Reconstruction of the Primordial Power Spectrum Using Multiple Data Sets

Paul Hunt

IFT UW

16th June 2009

Outline

1. Introduction
2. Methods of estimating the PPS.
3. Deconvolution approach.
4. Deconvolution as an ill-posed problem and the need for regularisation.
5. Tikhonov regularisation.
    - Theory.
    - Results.
6. Backus-Gilbert method.
    - Theory.
    - Results.
7. Conclusion.

Introduction

The WMAP results show that the primordial density perturbations are coherent, predominantly adiabatic and generated on superhorizon scales.

Why is the PPS $\mathcal{P}_{\mathcal{R}}(k)$ important?

- It can discriminate between models of inflation.
- Cosmological parameter estimation depends on the PPS (e.g. an EdeS model can fit the WMAP data if there is a 'bump' in the PPS Hunt & Sarkar 2007, 2008).

Usually the PPS is assumed to be a power-law with $\mathcal{P}_{\mathcal{R}}(k) \propto k^{n_s - 1}$.

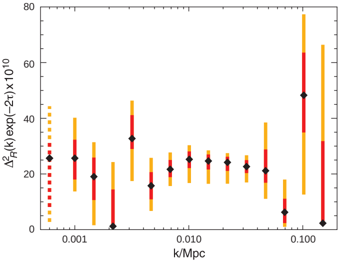
However, inflationary models involving abnormal initial conditions (e.g. Brandenberger and Martin 2001), interruptions to slow-roll evolution (Starobinsky 1992, 1998) or additional dynamical degrees of freedom (e.g. Salopek, Bond and Bardeen 1989) produce a wide variety of spectra.

Given our ignorance of the physics behind inflation a model-independent method of estimating the PPS is essential.

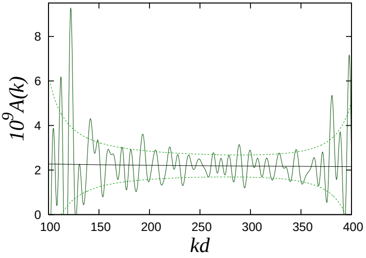
How to estimate $\mathcal{P}_{\mathcal{R}}(k)$?

Model-independent approaches fall into two classes:

- Parametric methods The number of parameters describing the PPS is much less than the number of data points Mukherjee and Wang 2005, Sealfon, Verde and Jimenez 2005, Leach 2006, Bridges, Lasenby and Hobson 2006, Verde and Peiris 2008, Bridges *et al.* 2008
- Deconvolution techniques The background cosmology is assumed to be known Kogo, Sasaki and Yokoyama 2005, Tocchini-Valentini, Douspis and Silk 2005, Tocchini-Valentini, Hoffman and Silk 2006, Shafieloo *et al.* 2007, Shafieloo and Souradeep 2007, 2009



Spergel *et al.* 2007



Nagata and Yokoyama 2008

Deconvolution as an ill-posed problem

The data points of CMB anisotropy, galaxy clustering, Lyman α forest, cluster abundance or weak lensing measurements can be written as

$$d_a = \int_0^\infty K_a(k)\,\mathcal{P}_\mathcal{R}(k)\,\mathrm{d}k + n_a. \qquad (1)$$

We wish to estimate $\mathcal{P}_\mathcal{R}(k)$ from the data, assuming $K_a(k)$ is known.

However, the convolution with $K_a(k)$ acts as a smoothing operation.

Conversely, noise in the data is amplified in the reconstructed $\mathcal{P}_\mathcal{R}(k)$.

Therefore the inverse problem of recovering the PPS has no unique stable solution and is mathematically ill-posed.

Regularisation schemes can be used to obtain approximate solutions to ill-posed problems, usually by employing prior information.

Tikhonov regularisation

Our strategy is to find the smoothest PPS consistent with the data.

Discretising the integral in eq.(1) produces

$$d_a = \sum_i W_{ai}\, s_i + n_a, \qquad s_i \equiv \mathcal{P}_{\mathcal{R}}\left(k_i\right).$$

The estimate $\hat{\mathbf{s}}$ of $\mathbf{s}$ is taken to be the vector which minimises

$$Q\left(\mathbf{s}, \mathbf{d}\right) = -2 \ln P\left(\mathbf{d}|\mathbf{s}\right) + \lambda R\left(\mathbf{s}\right).$$

The first term on the rhs enforces fidelity to the data while the second enforces smoothness.

