

**Maximum entropy deconvolution of primordial power spectrum**

Gaurav Goswami\* and Jayanti Prasad†

*IUCAA, Post Bag 4, Ganeshkhind, Pune 411007, India*

(Received 24 April 2013; published 18 July 2013)

It is well known that cosmic microwave background (CMB) temperature anisotropies and polarization can be used to probe the metric perturbations in the early universe. Presently, there exists no observational detection of tensor modes of primordial metric perturbations or of primordial non-Gaussianity. In such a scenario, the primordial power spectrum of scalar metric perturbations is the only correlation function of metric perturbations (presumably generated during inflation) whose effects can be directly probed through various observations. To explore the possibility of any deviations from the simplest picture of the era of cosmic inflation in the early universe, it thus becomes extremely important to uncover the amplitude and shape of this correlation sufficiently well. In the present work, we attempt to reconstruct the primordial power spectrum of scalar metric perturbations using the binned (uncorrelated) CMB temperature anisotropies data from WMAP 9-Year, using the maximum entropy method or MEM to solve the corresponding inverse problem. Our analysis shows that, given the current CMB data, there are no convincing reasons to believe that the primordial power spectrum of scalar metric perturbations has any significant features.

DOI: [10.1103/PhysRevD.88.023522](https://doi.org/10.1103/PhysRevD.88.023522)

PACS numbers: 98.80.Es, 98.70.Vc

**I. INTRODUCTION**

Observations of cosmic microwave background (CMB) temperature anisotropies as well as polarization [1,2] can be used to uncover the physics of the early universe, e.g., of cosmic inflation [3]. However, calculations of power spectra of CMB anisotropies and polarization [4,5] involve making a number of assumptions, e.g., about the reionization history of the universe, the equation of state of dark energy etc. It is also usually assumed that the primordial power spectrum of scalar metric perturbations (denoted by sPPS in this work) is a power law (with a small running). One can then use the CMB observational data to put constraints on the values of various cosmological parameters [6] including the ones specifying sPPS (usually denoted by  $A_S$ ,  $n_S$  etc). Since this procedure leads to “reasonable” values of these parameters, it is often said that a power law sPPS is consistent with the observed data. But it is worth noticing that this is just an assumption.

Cosmic inflation is the most actively investigated paradigm for explaining the origin of anisotropies in CMB sky as well as the large scale structure of the universe. The simplest versions [3,7,8] of inflationary models give a smooth, nearly scale-invariant (tilted red) sPPS. But there are other models which are capable of giving more complicated forms of sPPS (abnormal initial conditions, multifield models, interruptions to slow roll evolution, phase transition during inflation, see e.g., [9–14]). Are these models ruled out by the present data? Even though power law sPPS is consistent with the data, the assumption of a power law PPS (with small running) is just that: a well-motivated assumption. It is worth checking how the models

in which sPPS is not just a simple power law with a small running fare against the present available data.

This can be done in various ways, e.g., one could try to redo cosmological parameter estimation with the actual form of sPPS left free (see, e.g., [15]). Another option is to work with inflationary models which lead to features in sPPS and redoing parameter estimation for those models (see, e.g., [9,16]). This exercise illustrates that (i) models in which sPPS is not this simple also do fit the data and (ii) very often, with these models, one can get a better fit to the data than with a power law with small running.

Given this situation, a reasonable possibility is to try to directly deconvolve sPPS from observed CMB anisotropies (i.e.,  $C_\ell$ s). Previous attempts [17], (see also [18]), at doing so seem to suggest the existence of features in sPPS (the statistical significance of which is still being assessed [19]), e.g., a sharp infrared cutoff on the horizon scale, a bump (i.e., a localized excess just above the cutoff) and a ringing (i.e., a damped oscillatory feature after the infrared break). This is consistent with many existing models of inflation and this has also motivated theorists to build models of inflation that can give large and peculiar features in primordial power spectrum (see [9–14]).

Given the fact that primordial power spectrum of scalar metric perturbations is the only cosmological correlation whose effect is, at this stage, observable in the universe (primordial non-Gaussianity is yet to be detected in CMB data, so are B-modes of polarization of CMB due to inflationary gravitational waves), it becomes important to settle this issue of possible existence of features.

The sPPS cannot be modeled as any arbitrary function since the noise at small scale starts dominating the recovered power spectrum [20]. Only a small subset of functions have been used for modeling sPPS and featureless sPPS with (small) running are the simplest ones which add only

\*gaurav@iucaa.ernet.in  
†jayanti@iucaa.ernet.in

one extra parameter which can be estimated using, for instance, Markov chain Monte Carlo (MCMC) analysis. In order to reconstruct sPPS, the Planck team considered a binned power spectrum with external constraints and maximized a penalized likelihood which is essentially a form of regularization [20]. As has been argued in [15], if we let the power in  $k$ -space bins free, then the number of parameters becomes too large to be handled with MCMC. The Planck team has used the Newton-Raphson method to estimate sPPS in bins (keeping the other parameters fixed to their best fit values). We also consider a binned sPPS for our reconstruction and use external constraints in the form of maximum entropy regularization. However, our regularization function (entropy) is nonlinear, therefore we use a much more complex scheme for optimization (which is inspired by the work of Skilling and Bryan [21]) rather than the Newton-Raphson method.

Thus, in the present work, we try a new method of probing the shape of primordial power spectrum, the maximum entropy method, or MEM [22–24]. We begin in Sec. II by broadly describing the problem and its various attempted solutions. Then, in Sec. III, we describe in detail the algorithm that we have used. This is followed by Sec. IV in which we apply the algorithm to binned CMB temperature anisotropies data. We conclude in Sec. V with a discussion of salient features, limitations and future prospects for the work. In Appendix A, we present the results of applying the method on a toy problem and in the process illustrate the use of the algorithm.

