EXPECTED PRINCIPAL COMPONENT ANALYSIS OF COSMIC MICROWAVE BACKGROUND ANISOTROPIES

SAMUEL LEACH

Astrophysics Sector, SISSA–ISAS, via Beirut 4, Trieste 34014, Italy

E-mail: leach@sissa.it

We implement and test an approach for measuring the primordial power spectrum of density perturbations given observations of the cosmic microwave background anisotropy spectrum. The method depends on exploiting the fact that the linear response of the CMB anisotropy spectrum with respect to the primordial power spectrum model parameters is well understood, as well as the noise properties of the CMB detectors. This puts us in the luxurious position of being able to precompute an accurate and useful representation of a Fisher matrix, from which a set of orthonormal power spectrum modes can be obtained. The full power spectrum mode plus nuisance parameter space can be integrated out using Markov chain Monte Carlo, and all the information concerning the primordial power spectrum is compressed onto a series of mode amplitudes which can then be easily compared with theoretical models.

1 Introduction

High signal to noise, high resolution, multifrequency observations of the Cosmic Microwave Background (CMB) are providing us with a fascinating opportunity to probe many diverse sectors of astrophysics and of our cosmological model. In the near term future an ensemble of ground-based, balloon-borne and satellite observations of CMB temperature and polarization anisotropies will provide us with a window on the basic model of linear perturbations to a photon–baryon fluid coupled to dark matter potentials via gravity, on the reionization epoch, on re-scattering of CMB photons by hot cluster gas (the Sunyaev–Zel'dovich effect), and on gravitational lensing of the CMB by the intervening dark matter distribution. Each of these phenomena poses interesting challenges for data analysis, the most fundamental of which is the fact that CMB data is *correlated*, and hence a global analysis of the entire data set must be attempted in order to fully exploit the science—see the monologue by Dodelson[1] for a recent treatment of the physics and data analysis of the CMB.

Here our focus is on measuring the *primordial power spectrum* which seeds both the oscillations in the photon–baryon fluid and gravitational instability in the dark matter sector, leading to structure formation. The basic hope is that the details of the primordial power spectrum (its shape, its Gaussianity or otherwise) will shed light on whatever mechanism in the early universe is responsible for actually generating the primordial power spectrum itself. At present the dominant early universe paradigm that emerged back in the late 1970's and early 1980's, and is by now not without many observational successes, is the celebrated *inflation* model.

2 The basic problem

The desired primordial power spectrum $\mathcal{P}(k)$ is related to the observed anisotropy spectrum of the surface of last scattering C_ℓ via

$$C_\ell = \frac{2\pi}{\ell(\ell+1)} \int d\ln k\, \mathcal{P}(k)\, T_\ell^2(k; \{\omega_i\}) + N_\ell, \quad (1)$$

where the dependence of the numerically calculable CMB transfer functions $T_\ell(k)$ on a set of cosmological parameters $\{\omega_i\}$ has been written in explicitly as well a Gaussian isotropic noise term, N_ℓ. Amongst the main science goals of all recent CMB observations has been the determination of these cosmological parameters, which is made possible by assuming some reasonable form for $\mathcal{P}(k)$ such as a smooth power-law. What kind of approaches are possible if we drop these model-motivated assumptions?

There is in fact a satisfactory solution to this rather generic data analysis problem given by Hu and Okamoto[2] in which a *Fisher matrix principal component analysis* (PCA) approach can be taken. We have implemented this method[3], a reference which also contains more of the details as well as an entry point to the literature for other approaches to the same reconstruction problem that have been investigated.