Following Tocchini-Valentini, Douspis and Silk 2005, Tocchini-Valentini, Hoffman and Silk 2006 we choose

$$R\left(\mathbf{s}\right) = \mathbf{s}^t\, \mathsf{L}^t \mathsf{L}\, \mathbf{s} \propto \int \left(\frac{\mathrm{d}\mathcal{P}_{\mathcal{R}}}{\mathrm{d}\ln k}\right)^2 \frac{\mathrm{d}k}{k},$$

where L is a discrete approximation to the 1st-order derivative operator.

Bayesian analysis

Tikhonov regularisation has a natural Bayesian interpretation.

Bayes' theorem states

$$P(\mathbf{s}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{s})\, P(\mathbf{s}) \,.$$

Posterior distribution — Likelihood function — Prior distribution

If $P(\mathbf{s}) \propto \exp[-\lambda R(\mathbf{s})/2]$ then minimising $Q(\mathbf{s}, \mathbf{d})$ is equivalent to maximising $P(\mathbf{s}|\mathbf{d})$, and $\hat{\mathbf{s}}$ corresponds to the mode of $P(\mathbf{s}|\mathbf{d})$.

For most data sets $P(\mathbf{d}|\mathbf{s})$ is approximately Gaussian $\Rightarrow \langle \mathbf{s} \rangle = \hat{\mathbf{s}}$ and the scatter about $\langle \mathbf{s} \rangle$ is described by the covariance matrix

$$\Sigma_B \equiv \langle (\mathbf{s} - \langle \mathbf{s} \rangle)(\mathbf{s} - \langle \mathbf{s} \rangle)^t \rangle = \mathrm{H}^{-1}(\hat{\mathbf{s}}) \,, \qquad H_{ij} \equiv \frac{1}{2} \frac{\partial^2 Q}{\partial s_i \partial s_j} .$$

Frequentist analysis

Suppose that $\hat{\mathbf{s}} = \mathsf{M}\mathbf{d}$.

Using $\mathbf{d} = \mathsf{W}\mathbf{s}_T + \mathbf{n}$ where $\mathbf{s}_T$ is the true PPS this gives

$$\hat{\mathbf{s}} = \mathbf{s}_T + (\mathsf{R} - \mathsf{I})\,\mathbf{s}_T + \mathsf{M}\mathbf{n},$$

where $\mathsf{R} \equiv \mathsf{M}\mathsf{W}$ is known as the resolution matrix.

In the frequentist approach we imagine an ensemble of observers, each measuring the data and estimating the PPS in the same way.

Then $\hat{\mathbf{s}}$ has a distribution $P\left(\hat{\mathbf{s}}|\mathbf{s}_T\right)$ with mean $\langle\hat{\mathbf{s}}\rangle = \mathsf{R}\mathbf{s}_T$ and covariance matrix

$$\Sigma_F \equiv \left\langle \left(\hat{\mathbf{s}} - \langle\hat{\mathbf{s}}\rangle\right)\left(\hat{\mathbf{s}} - \langle\hat{\mathbf{s}}\rangle\right)^t \right\rangle = \mathsf{M}\mathsf{N}\mathsf{M}^t.$$

The bias of $\hat{\mathbf{s}}$ is

$$\mathrm{Bias}\left(\hat{\mathbf{s}}\right) \equiv \langle\hat{\mathbf{s}} - \mathbf{s}_T\rangle = (\mathsf{R} - \mathsf{I})\,\mathbf{s}_T.$$

Regularisation reduces the variance but introduces a bias.

Multiple data sets

Most deconvolution attempts have used the WMAP TT data alone.

By using multiple data sets we can recover $\mathcal{P}_{\mathcal{R}}(k)$ over a larger k range with increased accuracy.