## II. THE DECONVOLUTION PROBLEM

### A. Formulation as an inverse problem

The scalar primordial power spectrum (sPPS) is the (dimensionless) power spectrum of comoving curvature perturbation,  $\Delta_\zeta^2(k)$  and is defined by

$$\lim_{\eta \rightarrow 0} \langle \zeta(\eta, \mathbf{k}) \cdot \zeta(\eta, \mathbf{k}') \rangle = \frac{2\pi^2}{k^3} \delta^3(\mathbf{k} + \mathbf{k}') \Delta_\zeta^2(k), \quad (1)$$

where  $\eta$  is the conformal time (and at early times, all the modes of interest have wavelengths much bigger than the Hubble radius, so that all the correlations of comoving curvature perturbation,  $\zeta(t, \mathbf{x})$  freeze). For a power law sPPS,

$$\Delta_\zeta^2(k) = A_S \left( \frac{k}{k_0} \right)^{n_S-1}, \quad (2)$$

where  $A_S$  is the scalar spectral amplitude and  $n_S$  is the scalar spectral index and  $k_0$  is a pivot scale. We address the issue of reconstructing the shape of the sPPS by attempting to directly solve the (noisy) integral equations giving the CMB angular power spectrum using MEM. The observed CMB  $TT$  angular power spectrum is given by (see, e.g., [25]),

$$C_\ell^{TT}{}_{\text{obs}} = \int_0^\infty dk \left[ \frac{4\pi}{k} (\Delta_{T\ell}(k, \eta_0))^2 \right] \Delta_\Phi^2(k) + C_\ell^{TT}{}_{\text{noise}}. \quad (3)$$

Here,  $\ell$  is the multipole moment,  $k$  is the wave number, and the quantity in the square brackets is the radiation transfer function ( $\eta_0$  denotes the value of conformal time today) and  $\Delta_\Phi^2(k)$  is the dimensionless power spectrum of the (only) scalar metric perturbation in Newtonian gauge (often called Bardeen potential,  $\Phi$ ). Assuming a given set of values of background cosmological parameters, the radiative transport kernel can be found (see Sec. IV), we can then formulate the problem we are dealing with as the solution of a set of integral equations, i.e., as an inverse problem. In matter dominated universe (at the epoch of recombination), at linear order in perturbation theory,  $\Phi = (3/5)\zeta$ , so, for a power law PPS, theoretical  $C_\ell^{TT}$  (in  $\mu K^2$ ) shall be

$$C_\ell = A_S \left( \frac{3T_{\text{CMB}}}{5} \right)^2 \int_0^\infty \frac{4\pi dk}{k} (\Delta_{T\ell}(k, \eta_0))^2 \left( \frac{k}{k_0} \right)^{n_S-1}. \quad (4)$$

We shall now replace  $A_S \left( \frac{3T_{\text{CMB}}}{5} \right)^2 \left( \frac{k}{k_0} \right)^{n_S-1}$  by a general function  $f(k)$  and shall try to find it. We thus have

$$C_\ell = \int_0^\infty \frac{4\pi dk}{k} (\Delta_{T\ell}(k, \eta_0))^2 f(k), \quad (5)$$

with

$$f(k) = \left( \frac{3T_{\text{CMB}}}{5} \right)^2 \Delta_\zeta^2(k). \quad (6)$$

Given the temperature radiation transfer function ( $\Delta_{T\ell}(k, \eta_0)$ ), the theoretical  $C_\ell^{TT}$  can be found from Eq. (3), provided we know the sPPS. The  $\ell$  range for which we wish to evaluate the transfer function and the corresponding  $C_\ell$ s goes from  $\ell = 2$  to  $\ell = l_{\text{max}} = 1500$ . The typical behavior of the function

$$G(\ell, k) = dk \frac{4\pi}{k} (\Delta_{T\ell}(k, \eta_0))^2 \quad (7)$$

is shown in the Fig. 1 (with  $dk$  chosen such that the integral in the definition of  $C_\ell$  can be evaluated to a high enough accuracy).

For every given  $\ell$ , the radiation transport kernel is a highly oscillatory function of the wave number  $k$ . But for any  $\ell$ , it has significant (i.e., non-negligible) values only within a small range of  $k$  values. The brightness fluctuations roughly go as  $j_\ell[k(\eta_0 - \eta_*)]$  (where  $j_\ell$  is spherical Bessel function while  $\eta_*$  is the conformal time at the epoch of recombination), thus the minimum value of  $\ell$  sets a minimum value of  $k$  at which the kernel takes up non-negligible values. This procedure tells us that since the radiation transfer function is negligible for  $k < k_{\text{min}}$ , no matter how much power is there in sPPS at very small  $k$  values, the CMB anisotropies cannot be used to probe the

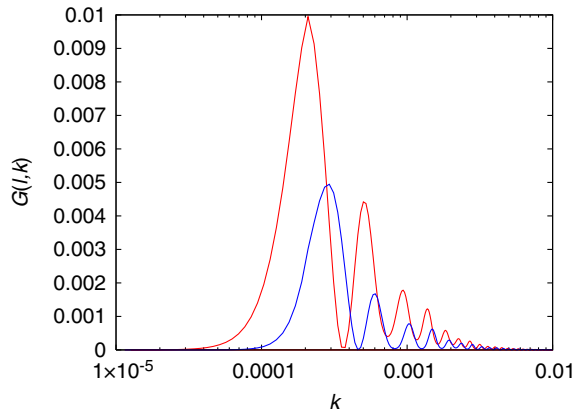


FIG. 1 (color online). The radiation transfer function  $G(l, k)$ , as defined by Eq. (7), is an oscillatory function of wave number  $k$  for all multipoles  $\ell$ . In this figure, the behavior of the kernel  $G(l, k)$  is shown for  $\ell = 2$  (bigger peaks) and  $\ell = 3$  (smaller peaks).

sPPS at these (very large) scales. This sets the  $k_{\min}$  below which we cannot probe the sPPS. Similarly, given the fact that we have observations only till a maximum value of  $\ell$ , this sets the maximum value of  $k$  up to which we need to sample the kernel; thus, the smallest possible angular resolution of a CMB experiment shall set the  $l_{\max}$  that we can probe which shall set a  $k_{\max}$ , i.e., sPPS at scales smaller than this scale can not be probed by CMB experiments. Thus,  $\ell = 2$  determines  $k_{\min}$  while  $\ell = l_{\max}$  determines  $k_{\max}$ . Within this range, one discretizes the  $k$  space in such a way that the transfer function can be sampled sufficiently well and the above integral can be performed to the desired accuracy.<sup>1</sup>

Apart from this consideration, the actual observed  $C_\ell$ s are also noisy (due to cosmic variance, instrumental noise and the effect of masking the sky). Thus Eq. (3) can be written as a set of linear equations

$$C_\ell = \sum_{k=1}^{n_s} G_{\ell k} f_k + C_\ell^N, \quad (8)$$

where  $n_s$  is the number of bins in  $k$  space and  $C_\ell^N$  is the noise term. Thus, the problem we wish to solve is: given the matrix  $G$ , the few observations ( $C_\ell$ s), the moments of the random variables  $C_\ell^N$ , how can we find the set of numbers  $f_k$ ? In this paper, we shall use the binned CMB data to find sPPS. The number of (binned and hence uncorrelated) data points (WMAP) is 45 (call it  $n_d$ ). To sample the kernel satisfactorily, we divide the  $k$  space into 6200 points ( $n_s$ ). Thus, we have a problem with a set of 45 noisy linear equations and 6200 unknowns to be determined.

