spontaneous symmetry breaking phase transitions, as the universe cools during inflation [22]. Before each phase transition the field is confined at the origin by a thermal barrier, but when the barrier disappears a Hubble-induced mass correction drives the field rapidly to the global minimum of its potential which is determined by stabilising higher dimensional operators. The gravitational coupling between the inflaton and these flat direction fields causes the inflaton mass to jump at each phase transition and thus produces characteristic features in the power spectrum [23]. Here the bump has parameters $k_1 = 0.03 \text{ Mpc}^{-1}$ and $k_2 = 0.08 \text{ Mpc}^{-1}$. It arises from 2 phase transitions which begin about an e-fold of inflation apart and cause fractional changes in the inflaton mass-squared of $\Delta m_1^2 = 0.151$ and $\Delta m_2^2 = 0.272$.¹³

For spectra A to C the background cosmology is taken to be flat Λ CDM with $\omega_b = 0.0223$, $\omega_c = 0.104$, $h = 0.73$, $\tau = 0.088$ and $b_{\text{LRG}} = 1.9$. The flat CHDM model of spectrum D has $\omega_b = 0.0165$, neutrino fraction $f_\nu = 0.127$, $h = 0.42$, $\tau = 0.074$ and $b_{\text{LRG}} = 2.1$.

B.4 Mock data

From each test spectrum we generate mock data using the following method. In the full sky limit the observed CMB multipoles are known to follow a Wishart distribution, being a

¹³This bump is located at a slightly higher wavenumber than that of the best-fit model in [15] to better illustrate the effects of including small angular scale data in the inversion.

sum of squared Gaussian spherical harmonic coefficients [185]. When mock WMAP TT data alone is required, we follow [142] and draw samples from the distribution

$$-2 \ln P_{W1}(\{d_\ell^{\text{TT}}\} | \{s_\ell^{\text{TT}}\}) = \sum_\ell n_{\text{dof}} \left(\ln \mathcal{S}_\ell^{\text{TT}} - \frac{n_{\text{dof}} - 2}{n_{\text{dof}}} \ln \mathcal{D}_\ell^{\text{TT}} + \frac{\mathcal{D}_\ell^{\text{TT}}}{\mathcal{S}_\ell^{\text{TT}}} \right), \quad (\text{B.38})$$

up to a constant, where $n_{\text{dof}} = (2\ell + 1) f_\ell^{\text{TT}}$. This is a one-dimensional Wishart distribution (i.e. a χ^2 distribution). The samples have the statistical properties:

$$\langle d_\ell^{\text{TT}} - s_\ell^{\text{TT}} \rangle = 0, \quad \langle (d_\ell^{\text{TT}} - s_\ell^{\text{TT}})^2 \rangle = \frac{2 (\mathcal{S}_\ell^{\text{TT}})^2}{(2\ell + 1) (f_\ell^{\text{TT}})^2}. \quad (\text{B.39})$$

Mock WMAP TT, TE and EE data is created by sampling the two-dimensional Wishart distribution

$$\begin{aligned} -2 \ln P_{W2}(\{d_\ell^{\text{X}}\} | \{s_\ell^{\text{X}}\}) = \sum_\ell n_{\text{dof}} \left\{ \ln \left[\mathcal{S}_\ell^{\text{TT}} \mathcal{S}_\ell^{\text{EE}} - (s_\ell^{\text{TE}})^2 \right] - \frac{n_{\text{dof}} - 3}{n_{\text{dof}}} \ln \left[\mathcal{D}_\ell^{\text{TT}} \mathcal{D}_\ell^{\text{EE}} - (d_\ell^{\text{TE}})^2 \right] \right. \\ \left. + \frac{\mathcal{S}_\ell^{\text{TT}} \mathcal{D}_\ell^{\text{EE}} + \mathcal{D}_\ell^{\text{TT}} \mathcal{S}_\ell^{\text{EE}} - 2s_\ell^{\text{TE}} d_\ell^{\text{TE}}}{\mathcal{S}_\ell^{\text{TT}} \mathcal{S}_\ell^{\text{EE}} - (s_\ell^{\text{TE}})^2} \right\}, \end{aligned} \quad (\text{B.40})$$

where $\mathcal{D}_\ell^{\text{EE}} \equiv d_\ell^{\text{EE}} + \mathcal{N}_\ell^{\text{EE}}$. In addition to eq.(B.39), the data satisfies

$$\langle d_\ell^{\text{X}} - s_\ell^{\text{X}} \rangle = 0, \quad \langle (d_\ell^{\text{TE}} - s_\ell^{\text{TE}})^2 \rangle = \frac{\mathcal{S}_\ell^{\text{TT}} \mathcal{S}_\ell^{\text{EE}} + (s_\ell^{\text{TE}})^2}{(2\ell + 1) (f_\ell^{\text{TT}})^2}, \quad \langle (d_\ell^{\text{EE}} - s_\ell^{\text{EE}})^2 \rangle = \frac{2 (\mathcal{S}_\ell^{\text{EE}})^2}{(2\ell + 1) (f_\ell^{\text{TT}})^2}, \quad (\text{B.41})$$

$$\langle (d_\ell^{\text{TT}} - s_\ell^{\text{TT}}) (d_\ell^{\text{TE}} - s_\ell^{\text{TE}}) \rangle = \frac{2 \mathcal{S}_\ell^{\text{TT}} s_\ell^{\text{TE}}}{(2\ell + 1) (f_\ell^{\text{TT}})^2}, \quad \langle (d_\ell^{\text{TT}} - s_\ell^{\text{TT}}) (d_\ell^{\text{EE}} - s_\ell^{\text{EE}}) \rangle = \frac{2 (s_\ell^{\text{TE}})^2}{(2\ell + 1) (f_\ell^{\text{TT}})^2}, \quad (\text{B.42})$$

$$\langle (d_\ell^{\text{TE}} - s_\ell^{\text{TE}}) (d_\ell^{\text{EE}} - s_\ell^{\text{EE}}) \rangle = \frac{2 s_\ell^{\text{TE}} \mathcal{S}_\ell^{\text{EE}}}{(2\ell + 1) (f_\ell^{\text{TT}})^2}, \quad (\text{B.43})$$

The variance of the mock WMAP data points thus matches the diagonal elements of the covariance matrices given in Sec. B.1.1. Although this approach neglects correlations between different multipoles induced when working on a cut sky, we do not expect this to significantly affect our results due to the high sky coverage ($f_\ell \sim 0.8$) of the WMAP experiment.

The quantities \mathcal{Z}_b^{X} (eq. B.9) were shown to be Gaussian distributed [186]. Thus to simulate ground-based (small angular scale) CMB experiments, \mathcal{Z}_b^{X} are drawn from a multivariate Gaussian distribution with vanishing mean and covariance matrix $\mathbf{V}^{\text{XX}'}$, where $\tilde{\mathbf{N}}^{\text{XX}'}$ is now calculated using theoretical bandpowers instead of data bandpowers. Similarly mock SDSS-4 LRG data is produced by contaminating \mathbf{s} from the test spectrum with Gaussian noise with covariance matrix $\tilde{\mathbf{N}}^{\text{LRG}}$.

Rather than use the full WMAP likelihood functions, we adopt less computationally expensive approximations when inverting the mock data. These are $L_{\text{TT}} \simeq L_{\text{W1}} + L_{\text{pTT}}$ for the WMAP temperature data alone and $L_{\text{TT}} + L_{\text{pol}} \simeq L_{\text{W2}} + L_{\text{pTT}} + L_{\text{pTE}}$ for the WMAP temperature and polarisation data. Here $L_{\text{W1,2}} \equiv -2 \ln P_{\text{W1,2}}$ are the likelihood functions corresponding to the Wishart distributions of eq.(B.38) and eq.(B.40). They are used for $\ell \leq 23$ while L_{pTT} and L_{pTE} are employed for $24 \leq \ell \leq 1000$ and $24 \leq \ell \leq 450$ respectively. Furthermore following [142] we neglect the off-diagonal elements of the matrices $(\mathbf{N}^{\text{TT}})^{-1}$.

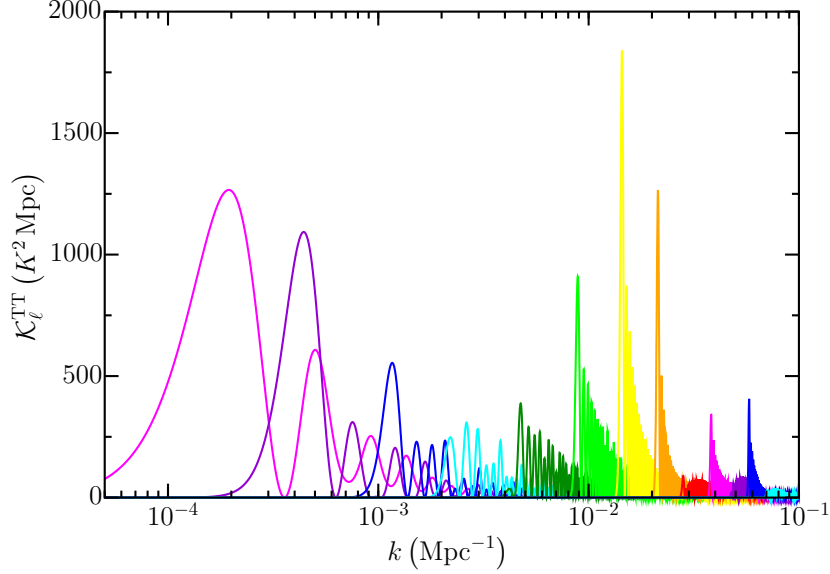


Figure 12. Integral kernels $\mathcal{K}_\ell^{\text{TT}}(k)$ (eq. 2.2) for $\ell = 2, 5, 15, 30, 60$ (corresponding to the Sachs-Wolfe plateau), $\ell = 120, 200, 300$ (first acoustic peak), $\ell = 400$ (first trough), $\ell = 540$ (second peak), $\ell = 680$ (second trough), $\ell = 820$ (third peak) and $\ell = 999$ (third trough), from left to right.

C Validation of the inversion method

C.1 The integral and resolution kernels

The integral kernels $\mathcal{K}_\ell^{\text{TT}}(k)$ for the CMB temperature angular power spectrum are fundamental to the properties of the inversion. Some examples of the kernels are presented in Fig. 12. In general the kernels become narrower for higher multipoles, so that convolution with them involves less smoothing and the relation between k -space and ℓ -space becomes more direct on small scales. However, the kernels which correspond to troughs in the angular power spectrum are broader than those associated with neighbouring acoustic peaks.

The estimated PPS $\hat{\mathcal{P}}_{\mathcal{R}}(k_0)$ cannot resolve features in the true PPS $\mathcal{P}_{\mathcal{R}}(k)$ which are narrower than the resolution kernel $R(k_0, k)$. Fig. 13 shows the first-order kernels (eq. A.6) for some selected values of the target wavenumber k_0 . Ideally the resolution kernel $R(k_0, k)$ would be sharply peaked at $k = k_0$ and negligible everywhere else. The resolution kernels depend on both the integral kernels and on the noise in the data. Consider first the resolution kernels for the WMAP-5 TT data alone. The resolution kernels are broadest on large scales due to the wide integral kernels at low ℓ , as well as the uncertainty in the data caused by cosmic variance. The resolution kernels narrow with higher wavenumber until the greatest resolution is attained at $k \simeq 0.02 \text{ Mpc}^{-1}$. This corresponds to $\ell \simeq 280$ on the right hand side of the first acoustic peak where the integral kernels are sharply peaked and the WMAP error is smallest. The resolution kernels at $k_0 = 0.03$ and 0.05 Mpc^{-1} are misshapen, reflecting the broad integral kernels of the first and second troughs in the TT angular power spectrum. The more strongly defined kernels at $k_0 = 0.04$ and 0.06 Mpc^{-1} correspond to the second and third acoustic peaks. The large WMAP measurement errors at high multipoles decrease the resolution on small scales. The resolution kernels are mostly centred closely on the target wavenumbers, showing that the reconstructed PPS is meaningful. However, the $k_0 =$

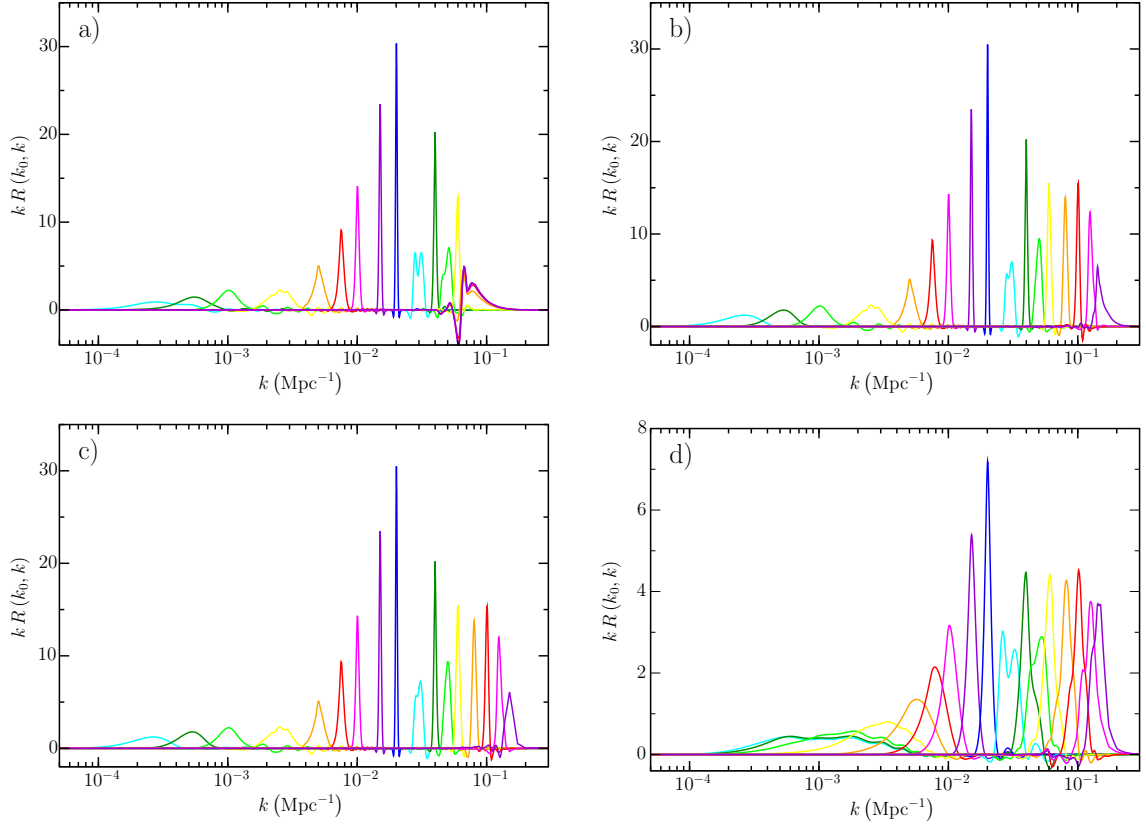


Figure 13. First-order resolution kernels $R(k_0, k)$ (eq. A.6) for $k_0 = 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 2.5 \times 10^{-3}, 5 \times 10^{-3}, 7.5 \times 10^{-3}, 0.01, 0.015, 0.02, 0.03, 0.04, 0.05, 0.06, 0.08, 0.1, 0.125, 0.15 \text{ Mpc}^{-1}$. The top left hand plot a) shows the kernels for the WMAP-5 TT data alone. The top right hand plot b) shows the kernels for the WMAP-5 TT, TE and EE data together with the small-scale CMB data. The bottom left hand plot c) shows the kernels for the WMAP-5, small-scale CMB and SDSS-4 LRG data. The bottom right hand plot d) is the same as the bottom left but with $\lambda = 1000$ instead of $\lambda = 10$.

10^{-4} Mpc^{-1} kernel is centred at $k \simeq 3 \times 10^{-4} \text{ Mpc}^{-1}$ instead. This is a consequence of the fact that at the very lowest wavenumbers, the estimated PPS is an extrapolation from higher wavenumbers. A similar phenomenon occurs at the very highest wavenumbers.

Adding the WMAP-5 polarisation data slightly improves the resolution on large scales. The $k_0 = 10^{-4} \text{ Mpc}^{-1}$ and $k_0 = 5 \times 10^{-4} \text{ Mpc}^{-1}$ kernels become slightly better localised. The resolution at high wavenumbers is greatly increased by the small angular scale CMB data. The kernels for $k_0 = 0.08, 0.1$ and 0.125 Mpc^{-1} which were close together with the WMAP-5 TT data become well separated and strongly peaked. The SDSS-4 LRG data improves the $k_0 = 0.03 \text{ Mpc}^{-1}$ and $k_0 = 0.05 \text{ Mpc}^{-1}$ kernels, causing them to become more peaked. Finally, Fig. 13 shows that increasing the regularisation parameter broadens the resolution kernels and decreases the resolution. The kernels on large scales become almost identical, which indicates that increased extrapolation occurs in the inversion.

The second-order resolution kernel (eq. A.8) is shown in Fig. 14. For $\lambda = 10$ the kernel is centred approximately at $k_1 = k_2 = k_0$. The kernel is broader for low k_0 and more localised for higher k_0 due to the narrower integral kernels at higher multipoles. For $\lambda = 1000$ the $k_0 = 10^{-3} \text{ Mpc}^{-1}$ kernel is centred at $k_1 = k_2 \simeq 4 \times 10^{-3} \text{ Mpc}^{-1}$. This is similar to the way

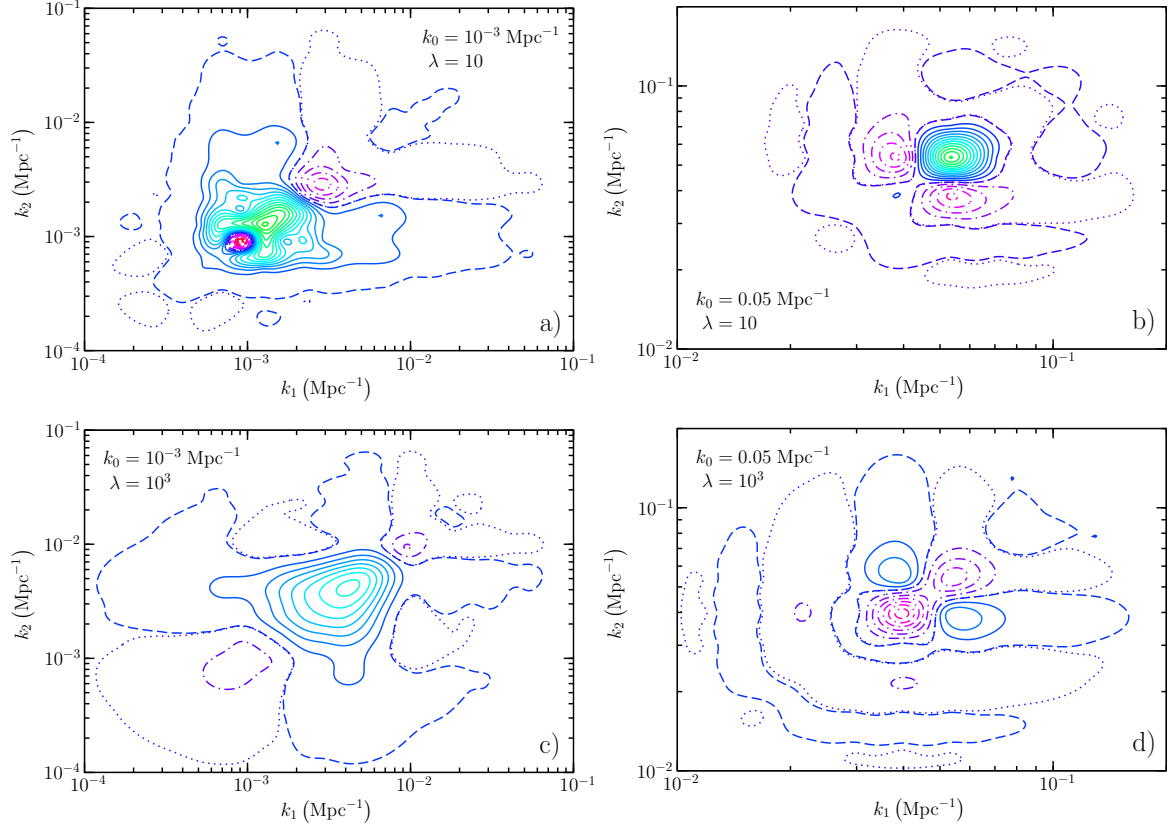


Figure 14. Second-order resolution kernels $Y(k_0; k_1, k_2)$ (eq. A.8) for the WMAP-5 TT data. The two left hand plots are for $k_0 = 10^{-3} \text{ Mpc}^{-1}$ and the right hand plots are for $k_0 = 0.05 \text{ Mpc}^{-1}$. The upper plots are for $\lambda = 10$ and the lower plots are for $\lambda = 1000$. The dashed contours are at $Y = 10^{-3}$ while the dotted contours are at $Y = -10^{-3}$. The solid contours are at $Y = 0.05, 0.1, 0.15, \dots$ and the dot-dashed contours are at $Y = -0.05, -0.1, -0.15, \dots$

in which the peak of $R(k_0; k_1)$, located at $k_1 = k_0$ for low λ , shifts as λ increases towards higher k_1 for low k_0 and lower k_1 for high k_0 . The second-order kernel is broader and lower in magnitude for higher λ .