For example, if

$$L(\mathbf{s}) = \sum_I (\mathsf{W}_I \mathbf{s} - \mathbf{d}_I)^t \, \mathsf{N}_I^{-1} (\mathsf{W}_I \mathbf{s} - \mathbf{d}_I),$$

then the estimate of the PPS is

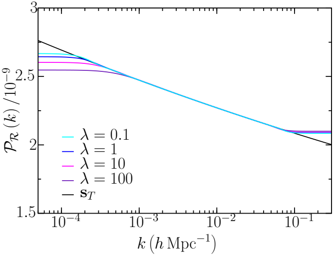
$$\hat{\mathbf{s}} = \Sigma_B \sum_I \mathsf{W}_I^t \mathsf{N}_I^{-1} \mathbf{d}_I,$$

and the covariance matrices are

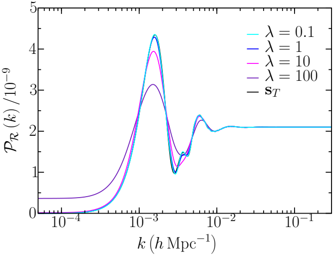
$$\Sigma_B^{-1} = \lambda \mathsf{L}_n^t \mathsf{L}_n + \sum_I \mathsf{W}_I^t \mathsf{N}_I^{-1} \mathsf{W}_I, \qquad \Sigma_F = \Sigma_B \left(\sum_I \mathsf{W}_I^t \mathsf{N}_I^{-1} \mathsf{W}_I \right) \Sigma_B^t.$$