<sup>1</sup>As we shall see in Sec. IV, this number is 6200, thus our  $G_{lk}$  matrix shall have dimensions  $1499 \times 6199$  ( $= 9292301$  entries).

## B. Bayesian inversion

Recovering the primordial power spectrum  $f_k$  from the observed  $C_l$  can be cast as a Bayesian inversion problem in the following way. The posterior probability  $P(f_k|C_l, G_{lk})$  of obtaining the primordial power spectrum  $f_k$ , given a kernel  $G_{lk}$  and observed  $C_l$  is given by:

$$P(f_k|C_l, G_{lk}) = \frac{P(C_l|f_k, G_{lk})P(f_k)}{P(C_l)}, \quad (9)$$

where  $P(C_l|f_k, G_{lk})$  is the likelihood and  $P(f_k)$  is the prior probability. For our case the denominator (evidence) works as just a normalization and we can ignore it.

For the case of Gaussian noise,<sup>2</sup> the likelihood function can be written as

$$P(C_l|f_k, G_{lk}) \propto \exp[-\chi^2/2], \quad (10)$$

where

$$\begin{aligned} \chi^2 &= (C_l - G_{lk}f_k)^T N^{-1} (C_l - G_{lk}f_k) \\ &= \sum_{l=2}^{l_{\max}} \frac{|C_l - G_{lk}f_k|^2}{\sigma_l^2}, \end{aligned} \quad (11)$$

for the case when the noise covariance matrix ( $N$ ) is diagonal.

Since for our problem, the number of unknowns, i.e.,  $f_k$ , is far more than the number of knowns, i.e.,  $C_l$ s; therefore, ordinary chi square minimization is of no use since it can make the chi square too low.<sup>3</sup> In order to avoid chi square taking unphysical values, we need some kind of regularization in the form of a prior. Instead of maximizing the likelihood function, we maximize the posterior probability.

It has been a common practice to consider the following form of prior for any regularization problem,

$$P(f_k) \rightarrow P(f_k, \lambda, S) = \exp[-\lambda S(f_k)/2], \quad (12)$$

where  $\lambda$  is the regularization parameter and  $S$  is the regularization function. Many forms of regularization functions (like quadratic form) have been explored in the literature. In the present work, we use an entropy function  $S(f_k)$  as a regularization function which is defined in the following way,

$$S(f_k) = - \sum_k f_k \left[ \ln \left( \frac{f_k}{A} \right) - 1 \right], \quad (13)$$

where  $A$  is a parameter that parametrizes the entropy functional we use. With the regularization function, the posterior probability distribution can be written as

<sup>2</sup>Even though the noise on  $C_\ell$ s is not Gaussian, we proceed pretending the noise to be Gaussian. This is justified because by the central limit theorem: since the chi-squared distribution is the sum of  $n_d$  independent random variables with finite mean and variance, it converges to a normal distribution for large  $n_d$ .

<sup>3</sup>even if we knew all the  $n_s$  parameters, the presence of noise shall ensure that  $\chi^2$  shall be a sum of  $n_d$  normalized Gaussians.

$$\begin{aligned}
P(f_k|G_{lk}, C_l) &= \exp\left[-\frac{\chi^2}{2}\right] * \exp\left[-\frac{\lambda S}{2}\right] \\
&= \exp\left[-\frac{1}{2}(\chi^2 + \lambda S)\right] = \exp[-M(f_k)],
\end{aligned}
\tag{14}$$

where

$$M(f_k) = \frac{1}{2}(\chi^2 + \lambda S(f_k, A)). \tag{15}$$

If we wish to minimize  $\chi^2$  and maximize  $S$ , then  $\lambda$  should be a negative number. To solve any constrained optimization problem, though it is common to use the formalism of Lagrange multipliers, it is not necessary. If one does use the formalism of Lagrange multiplier, the value of regularization parameter is also somehow to be determined and can not be arbitrarily set (refer to [26] to see how this is done for a quadratic form of regularization function, which leads to a linear problem). However, the algorithm that we shall use, is s.t. it does not cast the problem in terms of Lagrange multiplier (see Sec. 3.6 of [21]), and hence we directly solve the constrained optimization problem without using the regularization parameter explicitly.

Maximum entropy method is a particular (nonlinear) inversion method (for details see [22–24]). Here, the regularization function  $S(f_k, A)$  is nonquadratic so that the equations to be dealt with, to solve the optimization problem, shall turn out to be nonlinear. Without such a maximum entropy constraint, the inversion problem is ill posed (since the data can be satisfied by an infinity of primordial power spectra). The condition that the entropy be a maximum selects one among these (see Appendix A detail). There exist, in the literature, various arguments justifying the use of MEM over other ways of inversion (often using arguments from information theory<sup>4</sup>). At this stage, we treat it as just another nonlinear version of the general regularization scheme.

Using prior in the form of a regularization function makes it possible to incorporate features which we expect from our reconstructed sPPS. For example, Planck team has used a regularization function which penalizes any sPPS which have curvature and power in the modes which are not measured, i.e.,  $k < k_{\min}$  and  $k > k_{\max}$  [20]. In our case, we use entropy as our regularization function which also prefer a sPPS without features (having maximum entropy) as compared to a sPPS with features.

Our method differs from the one presented in [18] because they use a gradient based (smoothing) quadratic

<sup>4</sup>It is often argued that while the maximum likelihood method selects the spectrum that has the largest probability of reproducing the data, the maximum entropy method, instead, selects the positive spectrum to which is associated the largest number of ways of reproducing the data, i.e., the one that maximizes the information-theory definition of the entropy of the spectrum subject to the given constraints.

regularization function which leads to a linear problem. Also, we assumed the errors on  $C_\ell$ s to be Gaussian, the authors of [18] correct for the non-Gaussian nature of errors on  $C_\ell$ s at low  $\ell$  values.