C.2 Test spectra results

We start by inverting *mock* WMAP-5 TT “data” without noise. The reconstructed spectra resemble versions of the test spectra smoothed towards flatness, as seen in Fig. 15. The degree of smoothing increases with λ , in accordance with expectation. Since the resolution kernels integrate to unity, there are no spurious vertical scalings of the reconstructed spectra. For small values of λ the test spectra are recovered with impressive accuracy. (However since we are using first-order Tikhonov regularisation based on a H-Z spectrum, the *tilted* spectrum A is not fully recovered both at high and low k .) The loss of resolution with increased regularisation is clearly apparent for spectrum B, where the oscillations become increasingly smoothed with larger values of λ . For $\lambda = 0.1$ the oscillations are recovered least successfully at $k \simeq 3 \times 10^{-3} \text{ Mpc}^{-1}$ because the width of the oscillations there is comparable to that of the resolution kernels; at higher and lower k the oscillations are broader than the kernels. At very high and low k where the data is insensitive to the PPS (since the integral kernels

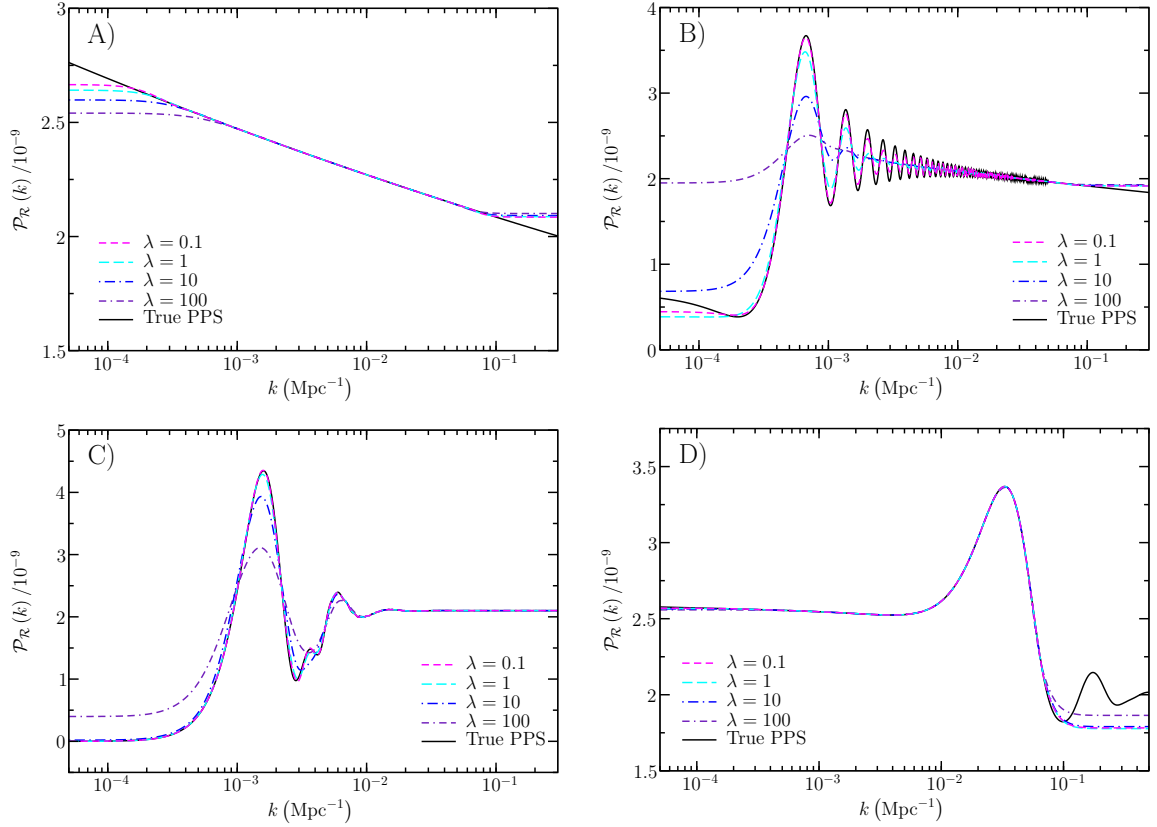


Figure 15. Spectra recovered from noise-free mock WMAP-5 TT “data” with regularisation parameter $\lambda = 0.1, 1, 10$ and 100 . Results are shown for spectra A to D (see Sec. B.3).

vanish), the estimated PPS is extrapolated horizontally from lower and higher k respectively. The amount of extrapolation increases with λ , as can be seen most clearly for spectrum A on large scales. Note that the WMAP data do not extend to small scales beyond $k \sim 0.1 \text{ Mpc}^{-1}$, hence the wiggle in spectrum D cannot be recovered there for any λ .

Fig. 16 shows typical estimated spectra from mock WMAP-5 TT data, but now including noise. The presence of noise greatly reduces the accuracy with which the PPS can be found. The noise-induced oscillatory features are broadest on large scales due to the wide integral kernels at low ℓ , and have high amplitude due to cosmic variance. The increased suppression of the noise-induced features can clearly be seen as λ is increased. Note that the best reconstructions are obtained using *different* values of λ for different test spectra.

We applied Tikhonov regularisation to 10^5 mock WMAP-5 TT data realisations and took the mean of the estimated spectra, with the results displayed in Fig. 17. On small and intermediate scales the mean reconstructions are equivalent to reconstructions obtained from noise-free data. However, on large scales this is not the case for small values of λ : the mean reconstructions have less power and are clearly biased.

Fig. 18 shows the mean reconstructed spectra taking $\lambda = 100$, but using all four data combinations to illustrate the gain when non-WMAP data is available, especially at high k .

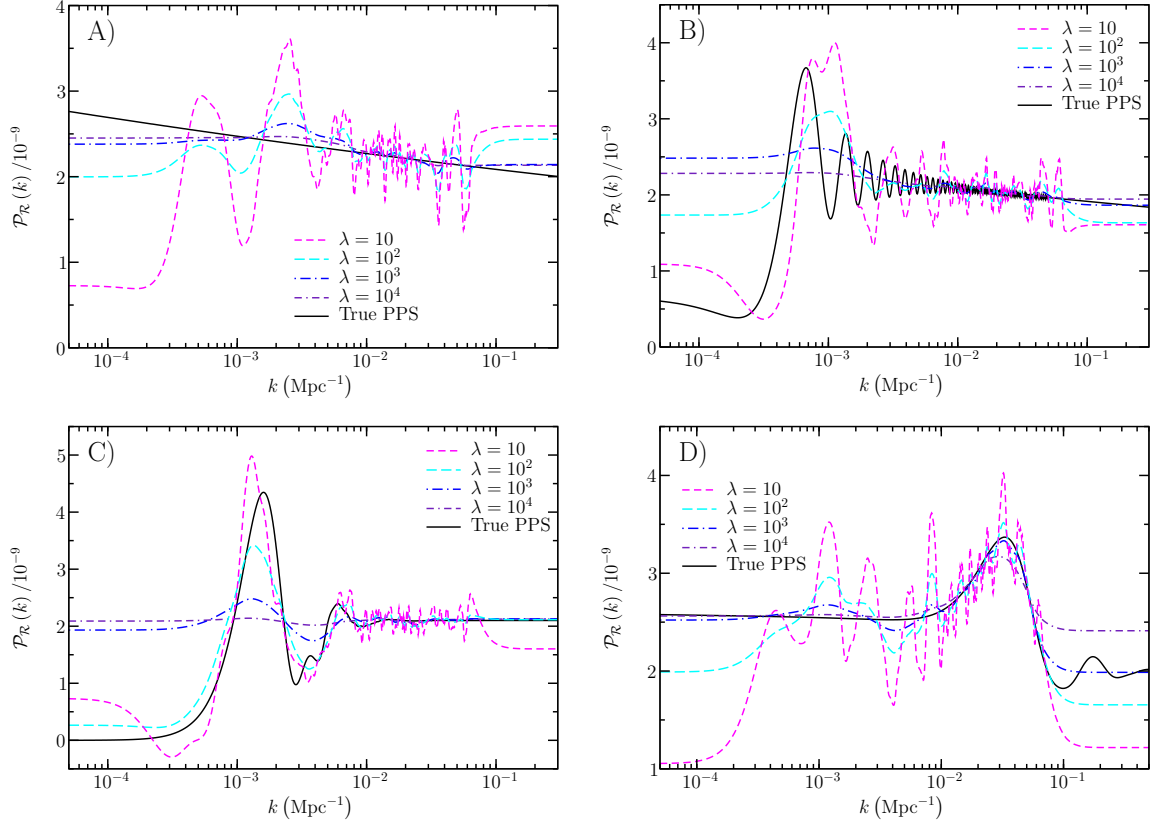


Figure 16. Spectra recovered from noisy mock WMAP-5 TT data with regularisation parameter $\lambda = 10, 10^2, 10^3$ and 10^4 . Results are shown for spectra A to D (see Sec. B.3).

C.3 The bias and variance

In order to understand the bias on large scales, the Taylor expansion eq.(A.2) is useful since Monte Carlo simulations do not provide immediate insights into the different sources of bias. Setting $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_t$ we define zeroth-, first- and second-order approximations to $\hat{\mathbf{p}}$,

$$\hat{\mathbf{p}}_i^{(z)} \equiv \mathcal{T}_i(\mathbf{p}_1, \boldsymbol{\theta}_t, \boldsymbol{\theta}_t, \mathbf{0}), \quad (\text{C.1})$$

$$\hat{\mathbf{p}}_i^{(f)} \equiv \hat{\mathbf{p}}_i^{(z)} + \delta \hat{\mathbf{p}}_i^{(f)}, \quad (\text{C.2})$$

$$\hat{\mathbf{p}}_i^{(s)} \equiv \hat{\mathbf{p}}_i^{(z)} + \delta \hat{\mathbf{p}}_i^{(f)} + \delta \hat{\mathbf{p}}_i^{(s)}, \quad (\text{C.3})$$

where

$$\delta \hat{\mathbf{p}}_i^{(f)} \equiv \sum_j R_{ij} \Delta \mathbf{p}_j + \sum_{\mathbb{Z}, a} M_{ia}^{(\mathbb{Z})} n_a^{(\mathbb{Z})}, \quad (\text{C.4})$$

$$\delta \hat{\mathbf{p}}_i^{(s)} \equiv \frac{1}{2} \sum_{j,k} Y_{ijk} \Delta \mathbf{p}_j \Delta \mathbf{p}_k + \sum_{\mathbb{Z}, j, a} Z_{ija}^{(\mathbb{Z})} \Delta \mathbf{p}_j n_a^{(\mathbb{Z})} + \frac{1}{2} \sum_{\mathbb{Z}, \mathbb{Z}', a, b} X_{iab}^{(\mathbb{Z}\mathbb{Z}')} n_a^{(\mathbb{Z})} n_b^{(\mathbb{Z}')}. \quad (\text{C.5})$$

In the left panel of Fig. 19 the approximations are compared to the full reconstruction for a simulated WMAP temperature and polarisation data set, with \mathbf{p}_1 equal to a H-Z spectrum of amplitude 2.41×10^{-9} . It can be seen that $\mathbf{p}^{(z)}$ exhibits significant departures from \mathbf{p}_1 ,

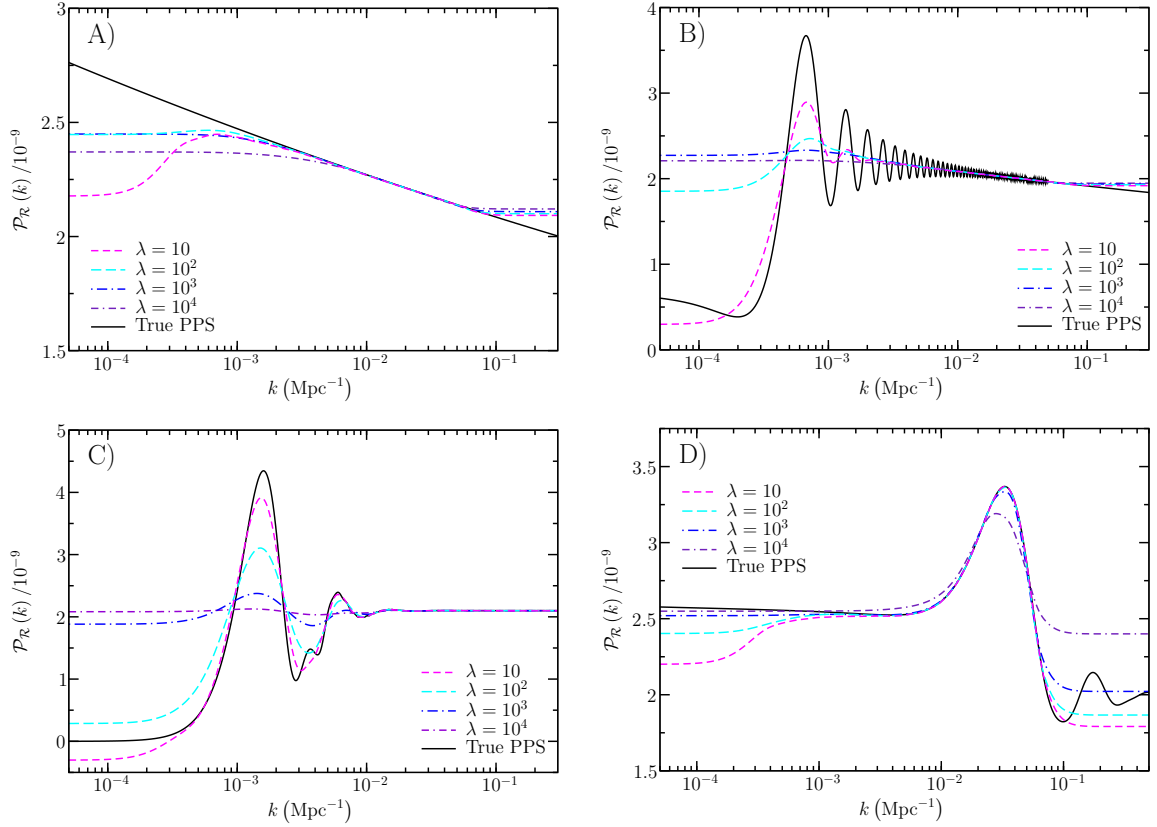


Figure 17. Mean reconstructed spectra from 10^5 mock WMAP-5 TT data realisations with regularisation parameter $\lambda = 10, 10^2, 10^3$ and 10^4 . Results are shown for spectra A to D (see Sec. B.3).

for the following reason. The likelihood function for Gaussian distributed data \mathbf{d}^G with a covariance matrix \mathbf{N}^G dependent on the PPS is

$$L_G = (\mathbf{s}^G - \mathbf{d}^G)^T \mathbf{N}^G (\mathbf{s}^G - \mathbf{d}^G) + \ln \det N^G, \quad (\text{C.6})$$

where $\mathbf{s}^G \equiv \mathbf{W}^G \mathbf{p}$. It has the derivative:

$$\frac{\partial L_G}{\partial p_\alpha} = 2 \sum_{a,a'} W_{a\alpha}^G (N^G)^{-1}_{aa'} (s_{a'}^G - d_{a'}^G) + \sum_{a,a'} [(s_a^G - d_a^G) (s_{a'}^G - d_{a'}^G) - N_{aa'}^G] \frac{\partial (N^G)^{-1}_{aa'}}{\partial p_\alpha}. \quad (\text{C.7})$$

The WMAP TE data for $\ell \geq 24$ has this type of likelihood function. The approximation $\mathbf{p}^{(z)}$ differs from \mathbf{p}_1 because the derivative does not vanish when $\mathbf{p} = \mathbf{p}_t$ for noise free data (recall that $\hat{\mathbf{p}}$ is defined by $\partial Q(\mathbf{p}, \mathbf{d}, \hat{\boldsymbol{\theta}}, \lambda) / \partial p_\alpha \big|_{\mathbf{p}=\hat{\mathbf{p}}} = 0$).

Due to the non-Gaussianity of the WMAP likelihood function at low multipoles, $\hat{\mathbf{p}}^{(s)}$ is closer to $\hat{\mathbf{p}}$ than $\hat{\mathbf{p}}^{(f)}$ for $k \lesssim 5 \times 10^{-3} \text{ Mpc}^{-1}$. The means of the first- and second-order approximations are

$$\langle \hat{p}_i^{(f)} \rangle = \hat{p}_i^{(z)} + \sum_j R_{ij} \Delta p_j, \quad (\text{C.8})$$

$$\langle \hat{p}_i^{(s)} \rangle \equiv \langle \hat{p}_i^{(f)} \rangle + \frac{1}{2} \sum_{j,k} Y_{ijk} \Delta p_j \Delta p_k + \frac{1}{2} \sum_{\mathbb{Z}, a, b} X_{iab}^{(\mathbb{Z}\mathbb{Z})} N_{ab}^{(\mathbb{Z})}. \quad (\text{C.9})$$

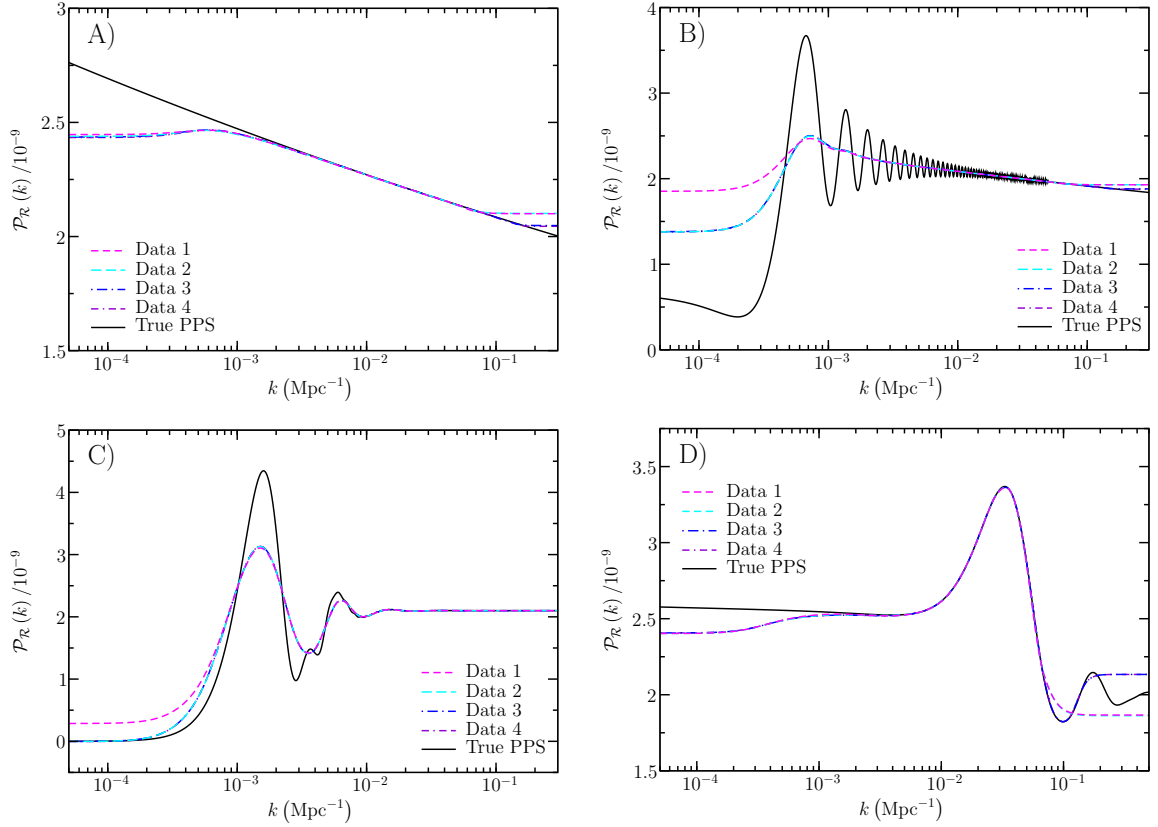


Figure 18. Mean reconstructed test spectra A–D (see Sec. B.3) from 10^5 mock realisations of the four different data set combinations (see Sec. 3.1), with $\lambda = 100$.