Deconvolution without noise

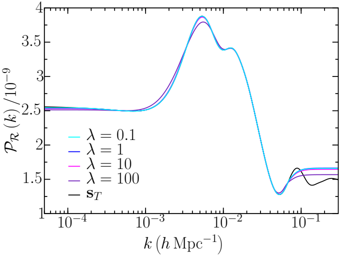
Dunkley *et al.* 2009



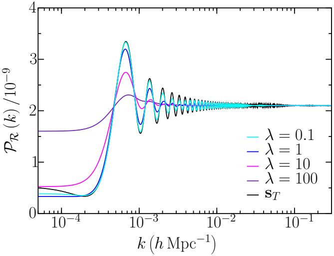
Langlois and Vernizzi 2005



Hunt and Sarkar 2007

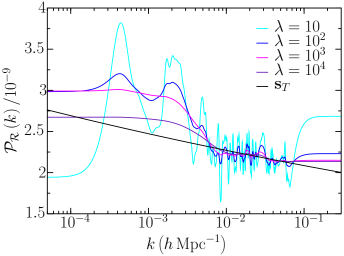


Starobinsky 1992

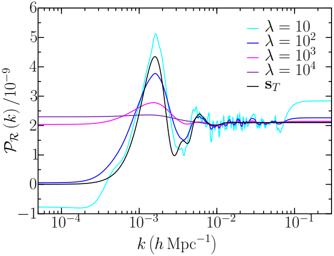


Deconvolution with noise

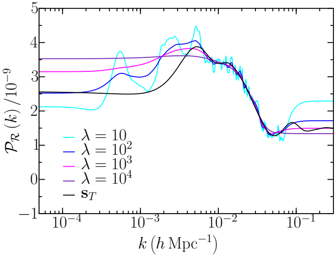
Dunkley *et al.* 2009



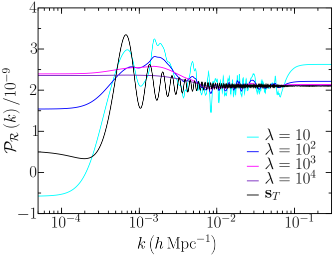
Langlois and Vernizzi 2005



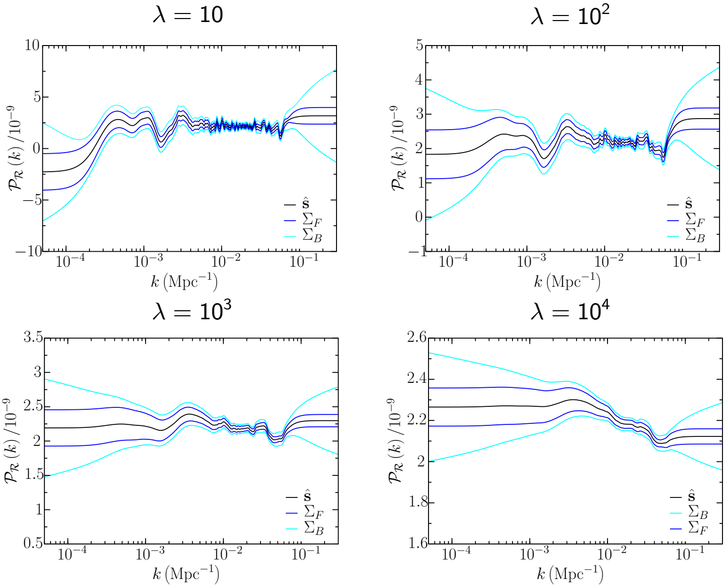
Hunt and Sarkar 2007



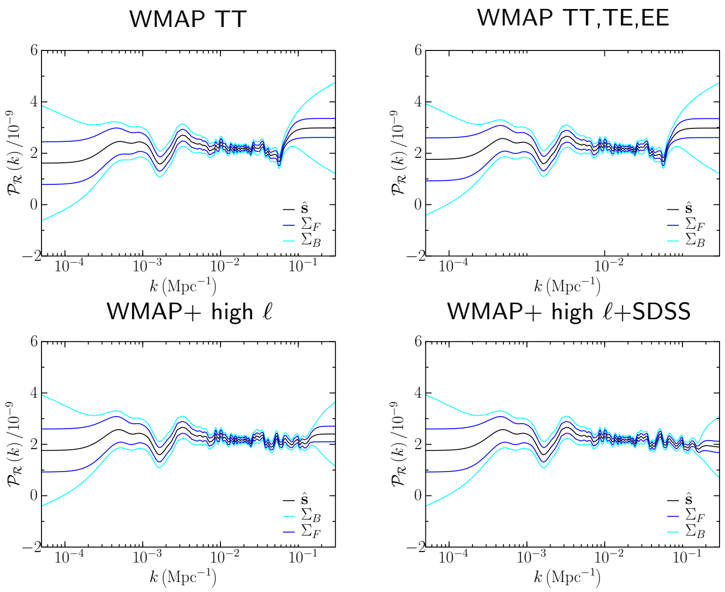
Starobinsky 1992



Varying λ, WMAP TT data

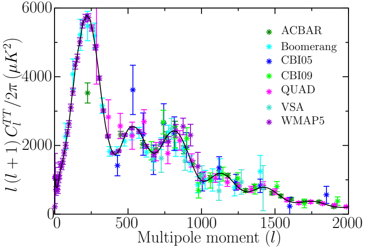


Adding data

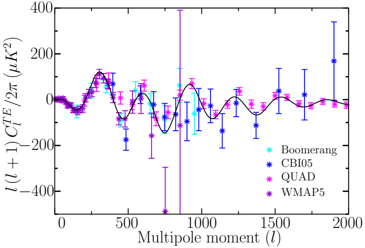

WMAP TT
WMAP TT,TE,EE
WMAP+ high ℓ
WMAP+ high ℓ+SDSS
$\mathcal{P}_{\mathcal{R}}(k)/10^{-9}$
$k\,(\mathrm{Mpc}^{-1})$
10^{-4} 10^{-3} 10^{-2} 10^{-1}
−2 0 2 4 6
$\hat{\mathbf{s}}$
Σ_F
Σ_B

Comparison with data

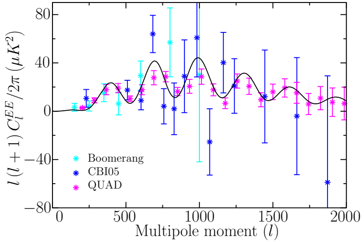
The fit to CMB TT data



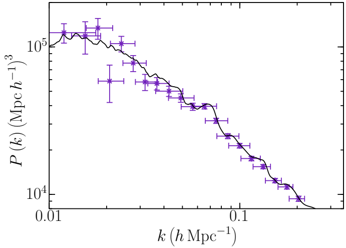
The fit to CMB TE data



The fit to CMB EE data



The fit to SDSS LRG data



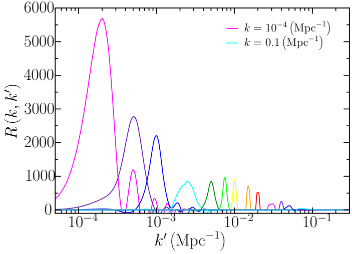
Resolution

We introduce resolution kernels $R(k,k')$ as the continuous analogues of the resolution matrix. These satisfy

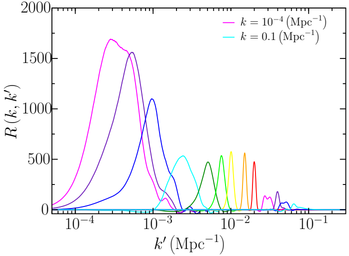
$$\langle \hat{\mathcal{P}}_{\mathcal{R}}(k) \rangle = \int_0^\infty R(k,k')\,\mathcal{P}_{\mathcal{R}}(k')\,\mathrm{d}k', \qquad \int_0^\infty R(k,k')\,\mathrm{d}k' = 1.$$

Clearly the closer $R(k,k')$ is to the Dirac delta function $\delta(k'-k)$ the better the resolution of the recovered PPS.