The essence of the method at hand is to con-

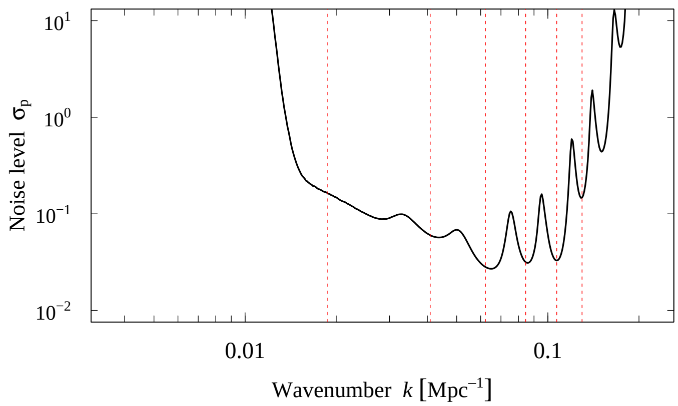

Figure 1. Illustrating the window of sensitivity to the primordial power spectrum for a *Planck*-like instrument. Here σ_p gives the approximate 1σ error on measurements of the primordial power spectrum using bandpowers with $\delta \ln k \sim 0.05$. The vertical lines indicate the position of the temperature acoustic peaks. The cosmological parameters have been fixed, so some degrading of the sensitivity is expected.

struct an *orthonormal power spectrum model*

$$\frac{\mathcal{P}(k)}{\mathcal{P}_0} = m_0 + \sum_{a=1}^{a_{\max}} m_a \mathcal{S}_a(k), \qquad (2)$$

which is designed to satisfy the expectation $\langle m_a m_b \rangle = \sigma_a^2 \delta_{ab}$. Which orthonormal basis should we choose? Clearly the variation in the power spectrum modes $\mathcal{S}_a(k)$ should reflect our expectations of where observations are at their most sensitive, and hence there is a link with the Fisher information matrix, which is often associated with forecasting the expected sensitivity of a given instrumental specification.

Before sketching the details of the method however, we state the broader working assumptions that we rely on, which may be relevant when trying to implement this method in other contexts outside the realm of CMB anisotropies:

1. The initial perturbations are pure Gaussian adiabatic modes entering in Eq. (1) via a single physical component $\mathcal{P}(k)$.

2. The transfer functions $T_\ell(k)$ can be accurately calculated and are fast to evaluate. We make use of the CMB anisotropy code CAMB[4].

3. The noise model N_ℓ is known, and hence the Fisher matrix $\mathbf{F}_{ij}$ (to be described below) for a given instrument can be calculated. Tegmark, Taylor and Heavens[5] give an exposition of the Fisher matrix formalism in the context of cosmology.

4. A method for exploring a 20–50 dimensional posterior parameter space is available. Here we make use of the Markov Chain Monte Carlo method, as implemented in the state-of-the-art CosmoMC[a] code by Lewis and Bridle[6].

5. The main science driver behind the PCA analysis, however, is the prospect of the large data set being gathered over the next five years or so.

[a]publicly available at:
http://cosmologist.info/cosmomc

Going more into the details, there is a basic pre-processing step which involves constructing the Fisher matrix

$$\mathbf{F}_{ij} = \sum_{\ell=2}^{\ell_{\max}} \frac{2\ell+1}{2} \mathrm{Tr}[\mathbf{D}_{\ell i} C_\ell^{-1} \mathbf{D}_{j\ell} C_\ell^{-1}] \,, \qquad (3)$$

where

$$\mathbf{D}_{\ell i} = \frac{\partial C_\ell}{\partial p_i}\Big|_{\rm fid} \qquad (4)$$
$$= \frac{2\pi}{\ell(\ell+1)} \int d\ln k \, \mathcal{P}_0 T_\ell^2(k) \, W_i(\ln k) \,,$$

which is evaluated for some fixed fiducial values for the cosmological parameters. We can take our power spectrum test function W_i to be the triangle window

$$W_i(\ln k) = \max\left[1 - \left|\frac{\ln k - \ln k_i}{\Delta \ln k}\right|, 0\right] \,. \qquad (5)$$