These are shown in the right panel of Fig. 19 together with the mean full reconstruction from the simulated data realisations. Since $\langle \partial L_G / \partial p_\alpha \rangle = 0$ for $\mathbf{p} = \mathbf{p}_t$ the approximation $\hat{\mathbf{p}}^{(s)}$ (and $\hat{\mathbf{p}}$) is unbiased for $5 \times 10^{-3} \lesssim k \lesssim 0.1 \text{ Mpc}^{-1}$, unlike $\hat{\mathbf{p}}^{(f)}$. As $\langle \hat{\mathbf{p}}_i^{(s)} \rangle$ is a good approximation to the mean reconstruction, we conclude that $\hat{\mathbf{p}}$ is biased low on large scales by the non-Gaussian likelihood function.

To further study the bias we introduce the four vectors

$$b_{1i} \equiv \mathcal{T}_i(\mathbf{p}_1, \boldsymbol{\theta}_t, \boldsymbol{\theta}_t, \mathbf{0}) - p_{1i}, \quad (\text{C.10})$$

$$b_{2i} \equiv \sum_j (R_{ij} - I_{ij}) \Delta p_j, \quad (\text{C.11})$$

$$b_{3i} \equiv \frac{1}{2} \sum_{j,k} Y_{ijk} \Delta p_j \Delta p_k, \quad (\text{C.12})$$

$$b_{4i} \equiv \frac{1}{2} \sum_{\mathbb{Z}, a, b} X_{iab}^{(\mathbb{Z}\mathbb{Z})} N_{ab}^{(\mathbb{Z})}, \quad (\text{C.13})$$

and their sum \mathbf{b} , the second-order approximation to the bias. In Fig. 20 the norm of these vectors is plotted as a function of λ for the various test spectra and combinations of data sets, again with \mathbf{p}_1 set to a H-Z spectrum of amplitude 2.41×10^{-9} . For the WMAP TT data alone $\mathbf{b}_1 = \mathbf{0}$ because the H-Z spectrum \mathbf{p}_1 is successfully recovered by the inversion.

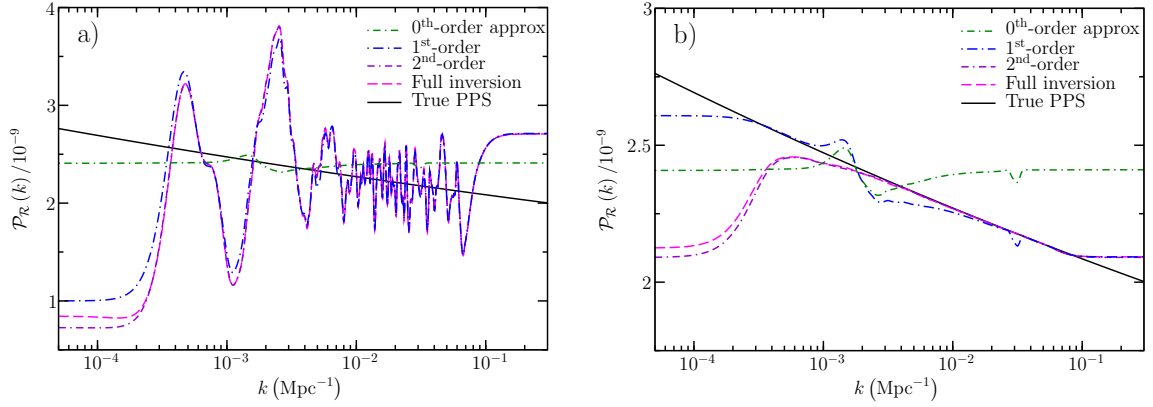


Figure 19. Panel (a) shows test spectrum A, the full reconstruction $\hat{\mathbf{p}}(\mathbf{d})$ and the zeroth-, first- and second-order approximations $\hat{\mathbf{p}}^{(z)}$ (eq. C.1), $\hat{\mathbf{p}}^{(f)}$ (eq. C.2) and $\hat{\mathbf{p}}^{(s)}$ (eq. C.3) using WMAP-5 temperature and polarisation data and $\lambda = 10$. Panel (b) again shows test spectrum A, and the *mean* full reconstruction $\langle \hat{\mathbf{p}}(\mathbf{d}) \rangle$ as well as the mean zeroth-, first- and second-order approximations $\langle \hat{\mathbf{p}}^{(z)} \rangle$, $\langle \hat{\mathbf{p}}^{(f)} \rangle$ (eq. C.8) and $\langle \hat{\mathbf{p}}^{(s)} \rangle$ (eq. C.9).

However, once the WMAP TE data is included \mathbf{b}_1 no longer vanishes as \mathbf{p}_1 is imperfectly recovered due to the determinant term in the WMAP TE likelihood function, as discussed above. The vector \mathbf{b}_1 is the least dependent on λ of the four we study.

The norm of \mathbf{b}_2 generally increases with λ as the resolution of the reconstruction is reduced. The exception occurs for spectrum D, where for $10 \lesssim \lambda \lesssim 1000$ the norm falls as λ increases. This is because, as can be seen in Fig. 17, increasing the regularisation parameter from $\lambda = 10$ to $\lambda = 1000$ brings the mean reconstruction closer to true spectrum on the smallest scales, decreasing the bias. Adding WMAP-5 polarisation data to the temperature data reduces the norm of \mathbf{b}_2 if there are features in the PPS on large scales, such as an infrared cutoff. Hence the reduction is largest for spectra B and C, less so for spectrum A, and smallest for spectrum D. Similarly, adding the small-scale CMB data reduces the extrapolation required in the estimated PPS at high wavenumbers, improving the mean reconstruction if the PPS departs from scale invariance there. Thus $\mathbf{b}_2^T \mathbf{b}_2$ decreases the most for spectrum D when the small-scale CMB data is used. The SDSS-4 LRG data does not cover higher wavenumbers than the small-scale CMB data. The latter constrains the PPS sufficiently well that adding the SDSS-4 LRG data does not further reduce the bias for small values of λ , as can be seen for spectrum D in Fig. 18. For higher λ the inversions are less sensitive to the data and the mean reconstruction using the small scale CMB data is smoothed away from the true PPS. Adding the SDSS-4 LRG data then reduces the smoothing and the bias, provided the true PPS is not scale invariant at high wavenumbers.

The vector \mathbf{b}_3 represents the bias component originating in the quadratic mapping from the true PPS to the reconstruction. It is a subdominant contribution and depends strongly on the true PPS due to the complicated structure of the second-order resolution kernel. The vector \mathbf{b}_4 arises from the non-Gaussianity of the WMAP likelihood function at low multipoles and corresponds to the resulting suppression of the recovered PPS on large scales. For our choice of roughness function $Q(\mathbf{p}, \mathbf{d})$ becomes more quadratic in \mathbf{p} as the regularisation parameter increases. Thus $\mathbf{b}_4^T \mathbf{b}_4$ is less for larger values of λ . Adding the WMAP TE data makes $Q(\mathbf{p}, \mathbf{d})$ more non-Gaussian and leads to a greater suppression on large scales. When

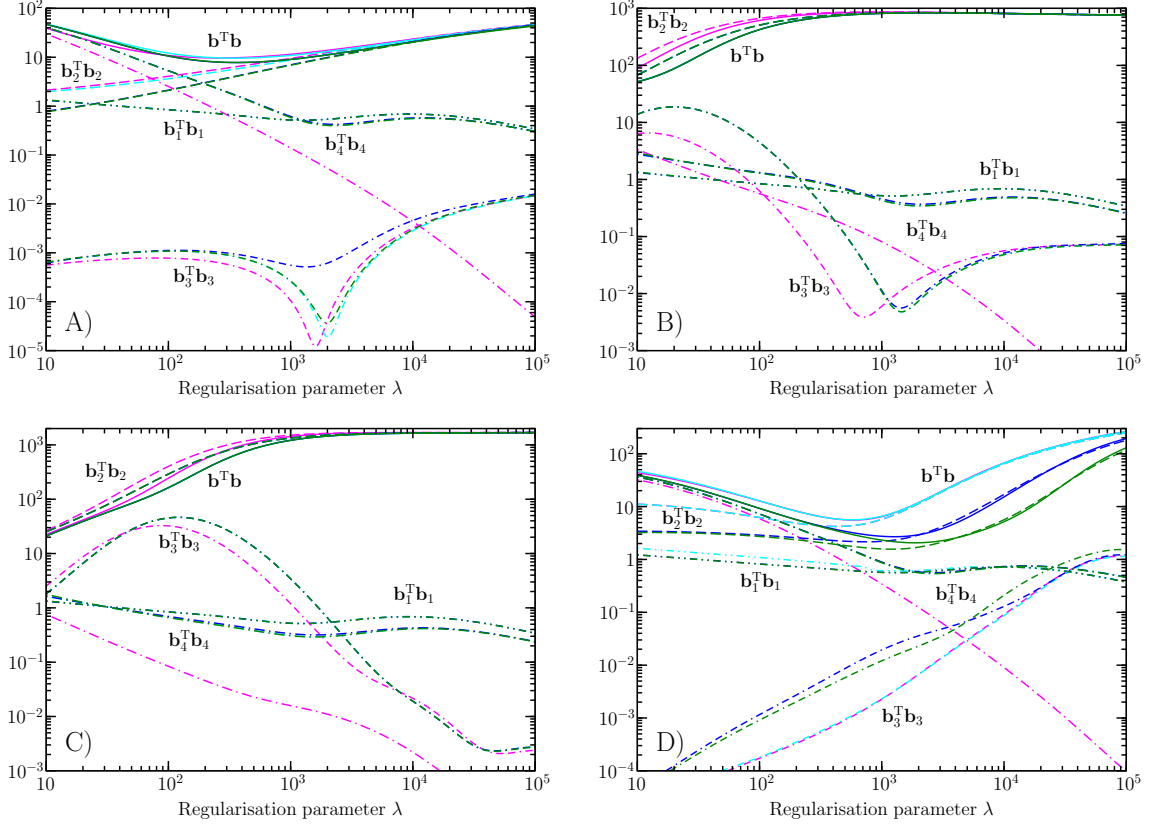


Figure 20. The norm of the bias vectors as labelled versus the regularisation parameter λ , for spectra A to D (see Sec. B.3), and various data combinations as defined in Sec. 3.1 (Combination 1: magenta lines, Combination 2: cyan lines, Combination 3: blue lines, Combination 4: green lines). The dot-dot-dashed lines show $\mathbf{b}_1^T \mathbf{b}_1$ (eq. C.10), the dashed lines $\mathbf{b}_2^T \mathbf{b}_2$ (eq. C.11), the dot-dash-dashed lines show $\mathbf{b}_3^T \mathbf{b}_3$ (eq. C.12), the dot-dashed lines $\mathbf{b}_4^T \mathbf{b}_4$ (eq. C.13), and the solid lines show $\mathbf{b}^T \mathbf{b}$.

the polarisation data is included \mathbf{b}_4 contains a component that cancels with \mathbf{b}_1 to leave the reconstruction unbiased on intermediate scales. Hence $\mathbf{b}_1^T \mathbf{b}_1$ and $\mathbf{b}_4^T \mathbf{b}_4$ are similar for high λ . For low λ values $\mathbf{b}_4^T \mathbf{b}_4$ is greater for spectra A and D than for B and C. This is because the low multipole diagonal elements of the covariance matrix \mathbf{N}^{TT} are larger for spectra without an infrared cutoff.

The frequentist covariance matrix of $\hat{\mathbf{p}}^{(f)}$ is given by Σ_F (eq. A.10), the elements of which are given by:

$$\Sigma_{F|ij} = \Sigma_{ij}^{(f)} \equiv \langle (\delta \hat{\mathbf{p}}_i^{(f)} - \langle \delta \hat{\mathbf{p}}_i^{(f)} \rangle) (\delta \hat{\mathbf{p}}_j^{(f)} - \langle \delta \hat{\mathbf{p}}_j^{(f)} \rangle) \rangle, \quad (\text{C.14})$$

$$= \sum_{\mathbb{Z}, a, b} M_{ia}^{(\mathbb{Z})} N_{ab}^{(\mathbb{Z})} M_{jb}^{(\mathbb{Z})}, \quad (\text{C.15})$$

in terms of the first-order differences $\delta \hat{\mathbf{p}}_i^{(f)}$ (eq. C.4). The frequentist covariance matrix of $\hat{\mathbf{p}}^{(s)}$ is similarly:

$$\Sigma^{(s)} \equiv \langle (\hat{\mathbf{p}}^{(s)} - \langle \hat{\mathbf{p}}^{(s)} \rangle) (\hat{\mathbf{p}}^{(s)} - \langle \hat{\mathbf{p}}^{(s)} \rangle)^T \rangle, \quad (\text{C.16})$$

$$= \Sigma^{(f)} + \Sigma^{(ss)} + 2\Sigma^{(fs)}, \quad (\text{C.17})$$

where

$$\Sigma_{ij}^{(\text{ss})} \equiv \langle (\delta \hat{\mathbf{p}}_i^{(\text{s})} - \langle \delta \hat{\mathbf{p}}_i^{(\text{s})} \rangle) (\delta \hat{\mathbf{p}}_j^{(\text{s})} - \langle \delta \hat{\mathbf{p}}_j^{(\text{s})} \rangle) \rangle, \quad (\text{C.18})$$

$$\begin{aligned} &= \frac{1}{4} \sum_{\mathbb{Z}, a, b, c, d} X_{iab}^{(\mathbb{ZZ})} \left(N_{abcd}^{(\mathbb{Z})4} - N_{ab}^{(\mathbb{Z})} N_{cd}^{(\mathbb{Z})} \right) X_{jcd}^{(\mathbb{ZZ})} \\ &\quad + \sum_{\mathbb{Z}, k, l, a, b} Z_{ika}^{(\mathbb{Z})} \Delta_{\text{Pk}} N_{ab}^{(\mathbb{Z})} \Delta_{\text{Pl}} Z_{jlb}^{(\mathbb{Z})} + \sum_{\mathbb{Z}, k, a, b, c} X_{iab}^{(\mathbb{Z})} N_{abc}^{(\mathbb{Z})3} \Delta_{\text{Pk}} Z_{jkc}^{(\mathbb{Z})}, \end{aligned} \quad (\text{C.19})$$

$$\Sigma_{ij}^{(\text{fs})} \equiv \langle (\delta \hat{\mathbf{p}}_i^{(\text{f})} - \langle \delta \hat{\mathbf{p}}_i^{(\text{f})} \rangle) (\delta \hat{\mathbf{p}}_j^{(\text{s})} - \langle \delta \hat{\mathbf{p}}_j^{(\text{s})} \rangle) \rangle, \quad (\text{C.20})$$

$$= \frac{1}{2} \sum_{\mathbb{Z}, a, b, c} M_{ia}^{(\mathbb{Z})} N_{abc}^{(\mathbb{Z})3} X_{jbc}^{(\mathbb{ZZ})} + \sum_{\mathbb{Z}, k, a, b} M_{ia}^{(\mathbb{Z})} N_{ab}^{(\mathbb{Z})} \Delta_{\text{Pk}} Z_{jkb}^{(\mathbb{Z})}. \quad (\text{C.21})$$

Here $N_{abc}^{(\mathbb{Z})3} \equiv \langle n_a^{(\mathbb{Z})} n_b^{(\mathbb{Z})} n_c^{(\mathbb{Z})} \rangle$ and $N_{abc}^{(\mathbb{Z})4} \equiv \langle n_a^{(\mathbb{Z})} n_b^{(\mathbb{Z})} n_c^{(\mathbb{Z})} n_d^{(\mathbb{Z})} \rangle$.

We repeated the following procedure for many values of λ : $\delta \hat{\mathbf{p}}^{(\text{f})}$ and $\delta \hat{\mathbf{p}}^{(\text{s})}$ were calculated for $\mathfrak{R} = 10^5$ data realisations and the quantities $\text{Tr } \Sigma^{(\text{f})}$, $\text{Tr } \Sigma^{(\text{ss})}$ and $\text{Tr } \Sigma^{(\text{fs})}$ were computed using the estimators

$$\begin{aligned} &\sum_{j=1}^{\mathfrak{R}} \left(\delta \hat{\mathbf{p}}_j^{(\text{f})} - \delta \bar{\mathbf{p}}^{(\text{f})} \right)^{\text{T}} \left(\delta \hat{\mathbf{p}}_j^{(\text{f})} - \delta \bar{\mathbf{p}}^{(\text{f})} \right) / \mathfrak{R}, \\ &\sum_{j=1}^{\mathfrak{R}} \left(\delta \hat{\mathbf{p}}_j^{(\text{s})} - \delta \bar{\mathbf{p}}^{(\text{s})} \right)^{\text{T}} \left(\delta \hat{\mathbf{p}}_j^{(\text{s})} - \delta \bar{\mathbf{p}}^{(\text{s})} \right) / \mathfrak{R}, \\ &\sum_{j=1}^{\mathfrak{R}} \left(\delta \hat{\mathbf{p}}_j^{(\text{f})} - \delta \bar{\mathbf{p}}^{(\text{f})} \right)^{\text{T}} \left(\delta \hat{\mathbf{p}}_j^{(\text{s})} - \delta \bar{\mathbf{p}}^{(\text{s})} \right) / \mathfrak{R}, \end{aligned} \quad (\text{C.22})$$

respectively. Here $\delta \hat{\mathbf{p}}_j^{(\text{f})}$ and $\delta \hat{\mathbf{p}}_j^{(\text{s})}$ are the vectors $\delta \hat{\mathbf{p}}^{(\text{f})}$ and $\delta \hat{\mathbf{p}}^{(\text{s})}$ for the j th data realisation and $\delta \bar{\mathbf{p}}^{(\text{f})} \equiv \sum_{j=1}^{\mathfrak{R}} \delta \hat{\mathbf{p}}_j^{(\text{f})} / \mathfrak{R}$ and $\delta \bar{\mathbf{p}}^{(\text{s})} \equiv \sum_{j=1}^{\mathfrak{R}} \delta \hat{\mathbf{p}}_j^{(\text{s})} / \mathfrak{R}$ are the means. This allowed us to plot $\text{Tr } \Sigma^{(\text{f})}$, $\text{Tr } \Sigma^{(\text{ss})}$, $2\text{Tr } \Sigma^{(\text{fs})}$ and their sum, $\text{Tr } \Sigma^{(\text{s})}$, as a function of the regularisation parameter for the various test spectra and combinations of data sets. The results are shown in Fig. 21, again for $\mathbf{p}_1 = 2.41 \times 10^{-9}$.

Noise-induced artifacts in a reconstruction are more suppressed for high values of λ . Since the artifacts vary with different data realisations, this means that $\text{Tr } \Sigma^{(\text{f})}$ and $\text{Tr } \Sigma^{(\text{s})}$ decrease as λ increases. At higher λ values second-order effects are smaller and $\text{Tr } \Sigma^{(\text{s})}$ is closer to $\text{Tr } \Sigma^{(\text{f})}$ because $Q(\mathbf{p}, \mathbf{d})$ becomes more quadratic for larger λ . Generally $2\text{Tr } \Sigma^{(\text{fs})}$ dominates $\text{Tr } \Sigma^{(\text{ss})}$ and determines how $\text{Tr } \Sigma^{(\text{s})}$ differs from $\text{Tr } \Sigma^{(\text{f})}$. For spectra A and D, $\text{Tr } \Sigma^{(\text{fs})}$ is predominantly negative so $\text{Tr } \Sigma^{(\text{s})}$ is mostly smaller than $\text{Tr } \Sigma^{(\text{f})}$. A slight bump in $\text{Tr } \Sigma^{(\text{s})}$ for spectra B and C corresponds to the maximum of $\text{Tr } \Sigma^{(\text{fs})}$. For small λ , adding the WMAP polarisation and the small-scale CMB data both reduce $\text{Tr } \Sigma^{(\text{f})}$ and $\text{Tr } \Sigma^{(\text{s})}$, but including the SDSS-4 LRG data increases them. For large λ , $\text{Tr } \Sigma^{(\text{f})}$ and $\text{Tr } \Sigma^{(\text{s})}$ increase with each additional data set.