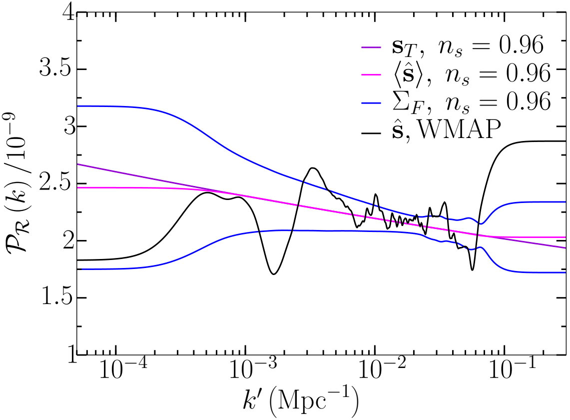
$\lambda = 10$



$\lambda = 10^2$



Is the PPS different from a power-law?



We find a cut-off at $5 \times 10^{-5} < k < 5 \times 10^{-4}$ Mpc at 0.7σ cl, a dip at $0.001 < k < 0.0025$ Mpc at 1.7σ cl and a peak at $0.0025 < k < 0.006$ at 1.4σ cl.

Backus-Gilbert method

We seek coefficients $q_a(k)$ such that

$$\hat{\mathcal{P}}_{\mathcal{R}}(k) = \sum_a q_a(k)\, d_a.$$

Combining this with eq.(1) gives

$$\hat{\mathcal{P}}_{\mathcal{R}}(k) = \int_0^\infty \sum_a q_a(k)\, K_a(k')\, \mathcal{P}_{\mathcal{R}}(k')\, \mathrm{d}k'.$$

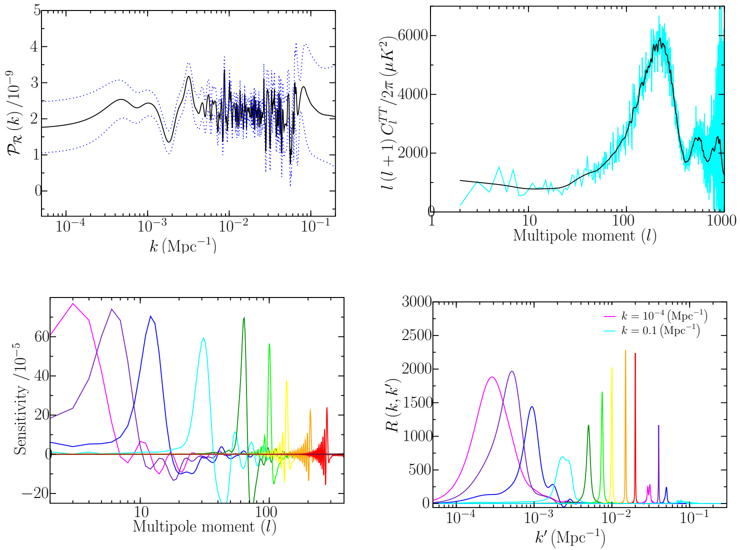
Thus we want $R(k,k') = \sum_a q_a(k)\, K_a(k')$ to approximate a delta function and so we impose $\int R(k,k')\, \mathrm{d}k' = 1$.

We also want to minimise the variance so we minimise

$$Q = \int_0^\infty (k'-k)^2 R^2(k,k')\, \mathrm{d}k' + \lambda \sum_{ab} q_a(k)\, N_{ab}\, q_b(k).$$

'Width' of $R(k,k')$ — Variance

Backus-Gilbert results



Conclusions

In the future as data sets improve we will be able to treat the PPS as a *function* instead of a set of parameters.

It is necessary to combine data sets to obtain the best results.

We have shown that Tikhonov regularisation and the Backus-Gilbert method produce high resolution reconstructions of the PPS from multiple noisy data sets.

Both methods give consistent results and have well-defined error estimates.

The recovered PPS show interesting features on large scales.

It should be possible to extend our method to isocurvature perturbations and nonlinear data sets.

Our ultimate goal is to determine *empirically* the cosmological parameters and the PPS.