### III. THE CAMBRIDGE MAXIMUM ENTROPY ALGORITHM

So, the problem that we wish to solve involves a highly under-determined system of linear equations. As was mentioned in the last section, one way in which we can attempt to solve this problem is to formulate it as a problem involving the optimization of a nonquadratic function (which will require solving a set of nonlinear equations) subject to a constraint. Since the number of unknowns is so large, we have to solve the corresponding constrained nonlinear optimization problem in a very large dimensional space. Also, we have other constraints that we need to take care of, e.g., the components of  $f$  are positive quantities (since  $f$  is a power spectrum), so the optimization algorithm that we use must not cause the components of  $f$  to become negative (this requirement rules out methods such as the steepest ascent). Similarly, since the objective function is quite different from a pure quadratic form, methods such as conjugate gradient method are not very useful.

Experience has shown that one of the strategies which work (despite being complicated) is the following: instead of searching for a minimum in a single search direction (e.g., in steepest ascent method), one searches in a small- (typically three-)dimensional subspace. This subspace is spanned by vectors that are calculated at each point in such a way as to avoid directions leading to negative values. The algorithm that we use is based on the one developed by Skilling and Bryan [21,27] and is sometimes referred to as The Cambridge maximum entropy algorithm. It has been extensively used in not only radio astronomy but also in other fields. Here we quickly review this algorithm for the sake of completeness.

#### A. Entropy and $\chi^2$

The problem to be solved involves finding a set of  $f_k$  ( $k = 1, 2, \dots, n_s$ ) (with maximum entropy) from a data set  $D_\ell$  ( $\ell = 1, 2, \dots, n_d$ ). For any  $f_k$ , let

$$F_\ell = \sum_k G_{\ell k} f_k. \tag{16}$$

We shall use the following definition of entropy (the nonlinear regularization function),

$$S(f_k) = -\sum_k f_k \left[ \ln\left(\frac{f_k}{A}\right) - 1 \right] = -\sum_k f_k \left[ \ln\left(\frac{f_k}{eA}\right) \right], \tag{17}$$

where  $A$  is a fixed number (sometimes called “the default”) that sets the normalization of  $f$ . Notice that

$S(\vec{0}) = 0$ ,  $S(f_k = A) = n_s \cdot A$ ,  $S(f_k = eA) = 0$ . This gives, (since  $A$  is fixed),

$$\partial S / \partial f_j = \log(A/f_j) \quad \partial^2 S / \partial f_i \partial f_j = -\delta_{ij}/f_j, \quad (18)$$

telling us that  $\partial_i S(\vec{0}) = \infty$ ,  $\partial_i S(f_k = A) = 0$  and  $\partial_i S(f_k = eA) = -1$ . It is easy to see that entropy surfaces are strictly convex. Also, the expression for the various derivatives of the entropy tell us that the solution  $f_i = A$  is the global maximum of entropy, this fact shall be important later. The measure of misfit that we shall use (in order to use the data) is the Chi-squared function

$$C(f) = \chi^2 = \sum_{\ell} (F_{\ell} - D_{\ell})^2 / \sigma_{\ell}^2, \quad (19)$$

from which we get the gradient of  $C$ ,

$$\partial C / \partial f_j = \sum_{\ell} G_{\ell j} 2(F_{\ell} - D_{\ell}) / \sigma_{\ell}^2, \quad (20)$$

and the Hessian,

$$\partial^2 C / \partial f_i \partial f_j = \sum_{\ell} G_{\ell j} \left( \frac{2}{\sigma_{\ell}^2} \right) G_{\ell i}. \quad (21)$$

For a linear experiment, the surfaces of constant chi-squared are convex ellipsoids in  $N$ -dimensional space. The largest acceptable value for  $\chi^2$  at 99% C.L. is about  $C_{\text{aim}} = n_d + 3.29\sqrt{n_d}$  (with  $n_d$  being the number of observations), see [21]. As the above equations show, quantities such as the gradient of  $C$  and Hessian of  $C$  can be easily evaluated (though finding the Hessian of  $C$  is the one of the most computationally expensive tasks since the matrix  $G_{\ell k}$  is  $45 \times 6200$  and the Hessian of  $C$  shall be  $6200 \times 6200$  matrix).

At every iteration, instead of searching for the maximum of  $S$  and minimum of  $C$  along a line, we search in an  $n$ -dimensional subspace of the parameter space. So, instead of

$$f_{(\text{new})}^i = f^i + x e^i, \quad (i = 1, 2, \dots, n_s), \quad (22)$$

we shall have (with  $e_{\mu}$  being  $n$  search directions)

$$f_{(\text{new})}^i = f^i + \sum_{\mu=1}^n x^{\mu} e_{\mu}^i. \quad (23)$$

Sufficiently near any point, every function can be approximated by a quadratic function (provided the higher-order terms in the Taylor expansion can be ignored). So, within the subspace we shall model the entropy and chi-squared by

$$S\left(f + \sum x e\right) \cong s(x), \quad (24)$$

$$C\left(f + \sum x e\right) \cong c(x), \quad (25)$$

where  $s(x)$  and  $c(x)$  are quadratic,

$$s(x) = s(0) + \sum_{\mu} s_{\mu} x^{\mu} - \sum_{\mu\nu} g_{\mu\nu} x^{\mu} x^{\nu} / 2, \quad (26)$$

$$c(x) = c(0) + \sum_{\mu} c_{\mu} x^{\mu} + \sum_{\mu\nu} h_{\mu\nu} x^{\mu} x^{\nu} / 2, \quad (27)$$

which correspond to the first three terms in the Taylor series expansion of  $S(f)$  and  $C(f)$ . The first-order term in the Taylor expansion of  $S$  is

$$\begin{aligned} \sum_{\mu} s_{\mu} x^{\mu} &= \sum_{i=1}^N \left( \frac{\partial S}{\partial f^i} \right) (f_{(\text{new})}^i - f^i), \\ &= \sum_{i=1}^N \left( \frac{\partial S}{\partial f^i} \right) \left( \sum_{\mu=1}^n x^{\mu} e_{\mu}^i \right), \\ &= \sum_{\mu=1}^n \left( \sum_{i=1}^N \frac{\partial S}{\partial f^i} e_{\mu}^i \right) x^{\mu}, \end{aligned}$$

which tells us what  $s_{\mu}$  should be. Similarly,  $c_{\mu}$ ,  $g_{\mu\nu}$ , and  $h_{\mu\nu}$  can be found,

$$c_{\mu} = \sum_i e_{\mu}^i \frac{\partial C}{\partial f^i}, \quad (28)$$

$$g_{\mu\nu} = - \sum_{ij} e_{\mu}^i e_{\nu}^j \frac{\partial^2 S}{\partial f^i \partial f^j}, \quad (29)$$

$$h_{\mu\nu} = \sum_{ij} e_{\mu}^i e_{\nu}^j \frac{\partial^2 C}{\partial f^i \partial f^j}. \quad (30)$$

Thus, if we know the basis vectors, we can find the quadratic functions  $s(x)$  and  $c(x)$ .