C.4 The MSE and MPE

Next, inversions were performed on 10^5 data realisations and the squared bias $\langle \hat{\mathbf{p}} - \mathbf{p}_t \rangle^{\text{T}} \langle \hat{\mathbf{p}} - \mathbf{p}_t \rangle$, variance $\langle (\hat{\mathbf{p}} - \langle \hat{\mathbf{p}} \rangle)^{\text{T}} (\hat{\mathbf{p}} - \langle \hat{\mathbf{p}} \rangle) \rangle$ and their sum, the MSE were computed using the estimators $(\bar{\mathbf{p}} - \mathbf{p}_t)^{\text{T}} (\bar{\mathbf{p}} - \mathbf{p}_t)$, $\sum_{j=1}^{\mathfrak{R}} (\hat{\mathbf{p}}_j - \bar{\mathbf{p}})^{\text{T}} (\hat{\mathbf{p}}_j - \bar{\mathbf{p}}) / \mathfrak{R}$, and $\sum_{j=1}^R (\hat{\mathbf{p}}_j - \mathbf{p}_t)^{\text{T}} (\hat{\mathbf{p}}_j - \mathbf{p}_t) / \mathfrak{R}$,

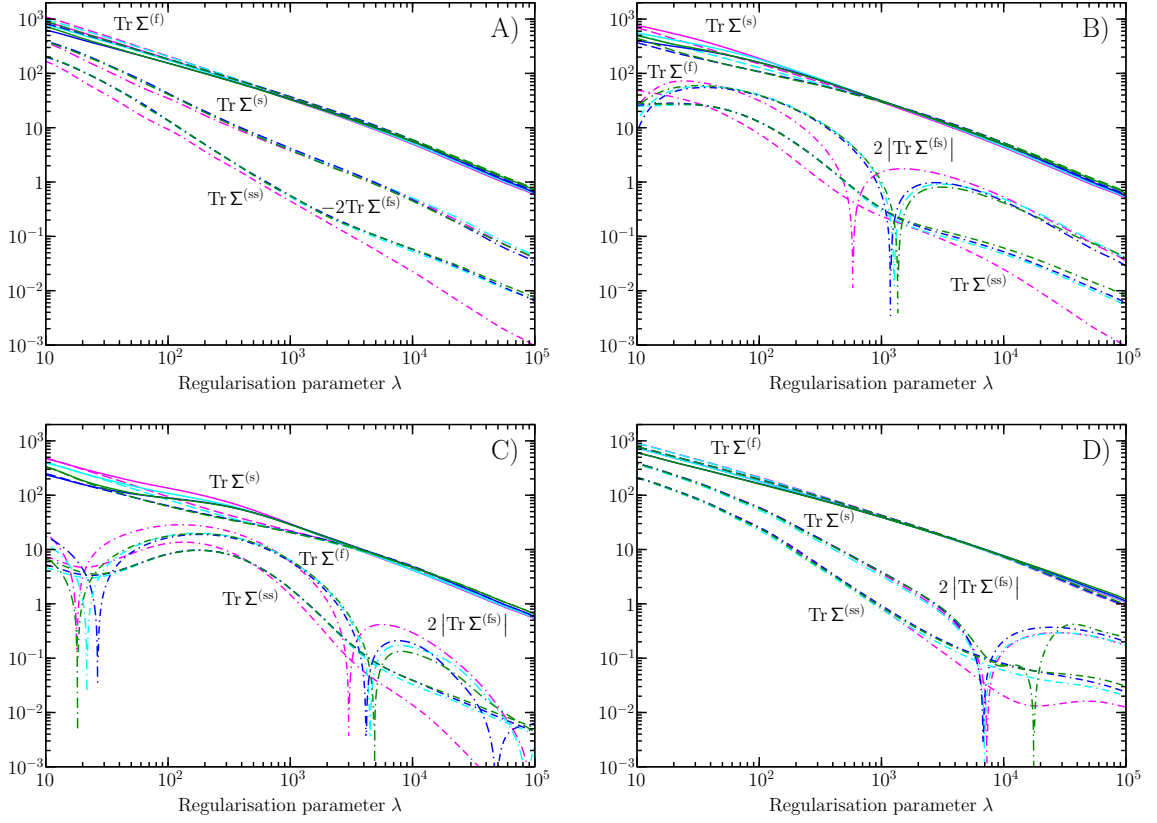


Figure 21. The trace of the frequentist covariance matrices versus the regularisation parameter λ for test spectra A to D (see Sec. B.3) and various data combinations as defined in Sec. 3.1 (Combination 1: magenta lines, Combination 2: cyan lines, Combination 3: blue lines, Combination 4: green lines). The dashed lines show $\text{Tr } \Sigma^{(f)}$ (eq. C.14), the dot-dash-dashed lines show $\text{Tr } \Sigma^{(ss)}$ (eq. C.18), the dot-dashed lines show $2|\text{Tr } \Sigma^{(fs)}|$ (eq. C.20), and the solid lines show $\text{Tr } \Sigma^{(s)}$ (eq. C.16).

respectively. Here $\hat{\mathbf{p}}_j$ is the j th of the $\mathfrak{R} = 10^5$ reconstructions and $\bar{\mathbf{p}} \equiv \sum_{j=1}^{\mathfrak{R}} \hat{\mathbf{p}}_j / \mathfrak{R}$ is the mean reconstruction. The MPE of the reconstructions was also calculated. This was repeated for many values of the regularisation parameter.

The squared bias, variance and MSE are shown in Fig. 22 as a function of λ for the various test spectra and combinations of data sets. The squared bias and variance found from the 10^5 data realisations exhibit similar behaviour to their second-order approximations $\mathbf{b}^T \mathbf{b}$ and $\text{Tr } \Sigma^{(s)}$. The contributions of the squared bias and the variance to the MSE vary with λ : for low values of λ the variance dominates, whereas for high values the bias dominates. Clearly there is a trade-off between bias and variance.¹⁴ Regularisation solves the inversion problem because for an appropriate value of λ a large reduction in the variance can be achieved by accepting a small bias, with the result that the MSE is greatly improved.

The value of λ which minimises the MSE depends on the test spectrum. It is smaller for spectra B and C since the bias increases more rapidly with λ for these spectra due to their large deviations from scale invariance. The minimum MSE is greater for spectra B and C due to the large bias and larger variance at low λ .

¹⁴The bias-variance trade-off is a general feature of model selection and is well known in estimation theory.

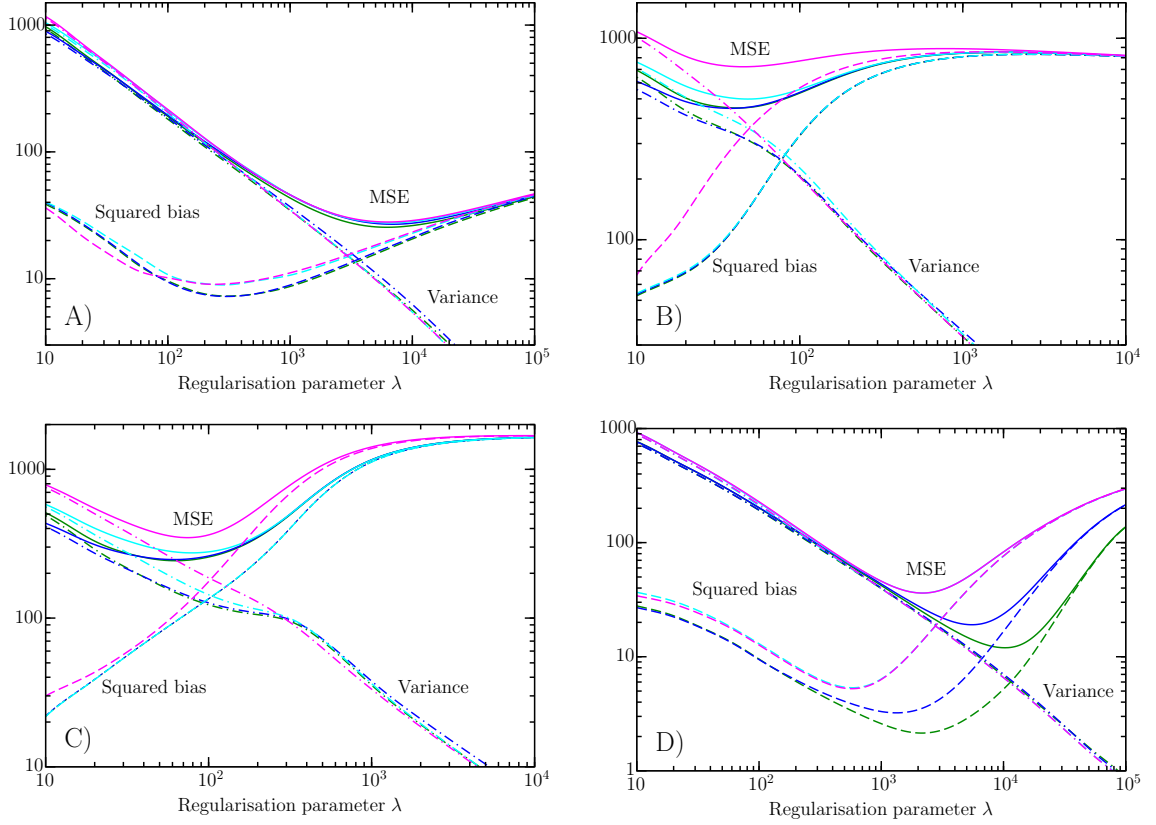


Figure 22. The squared bias, variance and (their sum) the mean squared error (eq. A.31) versus the regularisation parameter λ for spectra A–D (see Sec. B.3) and various data combinations as defined in Sec. 3.1 (Combination 1: magenta lines, Combination 2: cyan lines, Combination 3: blue lines, Combination 4: green lines). In each case the solid lines show the MSE, the dashed lines show the squared bias and the dot-dashed lines show the variance.

Fig. 23 shows the estimated MPE as a function of λ . For small λ the MPE is large as the predicted reconstructions are close to the noisy data rather than the noise-free data. For larger λ the predicted data resemble a smoothed version of the noisy data. Since this is closer to the noise-free data the MPE is reduced. The MPE increases again for very large λ because the predicted data are oversmoothed and approach that of the best-fit H-Z spectrum.

The MPE exhibits similar behaviour to the MSE for spectra A, C and D (e.g. the minimum of the functions are at the same λ value to within a factor of 2–3). However the MPE of spectrum B is very different to the corresponding MSE and instead resembles the MSE of spectrum A.

The MPE can be understood as a version of the MSE weighted towards wavenumbers where the PPS is more strongly constrained by the data. The MPE for the WMAP-5 TT data is most sensitive to the reconstructed PPS on intermediate scales, between about $k = 3 \times 10^{-3} \text{ Mpc}^{-1}$ and $k = 0.1 \text{ Mpc}^{-1}$.

Spectrum C has the largest features in this wavenumber range and so the value of λ which minimises the MPE is smallest for this spectrum. The λ value which minimises the MPE is higher for spectrum D since the ‘bump’ in this spectrum is less pronounced. The MPE is insensitive to the large scale cutoff in spectrum B. On intermediate scales where it

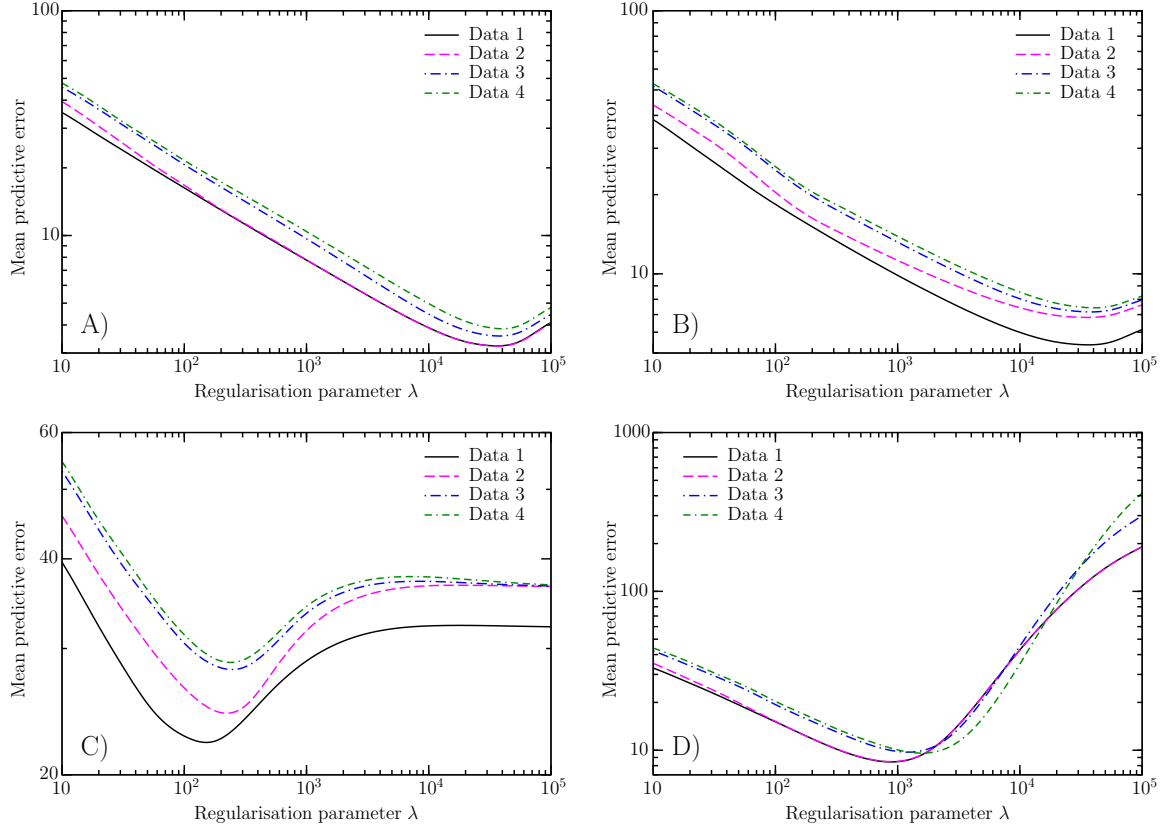


Figure 23. The mean predicted error (eq. A.32) versus the regularisation parameter λ . The plots are for spectrum A to D (see Sec. B.3).

exhibits low amplitude oscillations spectrum B is comparable to spectrum A and has the same tilt. This accounts for the similarity of the MPE of the two spectra.

The estimated PPS between $k = 3 \times 10^{-3} \text{ Mpc}^{-1}$ and $k = 0.04 \text{ Mpc}^{-1}$ largely determines the contribution to the MPE of the WMAP-5 polarisation data. Since spectra B and C possess features at the lower end of this wavenumber interval which are imperfectly recovered, the polarisation contribution is larger for these spectra than the two others.

The small-scale CMB and SDSS-4 LRG components of the MPE depend on the reconstructed PPS over the ranges $9 \times 10^{-3} \lesssim k \lesssim 0.2 \text{ Mpc}^{-1}$ and $0.02 \lesssim k \lesssim 0.2 \text{ Mpc}^{-1}$ respectively. For spectrum D adding the small scale CMB data significantly improves the inversion, due to the bump in the spectrum. This leads to a reduced WMAP-5 TT contribution to the MPE sufficient for intermediate values of λ to offset the additional small-scale CMB component, so that the MPE falls. Similarly including the SDSS-4 LRG data further decreases the WMAP-5 TT and small-scale CMB components for intermediate λ , and the MPE is again reduced.

D Choosing the regularisation parameter

For a successful reconstruction of the PPS the value of the regularisation parameter must be chosen correctly. If λ is too small $\hat{\mathbf{p}}$ will be dominated by artifacts caused by fitting the noise. On the other hand, if λ is too large $\hat{\mathbf{p}}$ will be oversmoothed and could miss features in the

true PPS. Choosing the regularisation parameter is akin to an exercise in model selection. The complexity of $\hat{\mathbf{p}}$ clearly decreases with increasing λ . For $\lambda = 0$ all the elements of $\hat{\mathbf{p}}$ are independent whereas for $\lambda = \infty$ the elements are totally correlated, so that there is effectively only one free parameter, the amplitude of the H-Z spectrum. One measure of the effective number of free parameters is the quantity $\nu_1 \equiv \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}})$ where $\mathbf{S}_{\mathbb{Z}} \equiv \mathbf{W}_{\mathbb{Z}}\mathbf{M}_{\mathbb{Z}}$ [187]. Its value decreases monotonically from $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) = N_j$ for $\lambda = 0$ to $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) = 1$ for $\lambda = \infty$. Other measures are $\nu_2 \equiv \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}^T \mathbf{S}_{\mathbb{Z}})$ and $\nu_3 \equiv \sum_{\mathbb{Z}} \text{Tr}(2\mathbf{S}_{\mathbb{Z}} - \mathbf{S}_{\mathbb{Z}}^T \mathbf{S}_{\mathbb{Z}})$.

D.1 Parameter selection methods

In general, the optimum value of the regularisation parameter depends on the noise level, the regularisation scheme and the object to be recovered. Unfortunately no universal strategy for selecting the regularisation parameter is known which works in all situations. Instead numerous methods have been proposed in the literature, based upon different principles. Here we consider five different methods: the discrepancy principle (DP), equivalent degrees of freedom (EDF), normalised cumulative periodogram (NCP), Mallows' C_p and generalised cross-validation (GCV). The aim of the first three techniques is to ensure that the estimated PPS has a statistically reasonable fit to the data, while the final two seek to minimise the PE. Although most of the methods were originally devised for single data sets with white noise, it is simple to recast them in forms suitable for multiple data sets with correlated noise. For comparison purposes we also include a method in which the SE is minimised, even though it requires knowledge of \mathbf{p}_t .

The discrepancy principle method [188] is also known as the constrained least squares method [189] and chi-squared method [190]. According to the discrepancy principle the estimated PPS should fit the data only to within the noise, which for multiple data sets with Gaussian noise is equivalent to requiring that

$$\chi^2(\hat{\mathbf{p}}) \equiv \sum_{\mathbb{Z}} (\mathbf{W}_{\mathbb{Z}}\hat{\mathbf{p}} - \mathbf{d}_{\mathbb{Z}})^T \mathbf{N}_{\mathbb{Z}}^{-1} (\mathbf{W}_{\mathbb{Z}}\hat{\mathbf{p}} - \mathbf{d}_{\mathbb{Z}}) = N_d. \quad (\text{D.1})$$

The value of λ for which eq.(D.1) holds is denoted by λ_{DP} . This method is popular due to its simplicity and was previously used in [141, 142]. However, it implicitly assumes that $\chi^2(\hat{\mathbf{p}})$, like $\chi^2(\mathbf{p}_t)$, has a χ^2 distribution with N_d degrees of freedom. In this case $\langle \chi^2(\hat{\mathbf{p}}) \rangle$ would equal N_d and $\chi^2(\hat{\mathbf{p}}) = N_d$ would signify an acceptable fit to the data. In fact this is untrue: it can be shown that

$$\langle \chi^2(\hat{\mathbf{p}}) \rangle = \sum_{\mathbb{Z}} \mathbf{p}_t^T \mathbf{W}_{\mathbb{Z}}^T (\mathbf{S}_{\mathbb{Z}}^T - \mathbf{I})^T \mathbf{N}_{\mathbb{Z}}^{-1} (\mathbf{S}_{\mathbb{Z}} - \mathbf{I}) \mathbf{W}_{\mathbb{Z}} \mathbf{p}_t + \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}^T \mathbf{N}_{\mathbb{Z}}^{-1} \mathbf{S}_{\mathbb{Z}} \mathbf{N}_{\mathbb{Z}}) - 2 \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) + N_d. \quad (\text{D.2})$$

We approximate the sum of the first two terms on the right of eq.(D.2) by $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}})$, to which it reduces as λ tends to zero, and require that

$$\chi^2(\hat{\mathbf{p}}) = N_{\text{eff}}, \quad (\text{D.3})$$

where $N_{\text{eff}} \equiv N_d - \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}})$. The value of λ which satisfies eq.(D.3) is denoted by λ_{EDF} . With the interpretation of $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}})$ as the effective number of parameters, N_{eff} represents the number of degrees of freedom of the reconstruction. This is known as the equivalent degrees of freedom method [190] and also the compensated discrepancy principle method [191]. A different Bayesian motivation for it can be found in [192].