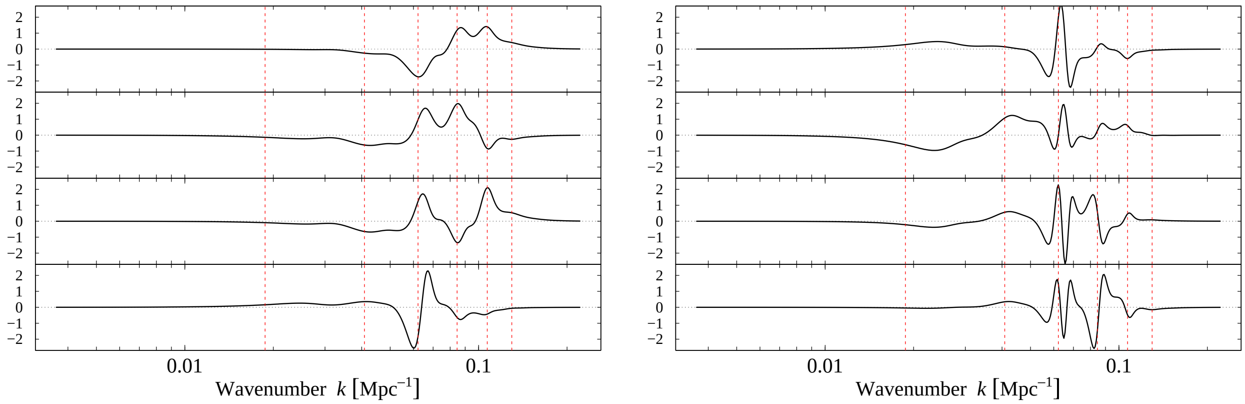

Figure 2. Illustrating PCA modes, $\mathcal{S}_a(k)$, 1–8 which have been generated assuming a *Planck*-like noise model. The vertical lines indicate the position of the temperature acoustic peaks.

The Fisher matrix Eq. (3) thus encodes the transfer of power from k-space to ℓ-space via the functions $\mathbf{D}_{\ell i}$, and the signal plus noise model via $C_\ell^{-1} = (S_\ell + N_\ell)^{-1}$. As usual the Cramer–Rao bound, given by the diagonal component of the inverse Fisher matrix, gives a useful handle on the best-case scenario for the sensitivity to the observables, and in this context reveals the possible observable range of scales, which we display in Figure 1.

The desired PCA modes $\mathcal{S}_a(k)$ are simply the (suitably normalised) orthonormal eigenvectors of the inverse Fisher matrix $\mathbf{F}_{ij}^{-1}$, and in Figure 2 we show the first eight PCA modes which resemble Fourier modes localised in the acoustic peak region $0.01 < k < 0.2\,\mathrm{Mpc}^{-1}$, displaying rapid oscillations at the acoustic peak scales. As a brief aside, the Fisher matrix pre-processing step was implemented using the R environment[7], which allows for matrix manipulations using the LAPACK linear algebra library.

The PCA mode amplitudes m_a can then be appended to the usual list of cosmological parameters to be integrated out using the now fairly standard and accessible MCMC technique. By construction the posterior distribution will be close to an uncorrelated Gaussian, which can then be used as a new likelihood function with respect to model-motivated power spectrum parametrisations. The model predictions for the PCA mode amplitudes are simply a convolution of the theoretical power spectrum over the PCA modes

$$m_a = \int d\ln k\, \mathcal{S}_a(k) \frac{\mathcal{P}}{\mathcal{P}_0}(k). \tag{6}$$

Incidentally, the likelihood evaluations over this compressed data set will be fast which opens up the possibility of performing thorough model selection studies.

3 Tests with simulated data

We have tested this method using simulated *Planck*-like data generated using various primordial power spectra including scale-invariant as well as a somewhat contrived Gaussian bump power spectra. Here we will present results assuming a scale-free input power spectrum with spectral slope $n_\mathrm{S} - 1 = -0.03$. We integrated out the parameter space consisting of five basic cosmological parameters and a further 20 PCA mode parameters, and the results are displayed in the first panel of Figure 3. The basic result here is that this kind of analysis is indeed feasible, requiring a total of around 10^6 MCMC likelihood evaluations to relax to a good representation of the posterior peak. We exploit the fact that CosmoMC can be executed across multiple CPUs with near perfect parallelisation. In addition, CosmoMC stores the cosmological parameter transfer functions during the movement through the power spectrum parameter space, meaning that movement in the "bulk" power spectrum parameter space is fast.