Obviously, the above definitions shall not be valid to arbitrary distances from the point in question. The quadratic models are reliable only in the vicinity of the current  $f$  where cubic and higher powers can be neglected. Thus, the step size at each iteration must be such that

$$|\delta f|^2 \leq l_0^2, \quad (31)$$

for some  $l_0$ . We thus need to define the concept of distance in this abstract space. Recall that this means we need to define a metric,

$$ds^2 \equiv \sum_{ij} \bar{g}_{ij} df^i df^j, \quad (32)$$

Note that the metric  $\bar{g}_{ij}$  is different from the function  $g_{\mu\nu}$  defined by Eq. (26). Experience (see [21]) has shown that the following definition of distance works well,

$$\bar{g}_{ij} = \frac{\delta_{ij}}{f^i}, \quad (33)$$

This needs to be compared with the expression for the Hessian matrix of entropy (notice that  $\bar{g}^{ij} = f^i \delta^{ij}$ ). It is straightforward to show that

$$ds^2 = \sum_{ij} \bar{g}_{ij} df^i df^j = \sum_{\mu\nu} g_{\mu\nu} x^\mu x^\nu, \quad (34)$$

while choosing  $l_0^2$  to be  $1/5$  of  $\sum f$  works well (see [21]). The algorithm works in the following way: at every iteration, when we are at a point in the  $f$  space, one considers a distance region such that the quadratic model is a good approximation in that region. We now find a subspace and within this subspace, we try to find the place where

- (1)  $s(x)$  is maximum,
- (2)  $c(x)$  equals some  $\tilde{C}_{\text{aim}}$ , and,
- (3) the distance of this new point from the old point is smaller than  $l_0$ .

### B. Construction of the subspace

So, how do we decide the basis vectors that span the subspace? One of our aims is to find the maximum of entropy on the surface of ellipsoid corresponding to  $\chi^2 = C_{\text{aim}}$ . So, naturally, the direction of the gradient of entropy must be one of the basis vectors. Since the metric in the space of interest is not Cartesian, there shall be a distinction between contravariant and covariant components of vectors in the space. Since the ‘‘position vector’’ of any point is  $f^i$ , a contravariant vector, gradient such as  $\partial S / \partial f^i$ , is going to be a covariant vector. So the first (contravariant) basis vector is

$$e_1^i = \sum_j \bar{g}^{ij} \frac{\partial S}{\partial f^j} = f^i \frac{\partial S}{\partial f^i}. \quad (35)$$

The meaning of this direction is easy to understand by recalling its equivalent in usual Cartesian space. In the usual situation,  $(\vec{\nabla}T) \cdot \hat{n} dr = dT$  (i.e., if we are at any point, and we go in the direction  $\hat{n}$  by a distance of  $dr$ , the change in the value of the function is  $dT$ ). It is obvious from this expression that when  $\hat{n}$  is parallel to the direction of gradient, the change  $dT$  is maximum. Thus, to maximize the change in  $f$ , we shall move in the direction parallel to  $\vec{\nabla}T$  so that

$$n^i = \sum_j \delta^{ij} \frac{\partial T}{\partial x^j}. \quad (36)$$

This equation should be compared with the definition of the first basis vector, Eq. (35), and since the Kronecker delta is the metric in a Cartesian space, the two equations are equivalent. Thus, the first basis vector tells us the direction in which the entropy change per unit distance is maximum.

Similarly, another basis vector could be

$$e_2^i = \sum_j \bar{g}^{ij} \frac{\partial C}{\partial f^j} = f^i \frac{\partial C}{\partial f^i}, \quad (37)$$

since we wish to change the  $\chi^2$  at every iteration so that we eventually reach the  $\chi^2 = C_{\text{aim}}$  surface. If we find what the two search directions (defined above) become after

incrementing by  $x^1 e_1 + x^2 e_2$ , the direction  $e_1$  shall stay within the subspace spanned by  $e_1$  and  $e_2$  but the direction  $e_2$  shall go out of the subspace (see [27]). This suggests that we choose more basis vectors such as

$$e_3^i = f^i \sum_j e_1^j \partial^2 C / \partial f^i \partial f^j, \quad (38)$$

$$e_4^i = f^i \sum_j e_2^j \partial^2 C / \partial f^i \partial f^j. \quad (39)$$

Experience has shown that a family of three or four such search directions gives quite a robust algorithm for solving the problem. In our problem, we chose the third search direction to be

$$e_3^i = f^i \sum_j \frac{\partial^2 C}{\partial f^i \partial f^j} \left( \frac{e_1^j}{L_s} - \frac{e_2^j}{L_c} \right), \quad (40)$$

where the following equations define the lengths  $L_s$  and  $L_c$  which are the gradient vectors

$$L_s = \left( \bar{g}^{ij} \frac{\partial S}{\partial x^i} \frac{\partial S}{\partial x^j} \right)^{\frac{1}{2}}, \quad L_c = \left( \bar{g}^{ij} \frac{\partial C}{\partial x^i} \frac{\partial C}{\partial x^j} \right)^{\frac{1}{2}}. \quad (41)$$

Our experience has shown that putting the factors of  $L_s$  and  $L_c$  in the definition of the third basis vector improves the speed of convergence of the answer.