The rationale for the NCP method is that the residual $\hat{\mathbf{d}} - \mathbf{d}$ should have the same statistical behaviour as the noise in the data when the PPS is correctly recovered [193, 194]. This suggests that the optimal value of λ is the one for which the statistical properties of the residual are closest to those of the noise. To apply the NCP method we construct random variables $y_a^{(\mathbb{Z})}$ from the predicted data intended to behave like white noise when $\hat{\mathbf{p}} = \mathbf{p}_t$. For the WMAP data these are

$$y_\ell^X = \sum_{\ell'} (J^X)_{\ell\ell'}^{-1} (\hat{s}_\ell^X - d_\ell^X), \quad (\text{D.4})$$

where $\hat{s}_\ell^X \equiv \sum_i W_{\ell i}^X \hat{p}_i$ and J^X is given by the Cholesky decomposition $\mathbf{N}^X = \mathbf{J}^X \mathbf{J}^{X\text{T}}$. The random variables for the small-scale CMB data are

$$y_b^X = \sum_{X', b'} (J^{XX'})_{bb'}^{-1} \hat{z}_b^{X'}. \quad (\text{D.5})$$

Here $\hat{z}_b^X = \hat{s}_b^X - d_b^X$ for a Gaussian bandpower and $\hat{z}_b^X = \ln(\hat{s}_b^X / \mathcal{D}_b^X)$ for a log-Gaussian bandpower, with $\hat{s}_b^X \equiv \sum_i T_{bi}^X \hat{p}_i$ and $\hat{\mathcal{S}}_b^X \equiv \hat{s}_b^X + \mathcal{N}_b^X$. The matrices $J^{XX'}$ are related to the matrices $\mathbf{V}^{XX'}$ by a Cholesky decomposition. For the SDSS-4 LRG data the variables are

$$y_a = \sum_{a'} (J^{\text{LRG}})_{aa'}^{-1} (\hat{s}_{a'} - d_{a'}), \quad (\text{D.6})$$

where $\hat{s}_a = \sum_i T_{ai}^{\text{LRG}} \hat{p}_i$ and J^{LRG} is given by the Cholesky decomposition of $\tilde{\mathbf{N}}^{\text{LRG}}$.

The quantities $y_a^{(\mathbb{Z})}$ of the different data sets are assembled into a single vector \mathbf{y} . If λ is too high or low \mathbf{y} will be dominated by low or high frequency components, respectively. The discrete Fourier transform (DFT) of \mathbf{y} is calculated using the fast Fourier transform algorithm after zero-padding the vector out to N_y , a convenient power of two. The sine and cosine coefficients of the DFT are

$$c_k = \sum_{a=1}^{N_y} y_a \cos \frac{2\pi (k-1)(a-1)}{N_y}, \quad (\text{D.7})$$

$$s_k = - \sum_{a=1}^{N_y} y_a \sin \frac{2\pi (k-1)(a-1)}{N_y}, \quad (\text{D.8})$$

for $k = 1, \dots, N_y/2 + 1$. The NCP of \mathbf{y} is the vector \mathbf{h} defined as

$$h_j = \frac{\sum_{k=1}^j c_k^2 + s_k^2}{\sum_{k=1}^{N_y/2+1} c_k^2 + s_k^2}. \quad (\text{D.9})$$

The NCP is a test for white noise, for which it lies close to the vector $v_j = 2(j-1)/N_y$. In the NCP method the value of the regularisation parameter is taken to be the one which minimises the function $\gamma_{\text{NCP}}(\lambda) \equiv [\mathbf{h}(\lambda) - \mathbf{v}]^T [\mathbf{h}(\lambda) - \mathbf{v}]$. It is labelled λ_{NCP} .

An alternative approach to the selection of the regularisation parameter involves attempting to minimise a loss function. It can be shown that

$$\text{MPE}(\hat{\mathbf{p}}) = \sum_{\mathbb{Z}} \mathbf{p}_t^T \mathbf{W}_{\mathbb{Z}}^T (\mathbf{S}_{\mathbb{Z}}^T - \mathbf{I})^T \mathbf{N}_{\mathbb{Z}}^{-1} (\mathbf{S}_{\mathbb{Z}} - \mathbf{I}) \mathbf{W}_{\mathbb{Z}} \mathbf{p}_t + \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}^T \mathbf{N}_{\mathbb{Z}}^{-1} \mathbf{S}_{\mathbb{Z}} \mathbf{N}_{\mathbb{Z}}), \quad (\text{D.10})$$

$$= \langle \chi^2(\hat{\mathbf{p}}) \rangle + 2 \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) - N_d. \quad (\text{D.11})$$

The second equality follows from eq.(D.2). Thus Mallows's C_p statistic

$$C_p(\lambda) \equiv \chi^2(\hat{\mathbf{p}}) + 2 \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) - N_d \quad (\text{D.12})$$

is an unbiased estimator of the MPE. Mallows's C_p method [195] chooses the value of λ which minimises $C_p(\lambda)$, denoted by λ_{CP} , since on average λ_{CP} minimises the PE. This is also referred to as the unbiased predictive risk estimator method [196] and as Stein's unbiased risk estimator method [197]. Note that the C_p statistic is essentially a special case of the Akaike information criterion used in model selection [198].

The idea behind cross-validation is that the optimum estimate of the PPS is the one best at predicting new or unused data. In leave-one-out cross-validation a statistic is employed which quantifies the accuracy with which $\hat{\mathbf{p}}$ predicts individual data points using all the other data. Let $\hat{\mathbf{p}}_{-a}^{(\mathbb{Z})}$ be the estimate of the PPS obtained when the data point $\mathbf{d}_a^{(\mathbb{Z})}$ is unused. Then $\hat{\mathbf{d}}_{-a}^{(\mathbb{Z})} \equiv \left(\mathbf{W}_{\mathbb{Z}} \hat{\mathbf{p}}_{-a}^{(\mathbb{Z})} \right)_a$ is the estimate of $\mathbf{d}_a^{(\mathbb{Z})}$ found using the other data points. It can be argued that optimum value of λ is the one which minimises the function

$$V_{\text{CV}}(\lambda) \equiv \sum_{\mathbb{Z}, a, b} \left(\hat{\mathbf{d}}_{-a}^{(\mathbb{Z})} - \mathbf{d}_a^{(\mathbb{Z})} \right) \mathbf{N}_{ab}^{(\mathbb{Z})-1} \left(\hat{\mathbf{d}}_{-b}^{(\mathbb{Z})} - \mathbf{d}_b^{(\mathbb{Z})} \right). \quad (\text{D.13})$$

Using a rank one update formula yields

$$\hat{\mathbf{d}}_{-a}^{(\mathbb{Z})} - \mathbf{d}_a^{(\mathbb{Z})} = \frac{(\mathbf{W}_{\mathbb{Z}} \hat{\mathbf{p}})_a - \mathbf{d}_a^{(\mathbb{Z})}}{1 - S_{aa}^{(\mathbb{Z})}}. \quad (\text{D.14})$$

In generalised cross-validation [199] the elements $S_{aa}^{(\mathbb{Z})}$ are replaced by their mean value $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) / N_d$ which leads to the more convenient statistic

$$V_{\text{GCV}}(\lambda) \equiv \frac{\chi^2(\hat{\mathbf{p}})}{\left[1 - N_d^{-1} \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) \right]^2}. \quad (\text{D.15})$$

The value of λ which minimises eq.(D.15) is written λ_{GCV} . The mean value of λ_{GCV} satisfies $d\langle V_{\text{GCV}}(\lambda) \rangle / d\lambda = 0$ which is equivalent to

$$\frac{d}{d\lambda} [\text{MPE}(\hat{\mathbf{p}})] - 2 \left[1 - \frac{\langle \chi^2(\hat{\mathbf{p}}) \rangle}{N_d - \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}})} \right] \frac{d}{d\lambda} \left[\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) \right] = 0. \quad (\text{D.16})$$

Provided the second term on the left above is small, which typically the case in practice [196], it can be seen that λ_{GCV} will on average equal λ_{CP} and minimise the PE.

Finally, λ_{LSE} is defined as the minimiser of the SE (eq. A.29). We refer to this as the least squared error (LSE) method and identify λ_{LSE} as the optimum value of the regularisation parameter. Although calculation of λ_{LSE} requires \mathbf{p}_t to be known, we include it as a benchmark for the other methods.

D.2 Application to test spectra

We investigate the regularisation parameter selection methods by applying them to mock data. Minimisation of $C_p(\lambda)$, $V_{\text{GCV}}(\lambda)$ and the SE is performed using Brent's method, while eqs.(D.1) and (D.3) are solved by Ridders' method [157]. Instead of using $\chi^2(\hat{\mathbf{p}})$ as defined in

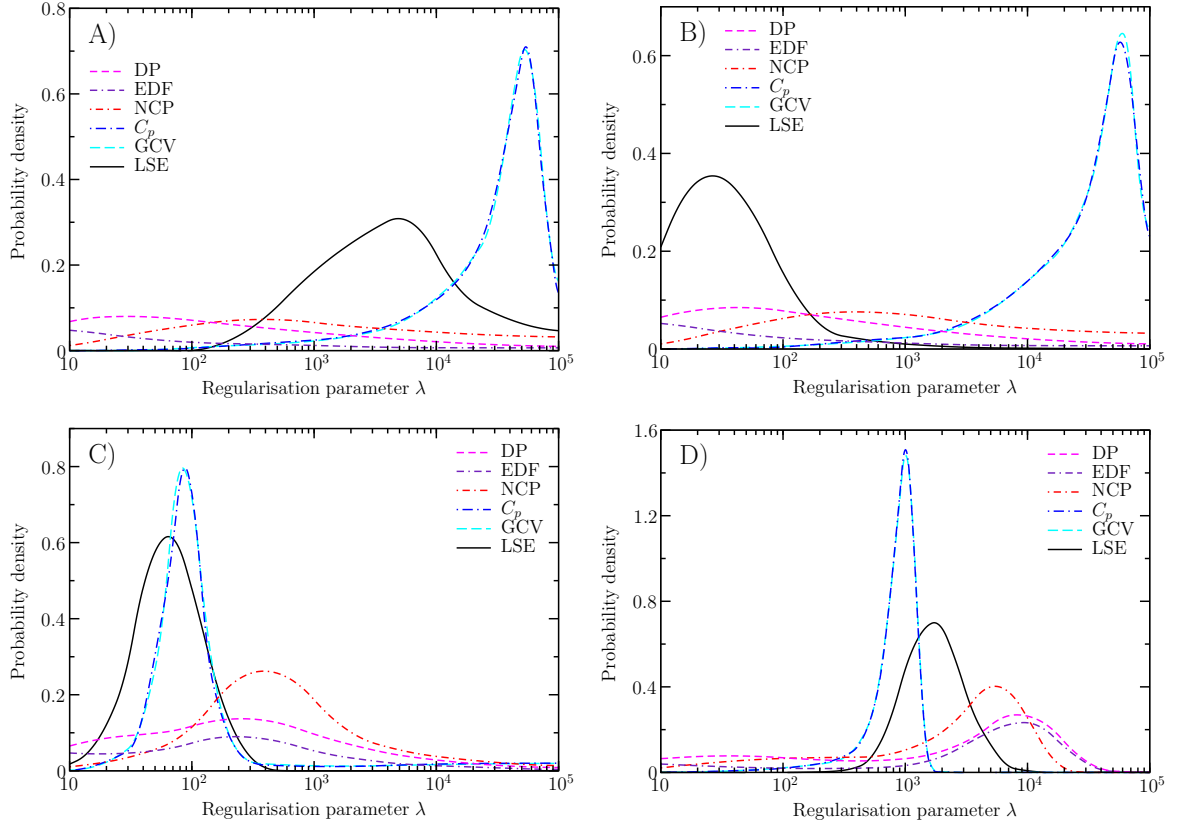


Figure 24. Distributions of the regularisation parameter chosen by the different methods for inversions using the WMAP-5 temperature data (combination 1, Sec. 3.1) and for spectra A–D Sec. B.3).

eq.(D.1), in practice we use the full log-likelihood function $L(\mathbf{p}, \mathbf{d})$ in eqs.(D.1), (D.3), (D.12) and (D.15), excluding the determinant terms. For each of the four test spectra and four data combinations we generate 10^4 mock data realisations, to which the selection methods are applied. This enables us to estimate the probability distributions of the chosen parameter values by constructing histograms of the results. These are shown in Figs. 24–27.

The optimal values of λ according to the LSE scheme are consistently small for spectra B and C due to their infrared cutoffs but are higher for spectra A and D. The peaks of the LSE distributions closely follow the minima of the MSE functions shown in Fig. 22. The variation in the λ_{LSE} values is related to the depth of the MSE minimum, e.g. the LSE distribution is narrower for spectrum D than for spectrum A, since the MSE minimum of spectra D is deeper. However, the ensemble mean of the λ_{LSE} values equals the λ value which minimises the MSE only if the λ versus SE curves of the ensemble are symmetric about their minima *and* translationally symmetric copies of each other. Since these conditions are not met, there is no absolute correspondence between the LSE distributions and the MSE functions.

The distributions for Mallows’ C_p method follow the minima of the MPE functions displayed in Fig. 23. This leads to relatively successful selection of λ for spectra C and D when the WMAP-5 data alone are used. However, with the addition of the small-scale CMB and SDSS-4 LRG the C_p method has a tendency to underestimate the correct value of λ for spectrum D and to overestimate it for spectrum C. The C_p method generally results in

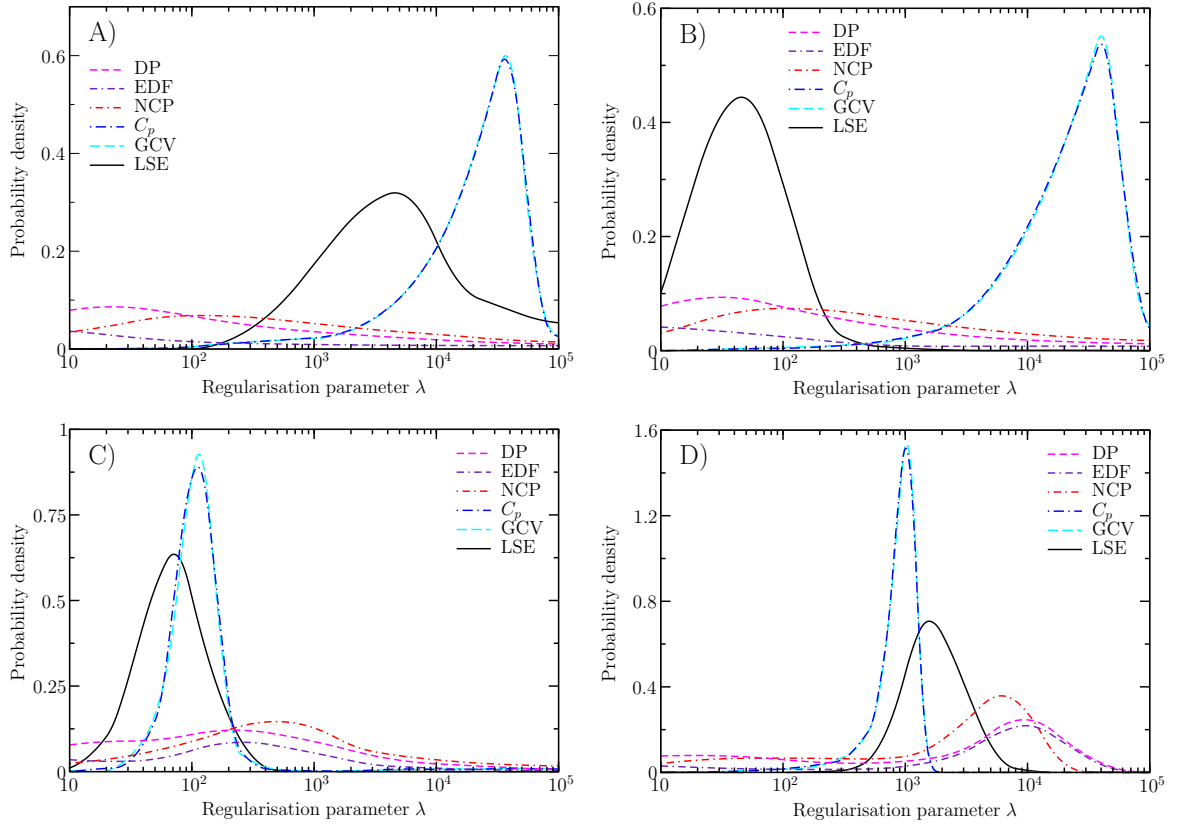


Figure 25. Same as Fig. 24 for inversions using the WMAP-5 temperature and polarisation data (combination 2, Sec. 3.1), and for spectra A–D Sec. B.3).

oversmoothed reconstructions for spectrum A, and performs even more poorly when applied to spectrum B, with extreme over smoothing.

The GCV distributions are almost identical to those of the C_p method. This is due to the similarity of the V_{GCV} and C_p statistics. In practice $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}})/N_d$ is of $\mathcal{O}(0.1)$ for reasonable reconstructions which permits the Taylor expansion

$$V_{\text{GCV}}(\lambda) = \chi^2(\hat{\mathbf{p}}) \left[1 + 2N_d^{-1} \sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) + \dots \right]. \quad (\text{D.17})$$

Since $\chi^2(\hat{\mathbf{p}})/N_d \simeq 1$ (see Fig. 28) we have $V_{\text{GCV}} \simeq C_p$, leading to the closeness of the results.

The peaks of the DP distributions are in approximately the correct location for spectra B and D when the WMAP-5, small-scale CMB and SDSS-4 LRG data are used together, but not for the other spectra. Moreover, the λ values chosen by the DP method are highly scattered. The DP results can be understood with reference to Fig. 28. This shows the λ dependence of the ensemble means $\langle \chi^2/N_d \rangle$ and $\langle \chi^2/N_{\text{eff}} \rangle$, estimated from 10^5 realisations of the WMAP-5, small-scale CMB and SDSS-4 LRG data sets. The ensemble mean of the λ_{DP} values is *not* equal to the value of λ for which $\langle \chi^2/N_d \rangle = 1$ because the $\chi^2/N_d - \lambda$ curves of the ensemble are not straight, parallel lines. The probability density is greatest where the $\chi^2/N_d - \lambda$ curves intersect the $\chi^2/N_d = 1$ line at the least acute angle. Thus the peak of the distribution is located where the $\chi^2/N_d - \lambda$ curves are steepest in the region where

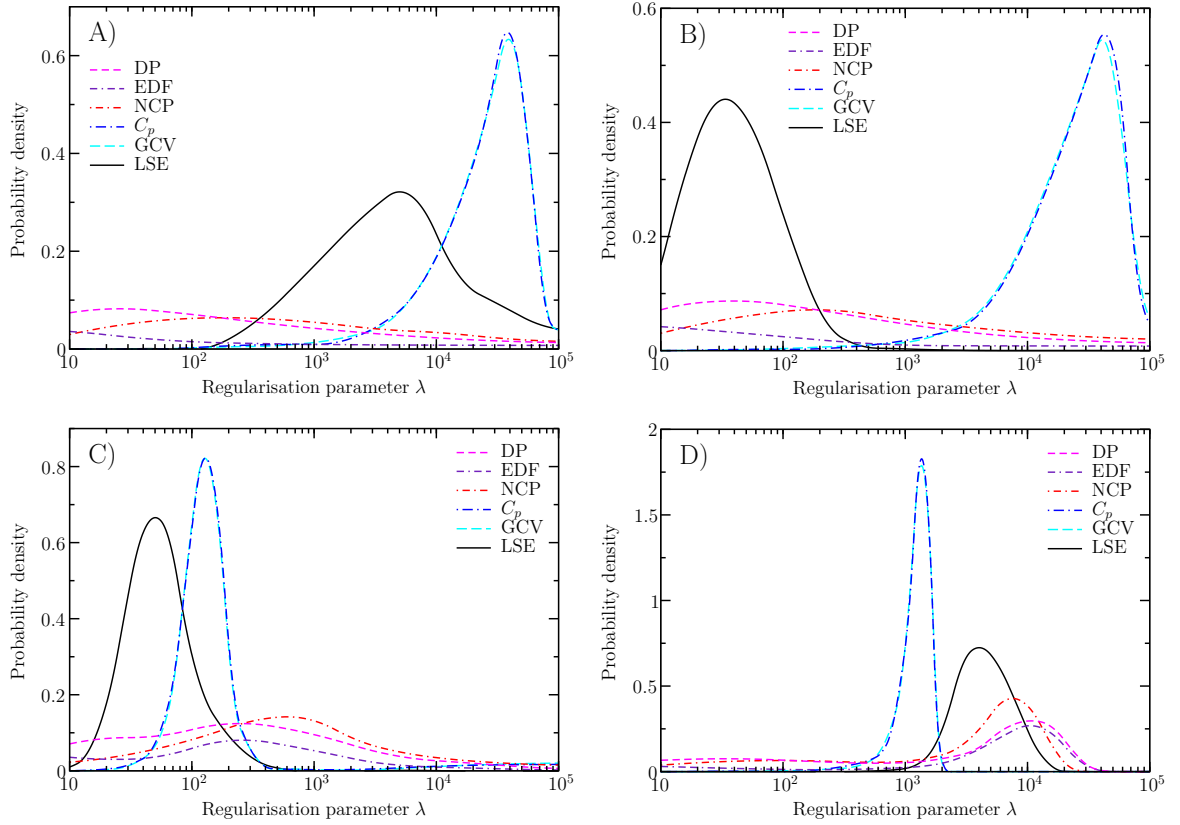


Figure 26. Same as Fig. 24 for inversions using the WMAP-5 and small-scale CMB data (combination 3, Sec. 3.1), and for spectra A–D Sec. B.3).