Finally, the measured PCA mode amplitudes can be used to constrain the power-law slope of the initial power spectrum and we recover to within one standard deviation the input power-law slope, shown in the second panel of Figure 3.

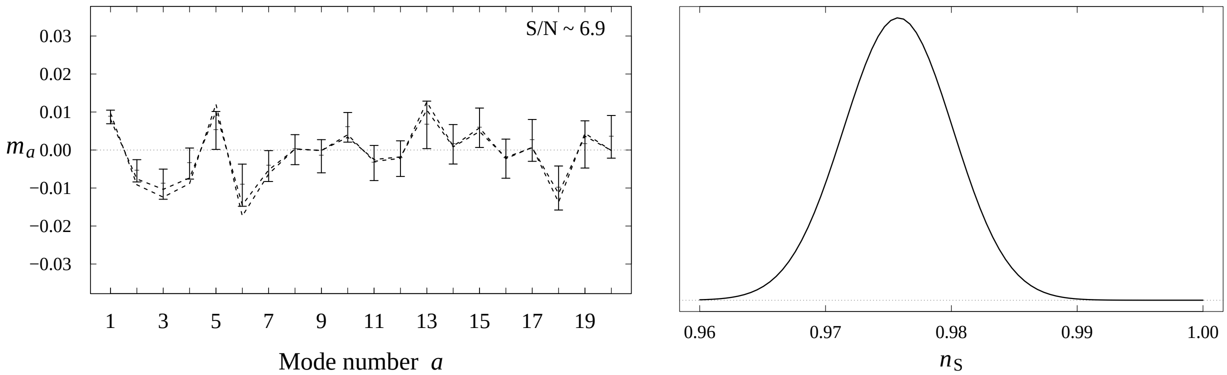

Figure 3. Illustrating the recovery of the first 20 principal component amplitudes from simulated *Planck* data with an input $n_S = 0.97$ spectrum (left panel). The models (dashed lines) correspond to power-law spectra with $n_S(k_0 = 0.05\ \mathrm{Mpc}^{-1}) = \{0.970, 0.985\}$ (bottom to top, mode 3). The PCA mode amplitudes can then be used to constrain more traditional power-law power spectra (right panel).

4 Final comments

The method has various extensions. For instance the PCA modes can be orthogonalised to the cosmological parameters in order to compensate for the fact that cosmological parameters degeneracies will break the desired statistical orthogonality of the recovered PCA mode amplitudes. In addition the PCA modes can be modified in order to search for deviations from scale-free spectra, not just deviations from scale-invariant spectra[2].

If one does require a (correlated) representation of the initial power spectrum and its covariance matrix in k-space given the measured PCA mode amplitudes, then an approach similiar to the "minimum variance map making" solution of Eq. (6) should be possible.

A detailed empirical approach such as the PCA method will always be an option for reconstructing some unknown function, particularly when the physics model underlying the data is thought to be basically well understood. The fact that it is derived from a Fisher matrix calculation also acts as a useful check on the consistency and scope of the final results.

Acknowledgements

The author would like to thank the organisers of PHYSTAT05, and Andrew Jaffe, Geoff Nicholls, and Roberto Trotta for their assistance and input during the conference.

References

1. S. Dodelson, *Modern Cosmology*, (Academic Press 2003).
2. W. Hu and T. Okamoto, *Phys. Rev.* D **69**, 043004 (2004).
3. S. Leach, preprint, `astro-ph/0506390`.
4. A. Lewis, A. Challinor and A. Lasenby, *ApJ* **538**, 473 (2002) (CAMB).
5. M. Tegmark, A. Taylor and A. Heavens, *ApJ* **480**, 20 (1997).
6. A. Lewis and S.L. Bridle, *Phys. Rev.* D **66**, 103511 (2002) (CosmoMC).
7. R Development Core Team, http://www.R-project.org, (2004).