### C. Optimization within the subspace

Once we have found the subspace (by finding the basis vectors in the space of all  $f$ s), we proceed as follows: we now wish to find the step, the coefficients  $x$  in Eq. (23). To do this, we shall solve a corresponding constrained optimization problem in the  $n$  dimensional subspace (as was stated in the previous subsection, we worked with  $n = 3$ , but we shall continue to explain the details for a general  $n$ ). Since the functions  $s(x)$  and  $c(x)$  are quadratic, the problem in the subspace is much simpler: it is a simple problem of quadratic programming (quadratic objective function with quadratic constraint). The only additional complication is that the quadratic model is not valid to arbitrary distances from the original point, so we need to satisfy an additional distance constraint.

Let us begin by recalling that both the matrices  $g$  and  $h$  are real-symmetric. Also, the matrix  $g$  is positive-definite. This is a result of the way we have defined the metric on the space [see Eq. (33)],

$$ds^2 = \sum_{ij} \bar{g}_{ij} df^i df^j = \sum_i \frac{(df^i)^2}{f^i} = \sum_{\mu\nu} g_{\mu\nu} x^\mu x^\nu \geq 0, \quad (42)$$

where in the last step we used Eq. (34). So, it is clear that  $g_{\mu\nu}$  is a positive-definite matrix (which implies that all its eigenvalues are positive). Thus if one of the eigenvalues of  $g_{\mu\nu}$  is a small positive number, numerical errors can cause

it to become negative. In our implementation of the algorithm, we choose to ignore any directions which are defined by eigenvalues which are too small. Additional simplification occurs if we simultaneously diagonalize the two matrices  $g$  and  $h$ .<sup>5</sup>

After simultaneous diagonalization, within the subspace, the quadratic model functions  $\tilde{S}$  and  $\tilde{C}$  are given by

$$\tilde{S}(x) = s_0 + \sum_{\mu} s_{\mu} x_{\mu} - \frac{1}{2} \sum_{\mu} x_{\mu}^2, \quad (43)$$

$$\tilde{C}(x) = C_0 + \sum_{\mu} c_{\mu} x_{\mu} + \frac{1}{2} \sum_{\mu} \gamma_{\mu} x_{\mu}^2. \quad (44)$$

the quantities  $s_{\mu}$ , etc. are now defined in terms of the new basis vectors (but the same old definitions). Since this causes the function  $g_{\mu\nu}$  to become a Kronecker delta, the distance constraint equation looks like

$$l^2 = \sum_{\mu} x_{\mu}^2 \leq l_0^2 \quad (\simeq 0.1 \sum f \quad \text{to} \quad 0.5 \sum f). \quad (45)$$

We chose the coefficient on the rhs to be 0.2 and we verified that the actual value of this number is unimportant. Typically, the function  $\tilde{C}$  is such that all its eigenvalues are (also) positive, then, the minimum value of the function  $\tilde{C}$  in the subspace (where the above definitions work) is

$$\tilde{C}_{\min} = C_0 - \frac{1}{2} \sum_{\mu} \frac{C_{\mu}^2}{\gamma_{\mu}}. \quad (46)$$

Thus, no matter what the global aim  $C_{\text{aim}}$  is, at a given iteration, within the subspace, we can not get to any values below  $\tilde{C}_{\min}$ . In fact, even trying to achieve  $\tilde{C}_{\min}$  is not a great idea since in that case we shall not use any information about  $\tilde{S}$ .

The real challenge in the subspace is to satisfy the distance constraint. Many different elaborate tricks have been mentioned in the literature to do this. We choose to not worry about getting a quick answer, hence we do the following: in order to ensure that the distance constraint always gets satisfied (i.e., we do not go too far from the present location in just one step), we shall choose to have a

<sup>5</sup>*Simultaneous diagonalization Theorem:* If  $A$  and  $B$  are real-symmetric matrices and  $B$  is positive-definite, then there exists an invertible matrix  $P$  s.t.  $P^T B P = I$  and  $P^T A P$  is diagonal. The diagonal entries of  $A$  are the roots of the polynomial  $\det(xB - A) = 0$ . Notice that this is different from finding a basis in which both are diagonal. Here, if  $B$  is chosen to be identity matrix, then  $P$  is orthogonal and  $P^T A P$  is the same as  $P^{-1} A P$ . This leads to the unique diagonal representation of the matrix (with the eigenvalues as the diagonal values). Recall that the eigenvalues of a matrix  $M$  are the solutions of the equation  $\det(M - \lambda I) = 0$ . There exist stable numerical algorithms to achieve this (see, e.g., p. 463 of [28]) which take in the two real-symmetric matrices  $A$  and  $B$  and returns the (nonsingular) matrix  $P$  and the diagonal matrix  $P^T A P$ .

$\tilde{C}_{\text{aim}}$  which is not too different from  $C_0$  (the present value of  $\chi^2$ ). We thus choose

$$\tilde{C}_{\text{aim}} = \max(a\tilde{C}_{\min} + (1-a)C_0, C_{\text{aim}}), \quad (47)$$

with  $a$  chosen to be a small number (e.g., 0.01). This causes the algorithm to take very small ‘‘baby steps’’ towards the answer. Numerical experience has shown that as far as our problem is concerned, this is good enough. Of course, the actual value of  $a$  or  $C_{\text{aim}}$  chosen is not important as long as the distance constraint gets satisfied.

The problem in the subspace is thus simplified to finding the point  $x_a$  such that the function  $\tilde{S}$  is maximum subject to the constraint that  $\tilde{C} = \tilde{C}_{\text{aim}}$  (and an additional constraint that the distance constraint must get satisfied). The technique of Lagrange’s undetermined multiplier is useful here: we wish to find the point on the curve  $\tilde{C} = \tilde{C}_{\text{aim}}$  where  $\tilde{S}$  is maximum; to find the desired point, we consider the set of points at which all the partial derivatives of the function

$$\tilde{Q} = \alpha\tilde{S} - \tilde{C}, \quad (48)$$

(for an undetermined  $\alpha$ ) vanish. For any  $\alpha$ , such points are given by

$$x_a = \frac{\alpha S_a - C_a}{\gamma_a + \alpha}. \quad (49)$$

So, for every value of  $\alpha$ , find the value of  $x_a$  and then the function  $\tilde{C}$ . We are after that value of  $\alpha$  which leads to  $\tilde{C} = \tilde{C}_{\text{aim}}$ , so we look for a solution of the equation  $\tilde{C}(\alpha) = \tilde{C}_{\text{aim}}$ . The function  $\tilde{C}(\alpha)$  is a monotonically increasing function of  $\alpha$  (see Fig. 2). Since the function  $\tilde{C}(\alpha) - \tilde{C}_{\text{aim}}$  often happens to be a quickly changing function of  $\alpha$  (especially while it is changing its sign), the solution for  $\alpha$  needs to be found to a high tolerance level.