χ^2/N_d is close to unity. For example, the $\chi^2/N_d - \lambda$ curves are steepest for spectrum D at high λ , flatten out, and then steepen again slightly at low λ . This leads to a well-defined peak at $\lambda \simeq 10^4$, a trough at intermediate λ , and a small secondary peak at low λ . The DP distributions are so broad because the change in χ^2 with λ is small compared with the variation in χ^2 values at fixed λ between different reconstructions in the ensemble.

While most of the λ_{EDF} values are close to the optimum for spectrum D when the WMAP-5, small-scale CMB and SDSS LRG data are used, some greatly underestimate it. The effective number of parameters decreases from $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) \simeq 60$ for $\lambda = 10$ down to $\sum_{\mathbb{Z}} \text{Tr}(\mathbf{S}_{\mathbb{Z}}) \simeq 3$ for $\lambda = 10^5$. Consequently the $\langle \chi^2/N_{\text{eff}} \rangle - \lambda$ curves of the test spectra are higher at low λ than the $\langle \chi^2/N_d \rangle - \lambda$ curves but are almost identical at high λ . This means that data realisations with low λ_{DP} values have even lower λ_{EDF} values, while realisations with high λ_{DP} values have approximately the same high values of λ_{EDF} . This is reflected in the EDF distributions which extend down to lower λ than their DP counterparts.

The NCP method gives similar results to the DP method. This is because for Gaussian distributed data the χ^2 of the reconstruction has the value $N_d \pm \sqrt{(2N_d)}$ when $\tilde{\mathbf{y}}$ behaves like white noise.

In summary, the λ values of the DP, EDF and NCP methods are too widely scattered to be meaningful and bear little relation to the optimum figure. Mallows C_p method and the GCV method perform very well in minimising the PE of the reconstruction. Unfortunately

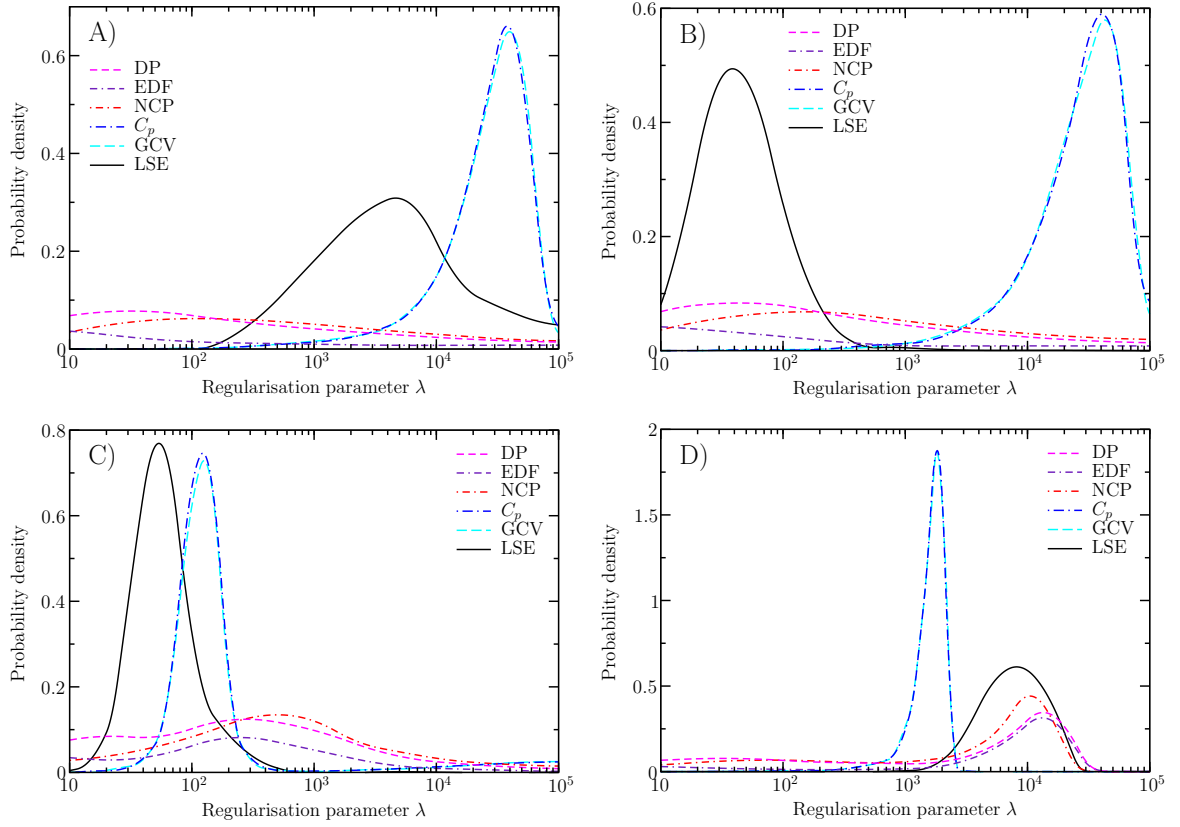


Figure 27. Same as Fig. 24 for inversions using the WMAP-5, small-scale CMB and SDSS-4 LRG data (combination 4, Sec. 3.1), and for spectra A–D Sec. B.3).

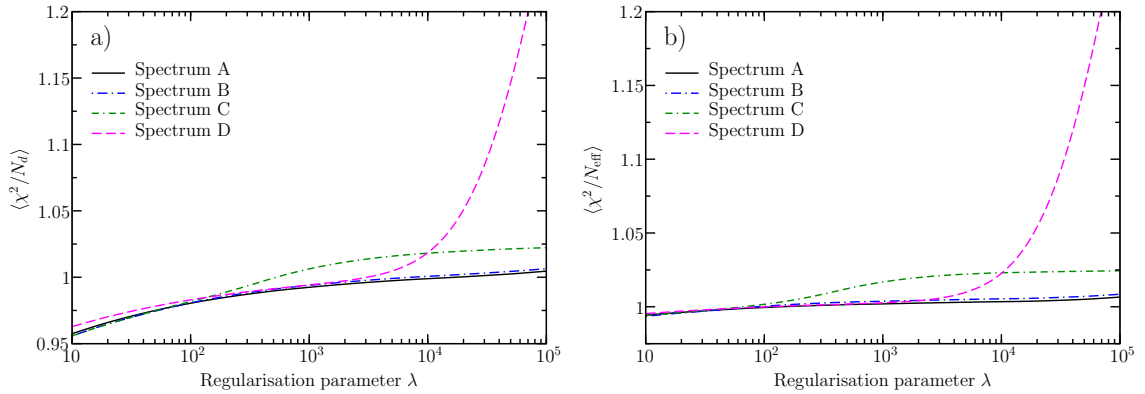


Figure 28. The left panel shows the ensemble mean of $\chi^2(\mathbf{p})/N_d$ as a function of the regularisation parameter, for reconstructions of the four test spectra using the WMAP-5, small-scale CMB and SDSS-4 LRG data. The right panel shows the same for the quantity $\chi^2(\mathbf{p})/N_{\text{eff}}$.

this is of limited use as the PE is usually different to the SE for CMB data.

E Glossary

We list below the notation and symbols used along with where they are first mentioned:

Symbol	Reference	Description
\mathbf{b}	Sec. C.3	Second-order approximation to the bias of $\hat{\mathbf{p}}$
$\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4$	Sec. C.3	Different components of \mathbf{b}
$C(k_0; k)$	eq. (3.1)	Correlation function of $\hat{\mathbf{p}}$
$d_a^{(\mathbb{Z})}$	eq. (2.2)	Data element a of data set \mathbb{Z} (whole vector is labelled $\mathbf{d}_{\mathbb{Z}}$)
\mathbf{d}	Sec. 2.1	Vector of all data sets $\mathbf{d}_{\mathbb{Z}}$
\mathbf{D}	Sec. 3.2	Matrix whose square is the (diagonal) covariance matrix of $\tilde{\mathbf{q}}$
\mathbf{G}	Sec. 3.2	Matrix relating $\tilde{\mathbf{q}}$ to \mathbf{q}
\mathbf{h}	eq. (D.9)	Normalised cumulative periodogram of \mathbf{y}
$\mathcal{K}_a^{(\mathbb{Z})}$	eq. (2.2)	Integral kernel for element a of data set \mathbb{Z}
L_{bps}	eq. (B.5)	Beam and point source component of L_{TT}
L_{Gibbs}	Sec. B.1.1	Low- ℓ Gibbs sampler component of L_{TT}
L_{LRG}	eq. (B.17)	SDSS LRG likelihood
L_{pol}	Sec. B.1.1	Total WMAP polarisation likelihood
L_{pix}	Sec. B.1.1	Low- ℓ pixel-based component of L_{pol}
L_{ss}	eq. (B.9)	Small angular scale CMB likelihood
L_{pTE}	eq. (B.6)	High- ℓ component of L_{pol}
L_{pTT}	eq. (B.3)	High- ℓ component of L_{TT}
L_{TT}	Sec. B.1.1	Total WMAP TT likelihood
$L(\mathbf{p}, \boldsymbol{\theta}, \mathbf{d})$	eq. (2.12)	$\equiv -2 \ln \mathcal{L}(\mathbf{p}, \boldsymbol{\theta} \mathbf{d})$
$\mathcal{L}(\mathbf{p}, \boldsymbol{\theta} \mathbf{d})$	eq. (2.9)	Likelihood of \mathbf{p} and $\boldsymbol{\theta}$ given \mathbf{d}
\mathbf{L}	eq. (2.14)	Matrix representation of the first-order derivative operator
$\text{MPE}(\hat{\mathbf{p}})$	eq. (A.32)	Mean predictive error of $\hat{\mathbf{p}}$
$\text{MSE}(\hat{\mathbf{p}})$	eq. (A.31)	Mean squared error of $\hat{\mathbf{p}}$
$n_a^{(\mathbb{Z})}$	eq. (2.2)	Noise element a of data set \mathbb{Z} (whole vector is labelled $\mathbf{n}_{\mathbb{Z}}$)
\mathbf{n}	Sec. 2.1	Vector of all noise sets $\mathbf{n}_{\mathbb{Z}}$
N	Sec. 2.1	Number of data sets
N_d	Sec. 2.1	Total number of data points
N_{eff}	Sec. D.1	Number of degrees of freedom of $\hat{\mathbf{p}}$
N_j		Number of basis functions $\phi_i(k)$
$N_{\mathbb{Z}}$	Sec. 2.1	Number of points in the set data set \mathbb{Z}
\mathbf{N}	Sec. 2.1	Matrix assembled from all $\mathbf{N}_{\mathbb{Z}}$
$\mathbf{N}_{\mathbb{Z}}$	Sec. 2.1	Covariance matrix of $\mathbf{n}_{\mathbb{Z}}$
$\mathcal{N}_{\mathcal{R}}(k)$	eq. (2.3)	Infinite-dimensional null space of functions
p_i	eq. (2.4)	Coefficients of the basis functions $\phi_i(k)$ (also labelled \mathbf{p})
P_{W1}	eq. (B.38)	One-dimensional Wishart distribution
P_{W2}	eq. (B.40)	Two-dimensional Wishart distribution
$\hat{\mathbf{p}}$	Sec. 2.1	Estimator of \mathbf{p}
$\hat{\mathbf{p}}_0$	Sec. 2.1	Vector with the minimum R value of those that maximise the likelihood. Given by $\hat{\mathbf{p}}_0 = \mathbf{W}_{\text{NR}}^\dagger \mathbf{d}$
$\hat{\mathbf{p}}^{(\text{f})}$	eq. (C.2)	First-order approximation to $\hat{\mathbf{p}}$
$\hat{\mathbf{p}}_{\text{MAP}}$	Sec. A.1	Maximum <i>a posteriori</i> estimate of \mathbf{p}
$\hat{\mathbf{p}}^{(\text{s})}$	eq. (C.3)	Second-order approximation to $\hat{\mathbf{p}}$

\mathbf{p}_t	Sec. A	True value of \mathbf{p}
$\hat{\mathbf{p}}^{(z)}$	eq. (C.1)	Zeroth-order approximation to $\hat{\mathbf{p}}$
$C_p(\lambda)$	eq. (D.12)	Mallow's C_p statistic
$\mathcal{P}_{\mathcal{R}}$	eq. (2.1)	Primordial power spectrum (PPS) of \mathcal{R}
$\text{PE}(\hat{\mathbf{p}})$	Sec. A.2	Predictive error of $\hat{\mathbf{p}}$
\mathbf{q}	Sec. 3.2	Correlated bandpowers of $\hat{\mathbf{p}}$
$\tilde{\mathbf{q}}$	Sec. 3.2	Uncorrelated bandpowers of $\hat{\mathbf{p}}$
$Q(\mathbf{p}, \mathbf{d}, \hat{\boldsymbol{\theta}}, \lambda)$	eq. (2.12)	$\equiv L(\mathbf{p}, \hat{\boldsymbol{\theta}}, \mathbf{d}) + \lambda R(\mathbf{p})$
$R(k_0; k)$	eq. (A.6)	First-order resolution function
$R(\mathbf{p})$	Sec. 2.1	Roughness function of \mathbf{p}
\mathbf{R}	eq. (A.3)	First-order resolution matrix
\mathcal{R}	eq. (2.1)	Primordial comoving curvature perturbation
\mathfrak{R}	Sec. C.3	Number of Monte Carlo simulations
s_k, c_k	Sec. D.1	Sine and cosine coefficients of the DFT of \mathbf{y}
$\text{SE}(\hat{\mathbf{p}})$	eq. (A.29)	Squared error of $\hat{\mathbf{p}}$
T_1	eq. (3.3)	Test statistic for a feature in the PPS (two-tailed)
T_2	eq. (3.5)	Test statistic for a feature in the PPS (one-tailed)
\mathbf{T}	Sec. 3.2	Matrix relating \mathbf{q} to $\hat{\mathbf{p}}$
$\mathcal{T}(\mathbf{p}_t, \boldsymbol{\theta}_t, \hat{\boldsymbol{\theta}}, \mathbf{n})$	eq. (A.1)	Transfer function relating $\hat{\mathbf{p}}$ to \mathbf{p}_t
u_α	Sec. 2.1	Element α of uncertainty in $\hat{\boldsymbol{\theta}}$ (whole vector is labelled \mathbf{u})
\mathbf{U}	Sec. 2.1	Covariance matrix for $\hat{\boldsymbol{\theta}}$
$V_{\text{GCV}}(\lambda)$	eq. (D.15)	Generalised cross-validation statistic
$W_{ai}^{(\mathbb{Z})}$	eq. (2.7)	Integral of $\mathcal{K}_a^{(\mathbb{Z})}$ over wavenumber bin i (also called $W_{\mathbb{Z}}$)
\mathbf{W}	eq. (2.8)	Matrix assembled from all $W_{\mathbb{Z}}$
$\mathbf{W}_{\text{NF}}^\dagger$	eq. (2.11)	Weighted Moore-Penrose inverse of \mathbf{W}
\mathbf{y}	Sec. D.1	Vector of all data sets $\mathbf{y}_{\mathbb{Z}}$
$\mathbf{y}_{\mathbb{Z}}$	Sec. D.1	Data set $\mathbf{d}_{\mathbb{Z}}$ whitened with a Cholesky decomposition
\mathbf{Y}	eq. (A.3)	Second-order resolution matrix
$Y(k_0; k_1, k_2)$	eq. (A.8)	Second-order resolution function
γ	eq. (B.15)	SDSS LRG scaling factor
$\gamma_{\text{NCP}}(\lambda)$	Sec. D.1	Normalised cumulative periodogram statistic
$\delta\hat{\mathbf{p}}^{(f)}$	eq. (C.4)	First-order approximation difference
$\delta\hat{\mathbf{p}}^{(s)}$	eq. (C.5)	Second-order approximation difference
λ	eq. (2.12)	Regularisation parameter
λ_{CP}	Sec. D.1	Mallow's C_p method λ value
λ_{DP}	Sec. D.1	Discrepancy principle method λ value
λ_{EDF}	Sec. D.1	Equivalent degrees of freedom method λ value
λ_{GCV}	Sec. D.1	Generalised cross-validation method λ value
λ_{LSE}	Sec. D.1	Least squared error method λ value
λ_{NCP}	Sec. D.1	Normalised cumulative periodogram method λ value
ν_1, ν_2, ν_3	Sec. D	Estimates of the effective number of free parameters of $\hat{\mathbf{p}}$
$\phi_i(k)$	eq. (2.5)	Basis functions for the reconstructed PPS
ρ	Sec. 2.1	Rank of \mathbf{W}
$\boldsymbol{\theta}$	eq. (2.2)	Set of parameters defining background cosmology

$\hat{\boldsymbol{\theta}}$	Sec. 2.1	Estimator of $\boldsymbol{\theta}$
$\boldsymbol{\theta}_t$	Sec. A	True value of $\boldsymbol{\theta}$
σ_i	Sec. 2.1	Weighted singular values of \mathbf{W}
$\xi(k)$	Sec. 3.4	Integral kernel specifying a feature in the PPS
Γ	Sec. 2.1	Matrix specifying \mathbf{R}
Σ_{B}	eq. (A.28)	Bayesian covariance matrix of $\hat{\mathbf{p}}$
$\Sigma^{(\text{f})}$	eq. (C.14)	Frequentist covariance matrix of $\hat{\mathbf{p}}^{(\text{f})}$
$\Sigma^{(\text{fs})}$	eq. (C.20)	Frequentist covariance matrix of $\delta\hat{\mathbf{p}}^{(\text{f})}$ and $\delta\hat{\mathbf{p}}^{(\text{s})}$
Σ_{F}	eq. (A.10)	Frequentist covariance matrix due to data noise
Σ_{N}	Sec. 3.2	Frequentist covariance matrix of \mathbf{q}
Σ_{P}	eq. (A.11)	Frequentist covariance matrix due to background parameter uncertainties
$\Sigma^{(\text{s})}$	eq. (C.16)	Frequentist covariance matrix of $\hat{\mathbf{p}}^{(\text{s})}$
$\Sigma^{(\text{ss})}$	eq. (C.18)	Frequentist covariance matrix of $\delta\hat{\mathbf{p}}^{(\text{s})}$
Σ_{T}	eq. (A.9)	Total frequentist covariance matrix of $\hat{\mathbf{p}}$

References

- [1] H. V. Peiris *et al.* [WMAP Collaboration], *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for inflation*, *Astrophys. J. Suppl.* **148**, 213 (2003) [arXiv:astro-ph/0302225].
- [2] S. Dodelson, *Coherent phase argument for inflation*, *AIP Conf. Proc.* **689**, 184 (2003) [arXiv:hep-ph/0309057].
- [3] E. Komatsu *et al.* [WMAP Collaboration], *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, *Astrophys. J. Suppl.* **180**, 330 (2009) [arXiv:0803.0547].
- [4] A. D. Linde, *Particle physics and inflationary cosmology*, *Contemp. Concepts Phys.* **5**, 1 (1990) [arXiv:hep-ph/0503203].
- [5] D. H. Lyth and A. Riotto, *Particle physics models of inflation and the cosmological density perturbation*, *Phys. Rept.* **314**, 1 (1999) [arXiv:hep-ph/9807278].
- [6] S. Weinberg, *The Cosmological Constant Problem*, *Rev. Mod. Phys.* **61**, 1 (1989).
- [7] S. Nobbenhuis, *Categorizing different approaches to the cosmological constant problem*, *Found. Phys.* **36** (2006) 613 [arXiv:gr-qc/0411093].
- [8] G. German, G. G. Ross and S. Sarkar, *Low-scale inflation*, *Nucl. Phys. B* **608**, 423 (2001) [arXiv:hep-ph/0103243].
- [9] S. Hotchkiss, G. German, G. G. Ross and S. Sarkar, *Fine tuning and the ratio of tensor to scalar density fluctuations from cosmological inflation*, *JCAP* **10**, 015 (2008) [arXiv:0804.2634].
- [10] S. Bird, H. V. Peiris and R. Easther, *Fine-tuning criteria for inflation and the search for primordial gravitational waves*, *Phys. Rev. D* **78**, 083518 (2008) [arXiv:0807.3745].
- [11] W. H. Kinney, *How to fool cosmic microwave background parameter estimation*, *Phys. Rev. D* **63**, 043001 (2001) [arXiv:astro-ph/0005410].
- [12] E. Komatsu *et al.* [WMAP Collaboration], *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, *Astrophys. J. Suppl.* **192**, 18 (2011) [arXiv:1001.4538].

- [13] A. Blanchard, M. Douspis, M. Rowan-Robinson and S. Sarkar, *An alternative to the cosmological ‘concordance model’*, Astron. Astrophys. **412**, 35 (2003) [arXiv:astro-ph/0304237].
- [14] P. Hunt and S. Sarkar, *Multiple inflation and the WMAP ‘glitches’ II. Data analysis and cosmological parameter extraction*, Phys. Rev. D **76**, 123504 (2007) [arXiv:0706.2443].
- [15] P. Hunt and S. Sarkar, *Constraints on large scale inhomogeneities from WMAP-5 and SDSS: confrontation with recent observations*, Mon. Not. Roy. Astron. Soc. **401**, 547 (2010) [arXiv:0807.4508].
- [16] S. Alexander, T. Biswas, A. Notari and D. Vaid, *Local Void vs Dark Energy: Confrontation with WMAP and Type Ia Supernovae*, JCAP **0909**, 025 (2009) [arXiv:0712.0370].
- [17] S. Nadathur and S. Sarkar, *Reconciling the local void with the CMB*, Phys. Rev. D **83**, 063506 (2011) [arXiv:1012.3460].
- [18] S. Nadathur, S. Hotchkiss and S. Sarkar, *The integrated Sachs-Wolfe imprints of cosmic superstructures: a problem for Λ CDM*, JCAP **1206**, 042 (2012) [arXiv:1109.4126].
- [19] P. A. R. Ade *et al.* [Planck Collaboration], *Planck 2013 results. XIX. The integrated Sachs-Wolfe effect*, [arXiv:1303.5079].
- [20] F. Finelli, J. Hamann, S. M. Leach and J. Lesgourgues, *Single-field inflation constraints from CMB and SDSS data*, JCAP **1004**, 011 (2010) [arXiv:0912.0522].
- [21] L. Alabidi and I. Huston, *An update on single field models of inflation in light of WMAP7*, JCAP **1008**, 037 (2010) [arXiv:1004.4794].
- [22] J. A. Adams, G. G. Ross and S. Sarkar, *Multiple inflation*, Nucl. Phys. B **503**, 405 (1997) [arXiv:hep-ph/9704286].
- [23] P. Hunt and S. Sarkar, *Multiple inflation and the WMAP ‘glitches’*, Phys. Rev. D **70**, 103518 (2004) [arXiv:astro-ph/0408138].
- [24] S. Hotchkiss and S. Sarkar, *Non-Gaussianity from violation of slow-roll in multiple inflation*, JCAP **1005**, 024 (2010) [arXiv:0910.3373].
- [25] D. J. H. Chung, E. W. Kolb, A. Riotto and I. I. Tkachev, *Probing Planckian physics: Resonant production of particles during inflation and features in the primordial power spectrum*, Phys. Rev. D **62**, 043508 (2000) [arXiv:hep-ph/9910437].
- [26] G. J. Mathews, D. J. H. Chung, K. Ichiki, T. Kajino and M. Orito, *Constraints on resonant particle production during inflation from the matter and CMB power spectra*, Phys. Rev. D **70**, 083505 (2004) [arXiv:astro-ph/0406046].
- [27] A. E. Romano and M. Sasaki, *Effects of particle production during inflation*, Phys. Rev. D **78**, 103522 (2008) [arXiv:0809.5142].
- [28] N. Barnaby, Z. Huang, L. Kofman and D. Pogosyan, *Cosmological fluctuations from infra-red cascading during inflation*, Phys. Rev. D **80**, 043501 (2009) [arXiv:0902.0615].
- [29] N. Barnaby, *Nongaussianity from particle production during inflation*, Adv. Astron. **2010**, 156180 (2010) [arXiv:1010.5507].
- [30] A. A. Starobinsky, *Spectrum of adiabatic perturbations in the universe when there are singularities in the inflation potential*, JETP Lett. **55**, 489 (1992)
- [31] S. M. Leach, M. Sasaki, D. Wands and A. R. Liddle, *Enhancement of superhorizon scale inflationary curvature perturbations*, Phys. Rev. D **64**, 023512 (2001) [arXiv:astro-ph/0101406].
- [32] J. O. Gong, *Breaking scale invariance from a singular inflaton potential*, JCAP **0507**, 015 (2005) [arXiv:astro-ph/0504383].
- [33] J. A. Adams, B. Cresswell and R. Easther, *Inflationary perturbations from a potential with a step*, Phys. Rev. D **64**, 123514 (2001) [arXiv:astro-ph/0102236].

- [34] X. Chen, R. Easther and E. A. Lim, *Large non-gaussianities in single field inflation*, JCAP **0706**, 023 (2007) [arXiv:astro-ph/0611645].
- [35] X. Chen, R. Easther and E. A. Lim, *Generation and characterization of large non-gaussianities in single field inflation*, JCAP **0804**, 010 (2008) [arXiv:0801.3295].
- [36] R. N. Lerner and J. McDonald, *Space-dependent step features: Transient breakdown of slow-roll, homogeneity and isotropy during inflation*, Phys. Rev. D **79**, 023511 (2009) [arXiv:0811.1933].
- [37] C. Dvorkin and W. Hu, *Generalized slow roll for large power spectrum features*, Phys. Rev. D **81**, 023518 (2010) [arXiv:0910.2237].
- [38] Z. -G. Liu, J. Zhang and Y. -S. Piao, *Phantom inflation with a steplike potential*, Phys. Lett. B **697**, 407 (2011) [arXiv:1012.0673].
- [39] P. Adshead, C. Dvorkin, W. Hu and E. A. Lim, *Non-gaussianity from step features in the inflationary potential*, Phys. Rev. D **85**, 023531 (2012) [arXiv:1110.3050].
- [40] H. M. Hodges, G. R. Blumenthal, L. A. Kofman and J. R. Primack, *Nonstandard primordial fluctuations from a polynomial inflaton potential*, Nucl. Phys. B **335**, 197 (1990).
- [41] S. M. Leach and A. R. Liddle, *Inflationary perturbations near horizon crossing*, Phys. Rev. D **63**, 043508 (2001) [arXiv:astro-ph/0010082].
- [42] A. Ashoorioon and A. Krause, *Power Spectrum and Signatures for Cascade Inflation*, [arXiv:hep-ph/0607001].
- [43] M. Joy, V. Sahni and A. A. Starobinsky, *A new universal local feature in the inflationary perturbation spectrum*, Phys. Rev. D **77**, 023514 (2008) [arXiv:0711.1585].
- [44] R. K. Jain, P. Chingangbam and L. Sriramkumar, *On the evolution of tachyonic perturbations at super-Hubble scales*, JCAP **0710**, 003 (2007) [arXiv:astro-ph/0703762].
- [45] R. Bean, X. Chen, G. Hailu, S. H. Tye and J. Xu, *Duality cascade in brane inflation*, JCAP **0803**, 026 (2008) [arXiv:0802.0491].
- [46] R. Saito, J. 'i. Yokoyama and R. Nagata, *Single-field inflation, anomalous enhancement of superhorizon fluctuations, and non-Gaussianity in primordial black hole formation*, JCAP **0806**, 024 (2008) [arXiv:0804.3470].
- [47] S. -H. H. Tye and J. Xu, *A meandering inflaton*, Phys. Lett. B **683**, 326 (2010) [arXiv:0910.0849].
- [48] A. Achucarro, J. -O. Gong, S. Hardeman, G. A. Palma and S. P. Patil, *Features of heavy physics in the CMB power spectrum*, JCAP **1101**, 030 (2011) [arXiv:1010.3693].
- [49] G. Goswami and T. Souradeep, *Power spectrum nulls due to non-standard inflationary evolution*, Phys. Rev. D **83**, 023526 (2011) [arXiv:1011.4914].
- [50] P. Brax and E. Cluzel, *Perturbation Theory in k -Inflation Coupled to Matter*, JCAP **1104**, 014 (2011) [arXiv:1102.1917].
- [51] F. Arroja, A. E. Romano and M. Sasaki, *Large and strong scale dependent bispectrum in single field inflation from a sharp feature in the mass*, Phys. Rev. D **84**, 123503 (2011) [arXiv:1106.5384].
- [52] J. Liu and Y. -S. Piao, *A multiple step-like spectrum of primordial perturbation*, Phys. Lett. B **705**, 1 (2011) [arXiv:1106.5608].
- [53] A. Chambers and A. Rajantie, *Lattice calculation of non-Gaussianity from preheating*, Phys. Rev. Lett. **100**, 041302 (2008) [Erratum-ibid. **101**, 149903 (2008)] [arXiv:0710.4133].
- [54] J. R. Bond, A. V. Frolov, Z. Huang and L. Kofman, *Non-Gaussian Spikes from Chaotic Billiards in Inflation Preheating*, Phys. Rev. Lett. **103**, 071301 (2009) [arXiv:0903.3407].

- [55] C. P. Burgess, J. M. Cline, F. Lemieux and R. Holman, *Are inflationary predictions sensitive to very high-energy physics?*, JHEP **0302**, 048 (2003) [arXiv:hep-ph/0210233].
- [56] Y. -S. Piao, B. Feng and X. -m. Zhang, *Suppressing CMB quadrupole with a bounce from contracting phase to inflation*, Phys. Rev. D **69**, 103520 (2004) [arXiv:hep-ph/0310206].
- [57] B. A. Powell and W. H. Kinney, *The pre-inflationary vacuum in the cosmic microwave background*, Phys. Rev. D **76**, 063512 (2007) [arXiv:astro-ph/0612006].
- [58] G. Nicholson and C. R. Contaldi, *The large scale CMB cut-off and the tensor-to-scalar ratio*, JCAP **0801**, 002 (2008) [arXiv:astro-ph/0701783].
- [59] A. Lasenby and C. Doran, *Closed universes, de Sitter space and inflation*, Phys. Rev. D **71**, 063502 (2005) [arXiv:astro-ph/0307311].
- [60] G. Dvali and S. Kachru, *Large scale power and running spectral index in new old inflation*, arXiv:hep-ph/0310244.
- [61] D. Langlois and F. Vernizzi, *From heaviness to lightness during inflation*, JCAP **0501**, 002 (2005) [arXiv:astro-ph/0409684].
- [62] R. H. Brandenberger and J. Martin, *The robustness of inflation to changes in super-Planck-scale physics*, Mod. Phys. Lett. A **16**, 999 (2001) [arXiv:astro-ph/0005432].
- [63] U. H. Danielsson, *A Note on inflation and transPlanckian physics*, Phys. Rev. D **66**, 023511 (2002) [arXiv:hep-ph/0203198].
- [64] K. Schalm, G. Shiu and J. P. van der Schaar, *The Cosmological vacuum ambiguity, effective actions, and transplanckian effects in inflation*, AIP Conf. Proc. **743**, 362 (2005) [arXiv:hep-ph/0412288].
- [65] B. Greene, K. Schalm, J. P. van der Schaar and G. Shiu, *Extracting new physics from the CMB*, eConf C **041213**, 0001 (2004) [arXiv:astro-ph/0503458].
- [66] R. Easther, W. H. Kinney and H. Peiris, *Boundary effective field theory and trans-Planckian perturbations: Astrophysical implications*, JCAP **0508**, 001 (2005) [arXiv:astro-ph/0505426].
- [67] X. Wang, B. Feng, M. Li, X. L. Chen and X. Zhang, *Natural inflation, Planck scale physics and oscillating primordial spectrum*, Int. J. Mod. Phys. D **14**, 1347 (2005) [arXiv:astro-ph/0209242].
- [68] N. Kaloper and M. Kaplinghat, *Primeval corrections to the CMB anisotropies*, Phys. Rev. D **68**, 123522 (2003) [arXiv:hep-ph/0307016].
- [69] R. Flauger, L. McAllister, E. Pajer, A. Westphal and G. Xu, *Oscillations in the CMB from Axion Monodromy Inflation*, JCAP **1006**, 009 (2010) [arXiv:0907.2916].
- [70] T. Biswas, A. Mazumdar and A. Shafieloo, *Wiggles in the cosmic microwave background radiation: echoes from non-singular cyclic-inflation*, Phys. Rev. D **82**, 123517 (2010) [arXiv:1003.3206].
- [71] M. G. Jackson and K. Schalm, *Model-Independent Signatures of New Physics in Slow-Roll Inflation*, arXiv:1104.0887.
- [72] X. Chen, *Primordial Features as Evidence for Inflation*, JCAP **1201**, 038 (2012) [arXiv:1104.1323].
- [73] K. Ichiki, R. Nagata and J. 'i. Yokoyama, *Cosmic discordance: Detection of a modulation in the primordial fluctuation spectrum*, Phys. Rev. D **81**, 083010 (2010) [arXiv:0911.5108].
- [74] K. Kumazaki, K. Ichiki, N. Sugiyama and J. Silk, *Exploring the origin of the fine structures in the CMB temperature angular power spectrum*, Phys. Rev. D **87**, 023008 (2013) [arXiv:1211.3097].
- [75] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith and R. S. Hill *et al.*, *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP)*

- Observations: Final Maps and Results*, *Astrophys. J. Suppl.* **208** (2013) 20 [arXiv:1212.5225].
- [76] C. R. Contaldi, M. Peloso, L. Kofman and A. Linde, *Suppressing the lower Multipoles in the CMB Anisotropies*, *JCAP* **0307**, 002 (2003) [arXiv:astro-ph/0303636].
 - [77] J. M. Cline, P. Crotty and J. Lesgourgues, *Does the small CMB quadrupole moment suggest new physics?*, *JCAP* **0309**, 010 (2003) [arXiv:astro-ph/0304558].
 - [78] B. Feng and X. Zhang, *Double inflation and the low CMB quadrupole*, *Phys. Lett. B* **570**, 145 (2003) [arXiv:astro-ph/0305020].
 - [79] M. Kawasaki and F. Takahashi, *Inflation model with lower multipoles of the CMB suppressed*, *Phys. Lett. B* **570**, 151 (2003) [arXiv:hep-ph/0305319].
 - [80] M. Bastero-Gil, K. Freese and L. Mersini-Houghton, *What can WMAP tell us about the very early universe? New physics as an explanation of suppressed large scale power and running spectral index*, *Phys. Rev. D* **68**, 123514 (2003) [arXiv:hep-ph/0306289].
 - [81] S. Hannestad and L. Mersini-Houghton, *A first glimpse of string theory in the sky?*, *Phys. Rev. D* **71**, 123504 (2005) [arXiv:hep-ph/0405218].
 - [82] M. Liguori, S. Matarrese, M. Musso and A. Riotto, *Stochastic inflation and the lower multipoles in the CMB anisotropies*, *JCAP* **0408**, 011 (2004) [arXiv:astro-ph/0405544].
 - [83] R. Sinha and T. Souradeep, *Post-WMAP assessment of infrared cutoff in the primordial spectrum from inflation*, *Phys. Rev. D* **74**, 043518 (2006) [arXiv:astro-ph/0511808].
 - [84] M. Joy, A. Shafieloo, V. Sahni and A. A. Starobinsky, *Is a step in the primordial spectral index favored by CMB data?*, *JCAP* **0906**, 028 (2009) [arXiv:0807.3334].
 - [85] R. K. Jain, P. Chingangbam, J. -O. Gong, L. Sriramkumar and T. Souradeep, *Punctuated inflation and the low CMB multipoles*, *JCAP* **0901**, 009 (2009) [arXiv:0809.3915].
 - [86] J. Martin and C. Ringeval, *Superimposed oscillations in the WMAP data?*, *Phys. Rev. D* **69**, 083515 (2004) [arXiv:astro-ph/0310382].
 - [87] J. Martin and C. Ringeval, *Exploring the superimposed oscillations parameter space*, *JCAP* **0501**, 007 (2005) [arXiv:hep-ph/0405249].
 - [88] M. Kawasaki, F. Takahashi and T. Takahashi, *Making waves on CMB power spectrum and inflaton dynamics*, *Phys. Lett. B* **605**, 223 (2005) [arXiv:astro-ph/0407631].
 - [89] L. Covi, J. Hamann, A. Melchiorri, A. Slosar and I. Sorbera, *Inflation and WMAP three year data: Features have a Future!*, *Phys. Rev. D* **74**, 083509 (2006) [arXiv:astro-ph/0606452].
 - [90] J. Hamann, L. Covi, A. Melchiorri and A. Slosar, *New constraints on oscillations in the primordial spectrum of inflationary perturbations*, *Phys. Rev. D* **76**, 023503 (2007) [arXiv:astro-ph/0701380].
 - [91] L. Hoi and J. M. Cline, *Testing for features in the primordial power spectrum*, *Int. J. Mod. Phys. D* **18**, 1863 (2009) [arXiv:0706.3887].
 - [92] D. K. Hazra, M. Aich, R. K. Jain, L. Sriramkumar and T. Souradeep, *Primordial features due to a step in the inflaton potential*, *JCAP* **1010**, 008 (2010) [arXiv:1005.2175].
 - [93] M. Nakashima, R. Saito, Y. -i. Takamizu and J. 'i. Yokoyama, *The effect of varying sound velocity on primordial curvature perturbations*, *Prog. Theor. Phys.* **125**, 1035 (2011) [arXiv:1009.4394].
 - [94] M. Aich, D. K. Hazra, L. Sriramkumar and T. Souradeep, *Oscillations in the inflaton potential: Exact numerical analysis and comparison with the recent and forthcoming CMB datasets*, *Phys. Rev. D* **87**, 083526 (2013) [arXiv:1106.2798].
 - [95] M. Benetti, M. Lattanzi, E. Calabrese and A. Melchiorri, *Features in the primordial spectrum: new constraints from WMAP7+ACT data and prospects for Planck*, *Phys. Rev. D* **84**, 063509

- (2011) [arXiv:1107.4992].
- [96] P. D. Meerburg, R. Wijers and J. P. van der Schaar, *WMAP 7 Constraints on Oscillations in the Primordial Power Spectrum*, Mon. Not. Roy. Astron. Soc. **421**, 369 (2012) [arXiv:1109.5264].
 - [97] K. Ichiki and R. Nagata, *Brute force reconstruction of the primordial fluctuation spectrum from five-year Wilkinson Microwave Anisotropy Probe observations*, Phys. Rev. D **80**, 083002 (2009).
 - [98] P. Mukherjee and Y. Wang, *Wavelet Band Powers of the Primordial Power Spectrum from CMB Data*, Astrophys. J. **593**, 38 (2003) [arXiv:astro-ph/0301058].
 - [99] P. Mukherjee and Y. Wang, *Direct wavelet expansion of the primordial power spectrum: results from pre-MAP CMB data*, Astrophys. J. **598**, 779 (2003) [arXiv:astro-ph/0301562].
 - [100] P. Mukherjee and Y. Wang, *Model-independent reconstruction of the primordial power spectrum from WMAP data*, Astrophys. J. **599**, 1 (2003) [arXiv:astro-ph/0303211].
 - [101] P. Mukherjee and Y. Wang, *Primordial power spectrum reconstruction*, JCAP **0512**, 007 (2005) [arXiv:astro-ph/0502136].
 - [102] S. Leach, *Measuring the primordial power spectrum: Principal component analysis of the cosmic microwave background*, Mon. Not. Roy. Astron. Soc. **372**, 646 (2006) [arXiv:astro-ph/0506390].
 - [103] C. Sealfon, L. Verde and R. Jimenez, *Smoothing spline primordial power spectrum reconstruction*, Phys. Rev. D **72**, 103520 (2005) [arXiv:astro-ph/0506707].
 - [104] L. Verde and H. V. Peiris, *On minimally-parametric primordial power spectrum reconstruction and the evidence for a red tilt*, JCAP **0807**, 009 (2008) [arXiv:0802.1219].
 - [105] H. V. Peiris and L. Verde, *The shape of the primordial power spectrum: a last stand before Planck*, Phys. Rev. D **81**, 021302 (2010) [arXiv:0912.0268].
 - [106] S. Bird, H. V. Peiris, M. Viel and L. Verde, *Minimally parametric power spectrum reconstruction from the lyman-alpha forest*, Mon. Not. Roy. Astron. Soc. **413**, 1717 (2011) [arXiv:1010.1519].
 - [107] C. Gauthier and M. Bucher, *Reconstructing the primordial power spectrum from the CMB*, JCAP **1210**, 050 (2012) [arXiv:1209.2147].
 - [108] Y. Wang, D. N. Spergel and M. A. Strauss, *Cosmology in the next millennium: Combining MAP and SDSS data to constrain inflationary models*, Astrophys. J. **510**, 20 (1999) [arXiv:astro-ph/9802231].
 - [109] Y. Wang and G. Mathews, *A measurement of the primordial power spectrum from Maxima and Boomerang data*, Astrophys. J. **573**, 1 (2002) [arXiv:astro-ph/0011351].
 - [110] S. L. Bridle, A. M. Lewis, J. Weller and G. Efstathiou, *Reconstructing the primordial power spectrum*, Mon. Not. Roy. Astron. Soc. **342**, L72 (2003) [arXiv:astro-ph/0302306].
 - [111] S. Hannestad, *Reconstructing the primordial power spectrum - a new algorithm*, JCAP **0404**, 002 (2004) [arXiv:astro-ph/0311491].
 - [112] M. Bridges, A. N. Lasenby and M. P. Hobson, *A Bayesian analysis of the primordial power spectrum*, Mon. Not. Roy. Astron. Soc. **369**, 1123 (2006) [arXiv:astro-ph/0511573].
 - [113] D. N. Spergel *et al.* [WMAP Collaboration], *Wilkinson Microwave Anisotropy Probe (WMAP) three year results: Implications for cosmology*, Astrophys. J. Suppl. **170**, 377 (2007) [arXiv:astro-ph/0603449].
 - [114] M. Bridges, A. N. Lasenby and M. P. Hobson, *WMAP 3-year primordial power spectrum*, Mon. Not. Roy. Astron. Soc. **381**, 68 (2007) [arXiv:astro-ph/0607404].
 - [115] T. Chantavat, C. Gordon and J. Silk, *Probing the Primordial Power Spectrum with Cluster*