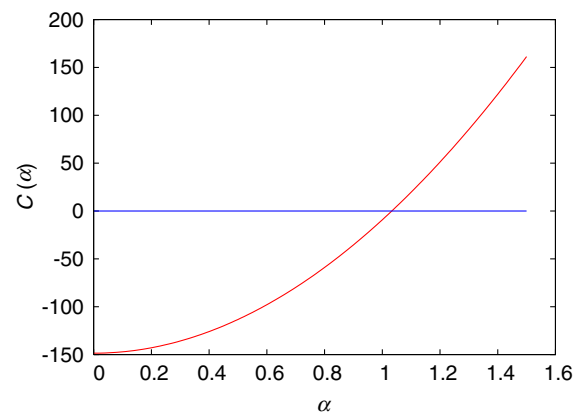


FIG. 2 (color online). The typical behavior of  $C(\alpha) = \tilde{C}(\alpha) - \tilde{C}_{\text{aim}}$  (the curve) as we change  $\alpha$ . The horizontal straight line corresponds to  $C(\alpha) = 0$ . The existence of a unique solution to the equation  $C(\alpha) = 0$  is absolutely necessary for the algorithm to work. Since  $C(\alpha)$  changes very quickly as we change  $\alpha$ , we need to find the root of  $C(\alpha) = 0$  to a high accuracy.

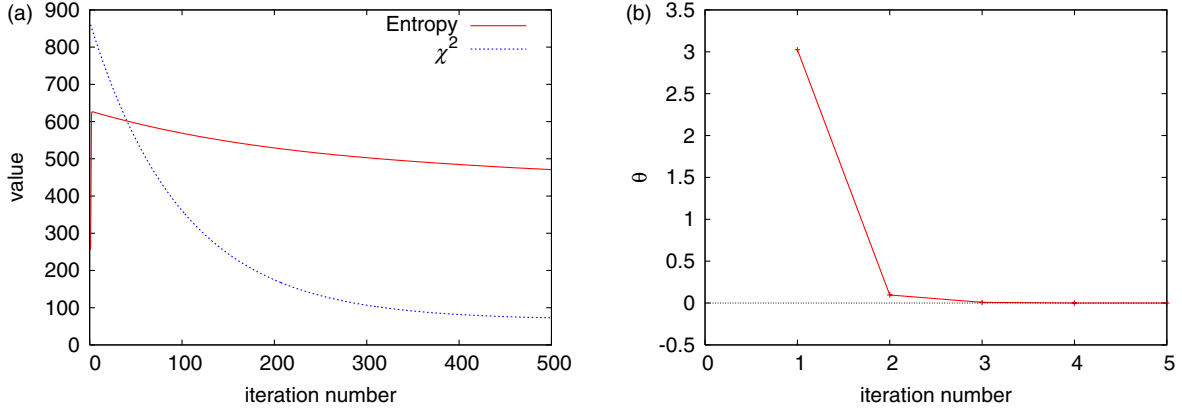


FIG. 3 (color online). (a) The typical evolution of entropy and  $\chi^2$  as the algorithm advances and (b) the illustration of the fact that the angle  $\theta$  (in radians) drops very quickly as the algorithm proceeds.

#### D. Stopping criterion

In solving the constrained optimization problem, one fact which becomes important is the following: at the point at which the constrained optimization problem gets solved, the gradient vectors of the two functions become parallel. Thus, if we find the unit vector in the direction of gradient of entropy and in the direction of gradient of chi squared function, the dot product of these two unit vectors (defined using the entropy metric) must become negligible as we head towards the point at which the constrained optimization problem gets solved. The unit vectors in the directions of gradients are (with  $L_s$  and  $L_c$  defined previously)

$$U_i^s = \frac{1}{L_s} \frac{\partial S}{\partial x^i}, \quad U_j^c = \frac{1}{L_c} \frac{\partial C}{\partial x^j}. \quad (50)$$

We thus expect the angle

$$\theta = \cos^{-1}(\bar{g}^{ij} U_i^s U_j^c), \quad (51)$$

to become too small (compared to a unit radian) as the algorithm proceeds (Fig. 3).

### IV. RECOVERING PRIMORDIAL POWER SPECTRUM

In this section, we shall (i) test the formalism presented in the previous section by trying to recover a featureless as well as feature-full sPPS from simulated noisy CMB data and (ii) apply the algorithm to actual WMAP 7-year as well as 9-year binned TT angular power spectrum [1] to recover the Primordial Power Spectrum. Thus, to begin with, we shall find out the radiation transfer function for the simplest set of assumptions, inject a featureless sPPS and get noise-free  $C_\ell^{TT}$  (which we shall refer to as theoretical  $C_\ell$ s). Next we shall add noise to these pure  $C_\ell$ s.

#### A. The radiative transport kernel

First, we need to set the values of the various cosmological parameters and get the corresponding radiation

transfer function. This can be done by making use of the codes such as CMBFAST [4], CAMB [5], or GTFast [29]. The results in this section are obtained from transfer function found using the code gTfast which itself is based on CMBFAST (version 4.0). It is important to notice that since in this work we shall only use the TT data, so, we only calculate the temperature radiation transfer function. To find out the transport kernel, we assume that the universe is spatially flat and dark energy is a cosmological constant (i.e., we have a spatially flat  $\Lambda$ CDM universe) and set the values of the cosmological parameters to their WMAP 9-year values [1] (WMAP9 + bao + h0): the values of various parameters to be fed into the code GTFast are given in Table I. We also assume that there are no tensor perturbations to the metric. We use Peebles recombination (rather than using RECFast) and assume that the primordial fluctuations are completely adiabatic. Finally, we shall not correct the transfer function for lensing of CMB, SZ effect or other effects that cause secondary anisotropies of CMB.