- Number Counts*, Phys. Rev. D **79**, 083508 (2009) [arXiv:0811.4371].
- [116] M. Bridges, F. Feroz, M. P. Hobson and A. N. Lasenby, *Bayesian optimal reconstruction of the primordial power spectrum*, Mon. Not. R. Astr. Soc. **400**, 1075 (2009) [arXiv:0812.3541].
 - [117] J. A. Vazquez, M. Bridges, M. P. Hobson and A. N. Lasenby, *Model selection applied to reconstruction of the Primordial Power Spectrum*, JCAP **1206**, 006 (2012) [arXiv:1203.1252].
 - [118] R. Hlozek, J. Dunkley, G. Addison, J. W. Appel, J. R. Bond, C. S. Carvalho, S. Das and M. Devlin *et al.*, *The Atacama Cosmology Telescope: a measurement of the primordial power spectrum*, Astrophys. J. **749**, 90 (2012) [arXiv:1105.4887].
 - [119] Z. -K. Guo, D. J. Schwarz and Y. -Z. Zhang, *Reconstruction of the primordial power spectrum from CMB data*, JCAP **1108**, 031 (2011) [arXiv:1105.5916].
 - [120] Z. -K. Guo and Y. -Z. Zhang, *Uncorrelated estimates of the primordial power spectrum*, JCAP **1111**, 032 (2011) [arXiv:1109.0067].
 - [121] Z. -K. Guo and Y. -Z. Zhang, *Primordial power spectrum versus extension parameters beyond the standard model*, Phys. Rev. D **85**, 103519 (2012) [arXiv:1201.1538].
 - [122] S. Hannestad, *Reconstructing the inflationary power spectrum from CMBR data*, Phys. Rev. D **63**, 043009 (2001) [arXiv:astro-ph/0009296].
 - [123] A. Berera and P. A. Martin, *Inverting the Sachs-Wolfe formula: an inverse problem arising in early-universe cosmology*, Inverse Problems **5**, 1393 (1999).
 - [124] M. Matsumiya, M. Sasaki and J. Yokoyama, *Cosmic inversion: Reconstructing primordial spectrum from CMB anisotropy*, Phys. Rev. D **65**, 083007 (2002) [arXiv:astro-ph/0111549].
 - [125] M. Matsumiya, M. Sasaki and J. Yokoyama, *Cosmic Inversion II –An iterative method for reproducing the primordial spectrum from the CMB data*, JCAP **0302**, 003 (2003) [arXiv:astro-ph/0210365].
 - [126] N. Kogo, M. Matsumiya, M. Sasaki and J. Yokoyama, *Reconstructing the primordial spectrum from WMAP data by the cosmic inversion method*, Astrophys. J. **607**, 32 (2004) [arXiv:astro-ph/0309662].
 - [127] N. Kogo, M. Sasaki and J. Yokoyama, *Reconstructing the Primordial Spectrum with CMB Temperature and Polarization*, Phys. Rev. D **70**, 103001 (2004) [arXiv:astro-ph/0409052].
 - [128] N. Kogo, M. Sasaki and J. Yokoyama, *Constraining cosmological parameters by the cosmic inversion method*, Prog. Theor. Phys. **114**, 555 (2005) [arXiv:astro-ph/0504471].
 - [129] R. Nagata and J. Yokoyama, *Reconstruction of the primordial fluctuation spectrum from the five-year WMAP data by the cosmic inversion method with band-power decorrelation analysis*, Phys. Rev. D **78**, 123002 (2008) [arXiv:0809.4537].
 - [130] G. Nicholson, C. R. Contaldi and P. Paykari, *Reconstruction of the primordial power spectrum by direct inversion*, JCAP **1001**, 016 (2010) [arXiv:0909.5092].
 - [131] A. Shafieloo and T. Souradeep, *Primordial power spectrum from WMAP*, Phys. Rev. D **70**, 043523 (2004) [arXiv:astro-ph/0312174].
 - [132] A. Shafieloo, T. Souradeep, P. Manimaran, P. K. Panigrahi and R. Rangarajan, *Features in the primordial spectrum from WMAP: A wavelet analysis*, Phys. Rev. D **75**, 123502 (2007) [arXiv:astro-ph/0611352].
 - [133] A. Shafieloo and T. Souradeep, *Estimation of primordial spectrum with post-WMAP 3 year data*, Phys. Rev. D **78**, 023511 (2008) [arXiv:0709.1944].
 - [134] G. Nicholson and C. R. Contaldi, *Reconstruction of the primordial power spectrum using temperature and polarisation data from multiple experiments*, JCAP **0907**, 011 (2009) [arXiv:0903.1106].

- [135] J. Hamann, A. Shafieloo and T. Souradeep, *Features in the primordial power spectrum? A frequentist analysis*, JCAP **1004**, 010 (2010) [arXiv:0912.2728].
- [136] C. Gibelyou, D. Huterer and W. Fang, *Detectability of large-scale power suppression in the galaxy distribution*, Phys. Rev. D **82**, 123009 (2010) [arXiv:1007.0757].
- [137] D. K. Hazra, A. Shafieloo and T. Souradeep, *Primordial power spectrum: a complete analysis with the WMAP nine-year data*, JCAP **1307**, 031 (2013) [arXiv:1303.4143].
- [138] D. K. Hazra, A. Shafieloo and T. Souradeep, *Cosmological parameter estimation with free-form primordial power spectrum*, Phys. Rev. D **87**, 123528 (2013) [arXiv:1303.5336].
- [139] G. Goswami and J. Prasad, *Maximum entropy deconvolution of primordial power spectrum*, Phys. Rev. D **88**, 023522 (2013) [arXiv:1303.4747].
- [140] M. Tegmark and M. Zaldarriaga, *Separating the early universe from the late universe: Cosmological parameter estimation beyond the black box*, Phys. Rev. D **66**, 103508 (2002) [arXiv:astro-ph/0207047].
- [141] D. Tocchini-Valentini, M. Douspis and J. Silk, *A new search for features in the primordial power spectrum*, Mon. Not. Roy. Astron. Soc. **359**, 31 (2005) [arXiv:astro-ph/0402583].
- [142] D. Tocchini-Valentini, Y. Hoffman and J. Silk, *Non-parametric reconstruction of the primordial power spectrum at horizon scales from WMAP data*, Mon. Not. Roy. Astron. Soc. **367**, 1095 (2006) [arXiv:astro-ph/0509478].
- [143] R. Nagata and J. 'i. Yokoyama, *Band-power reconstruction of the primordial fluctuation spectrum by the maximum likelihood reconstruction method*, Phys. Rev. D **79**, 043010 (2009) [arXiv:0812.4585].
- [144] J. A. Scales and R. Snieder, *The anatomy of inverse problems*, Geophysics **65**, 1708 (2000).
- [145] C. W. Groetsch, *The theory of tikhonov regularization for fredholm equations of the first kind* (Longman Science and Technology, 1984).
- [146] J. Hadamard, *Lectures on Cauchy's problem in linear partial differential equations* (Yale University Press, 1923).
- [147] P. C. Hansen, *Rank-deficient and discrete ill-posed problems* (SIAM, 1998).
- [148] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse problems* (Springer, 2000).
- [149] M. Rakhmanov, *Rank deficiency and Tikhonov regularization in the inverse problem for gravitational-wave bursts*, Class. Quant. Grav. **23**, S673 (2006) [gr-qc/0604005].
- [150] F. S. Kitaura and T. A. Ensslin, *Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization*, Mon. Not. R. Astron. Soc. **389**, 497 (2008) [arXiv:0705.0429 [astro-ph]].
- [151] R. L. Parker, *Geophysical inverse theory* (Princeton Univ. Press, 1994).
- [152] A. Ben-Israel and T. N. E. Greville, *Generalized inverses: Theory and applications* (Springer, 2003).
- [153] G. Wang, Y. Wei and S. Qiao, *Generalized inverses: Theory and computations* (Science Press, 2004).
- [154] C. F. Van Loan, *Generalizing the singular value decomposition*, SIAM. J. Numer. Anal. **13**, 76 (1976).
- [155] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems* (Winston and Sons, 1977).
- [156] R. C. Aster, B. Borchers and C. H. Thurber, *Parameter estimation and inverse problems* (Academic Press, 2004).
- [157] W. H. Press *et al.*, *Numerical recipes in FORTRAN 77: The art of scientific computing*

- (Cambridge University Press, 1992).
- [158] C. D. Rodgers, *Characterization and error analysis of profiles retrieved from remote sounding measurements*, J. Geophys. Res. **95**, 609 (1990).
 - [159] J. Christensen-Dalsgaard, J. Schou and M. J. Thompson, *A comparison of methods for inverting helioseismic data*, Mon. Not. Roy. Astron. Soc. **242**, 353 (1990)
 - [160] G. Backus and F. Gilbert, *The resolving power of gross Earth data*, Geophys. J. Roy. Astron. Soc. **16**, 169 (1968).
 - [161] R. Sneider, *An extension of Backus-Gilbert theory to nonlinear inverse problems*, Inverse Problems **7**, 409 (1991).
 - [162] J. Ory and R. G. Pratt, *Are our parameter estimators biased? The significance of finite-difference regularization operators*, Inverse Problems **11**, 397 (1995).
 - [163] A. Tarantola, *Inverse problem theory and methods for model parameter estimation* (SIAM, 2004).
 - [164] J. O. Ramsay and B. W. Silverman, *Functional data analysis* (Springer, 1997).
 - [165] <http://camb.info>
 - [166] A. Lewis and S. Bridle, *Cosmological parameters from CMB and other data: a Monte-Carlo approach*, Phys. Rev. D **66**, 103511 (2002) [arXiv:astro-ph/0205436].
 - [167] J. Dunkley *et al.* [WMAP Collaboration], *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: likelihoods and parameters from the WMAP data*, Astrophys. J. Suppl. **180**, 306 (2009) [arXiv:0803.0586].
 - [168] L. Verde *et al.*, *First Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Parameter estimation methodology*, Astrophys. J. Suppl. **148**, 195 (2003) [arXiv:astro-ph/0302218].
 - [169] G. Hinshaw *et al.*, *Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Temperature analysis*, Astrophys. J. Suppl. **170**, 288 (2007) [arXiv:astro-ph/0603451].
 - [170] L. Page *et al.*, *Three year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Polarization analysis*, Astrophys. J. Suppl. **170**, 335 (2007) [arXiv:astro-ph/0603450].
 - [171] C. Dickinson *et al.*, *High sensitivity measurements of the CMB power spectrum with the extended Very Small Array*, Mon. Not. Roy. Astron. Soc. **353**, 732 (2004) [arXiv:astro-ph/0402498].
 - [172] C. L. Reichardt *et al.*, *High resolution CMB power spectrum from the complete ACBAR data set*, Astrophys. J. **694**, 1200 (2009) [arXiv:0801.1491].
 - [173] J. L. Sievers *et al.*, *Cosmological results from five years of 30 GHz CMB intensity measurements with the Cosmic Background Imager*, [arXiv:0901.4540].
 - [174] W. C. Jones *et al.*, *A measurement of the angular power spectrum of the CMB temperature anisotropy from the 2003 flight of Boomerang*, Astrophys. J. **647**, 823 (2006) [arXiv:astro-ph/0507494].
 - [175] F. Piacentini *et al.*, *A measurement of the polarization-temperature angular cross power spectrum of the cosmic microwave background from the 2003 flight of BOOMERANG*, Astrophys. J. **647**, 833 (2006) [arXiv:astro-ph/0507507].
 - [176] T. E. Montroy *et al.*, *A measurement of the CMB Spectrum from the 2003 Flight of BOOMERANG*, Astrophys. J. **647**, 813 (2006) [arXiv:astro-ph/0507514].
 - [177] M. L. Brown *et al.*, *Improved measurements of the temperature and polarization of the CMB from QUaD*, Astrophys. J. **705**, 978 (2009) [arXiv:0906.1003].

- [178] 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. Roy. Astron. Soc. **335**, 1193 (2002) [arXiv:astro-ph/0112114].
- [179] M. Tegmark *et al.*, *Cosmological constraints from the SDSS luminous red galaxies*, Phys. Rev. D **74**, 123507 (2006) [arXiv:astro-ph/0608632].
- [180] D. J. Eisenstein, H. j. Seo and M. J. White, *On the robustness of the acoustic scale in the low-redshift clustering of matter*, Astrophys. J. **664**, 660 (2007) [arXiv:astro-ph/0604361].
- [181] D. J. Eisenstein and W. Hu, *Baryonic features in the matter transfer function*, Astrophys. J. **496**, 605 (1998) [arXiv:astro-ph/9709112].
- [182] D. J. Eisenstein and W. Hu, *Power spectra for cold dark matter and its variants*, Astrophys. J. **511**, 5 (1997) [arXiv:astro-ph/9710252].
- [183] S. Cole *et al.*, *The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications*, Mon. Not. Roy. Astron. Soc. **362**, 505 (2005) [arXiv:astro-ph/0501174].
- [184] J. Hamann, S. Hannestad, A. Melchiorri and Y. Y. Y. Wong, *Nonlinear corrections to the cosmological matter power spectrum and scale-dependent galaxy bias: implications for parameter estimation*, JCAP **0807**, 017 (2008) [arXiv:0804.1789].
- [185] W. J. Percival and M. L. Brown, *Likelihood methods for the combined analysis of CMB temperature and polarisation power spectra*, Mon. Not. Roy. Astron. Soc. **372**, 1104 (2006) [arXiv:astro-ph/0604547].
- [186] M. E. Abroe *et al.*, *Frequentist estimation of cosmological parameters from the MAXIMA-1 cosmic microwave background anisotropy data*, Mon. Not. Roy. Astron. Soc. **334**, 11 (2002) [arXiv:astro-ph/0111010].
- [187] T. J. Hastie and R. J. Tibshirani, *Generalized additive models* (Chapman and Hall/CRC, 1990).
- [188] V. V. Morozov, *Methods for solving incorrectly posed problems* (Springer, 1984).
- [189] B. R. Hunt, *The application of constrained least squares estimations to image restoration by digital computer* IEEE Trans. Comput. **C-22**, 805 (1973)
- [190] A. M. Thompson *et al.*, *A study of methods of choosing the smoothing parameter in image restoration by regularization* IEEE Trans. Pattern Anal. Machine Intell. **13**, 326 (1991)
- [191] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: Numerical Aspects of Linear Inversion* (SIAM, 1998).
- [192] A. M. Thompson and I. J. D. Craig, *Automatic strategies for astrophysical inverse problems* Astron. Astrophys. **262**, 359 (1992).
- [193] P. C. Hansen, M. E. Kilmer and R. H. D. Kjellden, *Exploiting residual information in the parameter choice for discrete ill-posed problems* BIT Numer. Math. **46**, 41 (2006).
- [194] B. W. Rust and D. P. D. O'Leary, *Residual periodograms for choosing regularization parameters for ill-posed problems* Inverse Problems **24**, 034005 (2008).
- [195] C. L. Mallows, *Some Comments on Cp* Technometrics **15**, 661 (1973).
- [196] C. R. Vogel, *Computational Methods for Inverse Problems* (SIAM, 2002).
- [197] C. M. Stein, *Estimation of the mean of a multivariate normal distribution* Ann. Statist. **9**, 1135 (1981).
- [198] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, Proc. of the 2nd Int. Symp. on Information Theory 267 (1973)

- [199] G. Golub, M. Heath and G. Wahba, *Generalized cross validation as a method for choosing a good ridge parameter*, Technometrics **21**, 215 (1979).
- [200] S. F. Gull and G. J. Daniell, *Image reconstruction from incomplete and noisy data*, Nature **272**, 686 (1978).
- [201] R. Narayan and R. Nityananda, *Maximum entropy image restoration in astronomy*, Ann. Rev. Astron. Astrophys. **24**, 127 (1986).
- [202] K. Horne, *Images of accretion discs-I: The eclipse mapping method*, Mon. Not. Roy. Astron. Soc. **213**, 129 (1985).
- [203] C. Pichon and E. Thiebaud, *Non parametric reconstruction of distribution functions from observed galactic disks*, Mon. Not. Roy. Astron. Soc. **301**, 419 (1997) [arXiv:astro-ph/9808197].
- [204] D. N. Spergel *et al.*, *First Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters*, Astrophys. J. Suppl. **148**, 175 (2003) [arXiv:astro-ph/0302209].
- [205] M. R. Nolta *et al.*, *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Angular power spectra*, Astrophys. J. Suppl. **180**, 296 (2009) [arXiv:0803.0593].
- [206] M. Tegmark *et al.* [SDSS Collaboration], *The 3D power spectrum of galaxies from the SDSS*, Astrophys. J. **606**, 702 (2004) [arXiv:astro-ph/0310725].
- [207] A. Lewis, A. Challinor and A. Lasenby, *Efficient Computation of CMB anisotropies in closed FRW models*, Astrophys. J. **538**, 473 (2000) [arXiv:astro-ph/9911177].
- [208] <http://cosmologist.info/cosmomc/>
- [209] A. J. S. Hamilton and M. Tegmark, *Decorrelating the power spectrum of galaxies*, Mon. Not. Roy. Astron. Soc. **312**, 285 (2000) [arXiv:astro-ph/9905192].
- [210] P. Paykari and A. H. Jaffe, *Optimal binning of the primordial power spectrum*, Astrophys. J. **711**, 1 (2010) [arXiv:0902.4399].
- [211] F. Bauer and M. A. Lukas, *Comparing parameter choice methods for regularization of ill-posed problems*, Math. Comput. Simulation **81**, 1795 (2011).
- [212] U. Hämarik, R. Palm and T. Raus, *A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level*, J. Comput. Appl. Math. **236**, 2146 (2012).
- [213] L. Reichel and G. Rodriguez, *Old and new parameter choice rules for discrete ill-posed problems*, Numer. Algorithms **63**, 65 (2013).
- [214] P. C. Hansen, *Discrete inverse problems: insight and algorithms* (SIAM, 2010).
- [215] P. A. R. Ade *et al.* [Planck Collaboration], *Planck 2013 results. XV. CMB power spectra and likelihood*, [arXiv:1303.5075].
- [216] S. Das, T. Louis, M. R. Nolta, G. E. Addison, E. S. Battistelli, J. R. Bond, E. Calabrese and D. C. M. J. Devlin *et al.*, *The Atacama Cosmology Telescope: Temperature and gravitational lensing power spectrum measurements from three seasons of data*, [arXiv:1301.1037].
- [217] K. T. Story, C. L. Reichardt, Z. Hou, R. Keisler, K. A. Aird, B. A. Benson, L. E. Bleem and J. E. Carlstrom *et al.*, *A measurement of the cosmic microwave background damping tail from the 2500-square-degree SPT-SZ survey*, Astrophys. J. **779**, 86 (2010) [arXiv:1210.7231].
- [218] P. A. R. Ade *et al.* [Planck Collaboration], *Planck 2013 results. XXII. Constraints on inflation*, [arXiv:1303.5082].
- [219] P. A. R. Ade *et al.* [Planck Collaboration], *Planck 2013 results. XVI. Cosmological parameters*, [arXiv:1303.5076].