TABLE I. The table shows the values of various parameters for the run of GTFast for which we shall display the various results. It is important to note that the kernel  $G$  [defined by Eq. (7)] does not change too much if we change the values of the parameters slightly.

| CMBfast Parameter  | Value          |
|--------------------|----------------|
| lmax               | 1500           |
| ketamax            | 3000           |
| $\Omega_b$         | 0.0472         |
| $\Omega_c$         | 0.2408         |
| $\Omega_\Lambda$   | 0.712          |
| $\Omega_\nu$       | 0.0            |
| $H_0$              | 69.33 km/s/Mpc |
| Tcmb               | 2.72548 K      |
| $Y_{\text{He}}$    | 0.308          |
| $N_\nu$ (massless) | 3.04           |
| $N_\nu$ (massive)  | 0.0            |
| $\tau$             | 0.088          |



This shall give us the radiation transfer function from which we can easily evaluate the matrix  $G_{\ell k}$ . For the case we are dealing with, the matrix  $G_{\ell k}$  shall have dimensions  $1500 \times 6200$ . Finally, we would like to state that the results one obtains and conclusions that one draws should better not depend on the exact values of these parameters.

### B. Recovering test spectra

We can now inject a test sPPS which is a power law with  $A_S = 2.427 \times 10^{-9}$ ,  $n_S = 0.971$  (with  $k_0 = 0.002 \text{ Mpc}^{-1}$  and  $T_{\text{cmb}} = 2.72548 \times 10^6 \text{ } \mu\text{K}$ ) and get the corresponding theoretical  $C_\ell$ s, and add noise. The noise we add is dominated by cosmic variance at low  $\ell$  (less than 600) values while for high  $\ell$  values, the noise is dominated by instrumental errors. Figure 4 shows the result of using the algorithm described in the previous section to recover the sPPS in the present case. The following points are worth noting:

- (1) To get the result shown in Fig. 4, we set the parameter  $A$  in Eq. (17) to be  $5.4 \times 10^4$ . As was stated, the solution  $f_k = A$  is the location of global maximum of entropy in the  $f$  space. From Eq. (6), for any  $f_k$ , the corresponding  $\Delta_\zeta^2(k)$  can be found and so it is clear that  $f_k = A = 54000$  corresponds to  $\Delta_\zeta^2(k)$  being (approx)  $2 \times 10^{-8}$ . Thus, this value of  $A$  corresponds to the situation in which  $\Delta_\zeta^2(k) = 2 \times 10^{-8}$  is the solution with the maximum value of entropy.
- (2) In an actual CMB experiment (such as WMAP) the amount of noise (instrumental as well as that due to cosmic variance) is not the same for all scales,

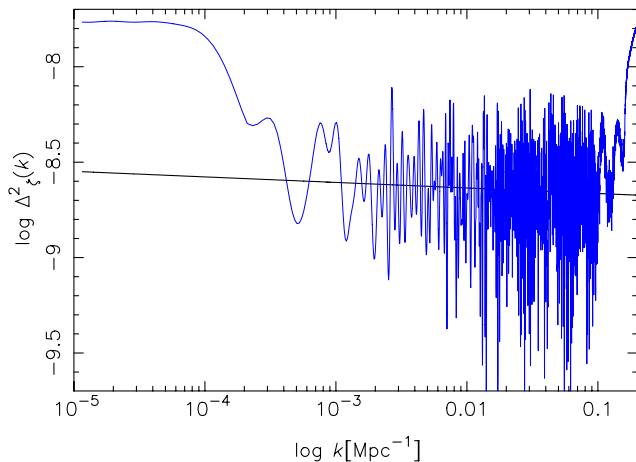


FIG. 4 (color online). Recovery (solid curve) of an injected featureless tilted red sPPS (the solid line) using simulated unbinned CMB data. The artificially added noise is dominated by cosmic variance for small (up to 600)  $\ell$  values and by instrumental noise at larger  $\ell$  values. This result is obtained when the parameter  $A$  in Eq. (17) is set to the value  $5.4 \times 10^4$ . This corresponds to assuming that the prior on  $A_S$  is  $2 \times 10^{-8}$ .

which means that our data is not equally good for all values of  $k$ . Figure 4 shows that at scales at which the noise is large (very low and very high  $\ell$  values which will correspond to very low and very high  $k$  values), the recovered  $f(k)$  tends to approach the value  $A$ , the recovery (at these scales) tends to be poor. Thus, at scales at which the noise is too large (or the kernel takes up negligible values), the recovery depends on what is the prior information we have about the solution. Thus the range of  $k$  values in which we can recover the PPS is too restricted.

- (3) Even at scales at which the noise is smaller (and at which we hope to recover well), we can have wiggly artificial features in the recovered PPS (in the form of peaks and dips). In the recovered power spectrum there could exist three kinds of features: (i) those that are actually there in the injected PPS (which are not there in the present case), (ii) those that are not there in the PPS but got introduced by the algorithm itself (these shall change as we change  $A$ ), and finally (iii) those that are artifacts of the added noise (a particular realization of the noise shall have outliers, if we consider different realizations of the noise, we shall get different recoveries).
- (4) The scales at which we typically introduce features in the sPPS are roughly  $10^{-3} \text{ Mpc}^{-1}$  to  $10^{-2} \text{ Mpc}^{-1}$ . We would like the recovery to be good at these scales. If we have data till very large value of  $\ell$ , and the noise at these large  $\ell$  values is very low compared to the noise at  $\ell$ s corresponding to the above scales, the algorithm shall ignore the few data points with larger noise and try to only take the data at the other scales seriously. Thus, if we wish to recover better at these scales we must focus on recovering the PPS using only the data from the  $\ell$  values corresponding to these scales.

In practice, the process of masking the sky causes the various  $C_\ell$ s to get correlated. The simplest situation in which we can hope to recover the sPPS is the one in which the data points corresponding to different  $\ell$  values are uncorrelated. This happens for the binned CMB data set (which has data only for 45  $\ell$  values). To make use of the binned data, we also work with a binned kernel which is defined by

$$G_{\ell k}^{\text{avg}} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \frac{G_{\ell k}}{N_\ell}, \quad (52)$$

where  $N_\ell$  is the number of  $\ell$  values in the bin (the binning of the kernel in  $\ell$  space is done exactly the way the  $C_\ell$ s are binned in  $\ell$  space by WMAP team). By using this averaged kernel and applying the algorithm to simulated binned data (with the added noise equal to the noise for WMAP 7-year binned data), we get the results shown in Fig. 5 (this time we show the results for many  $A$  values). We again get an

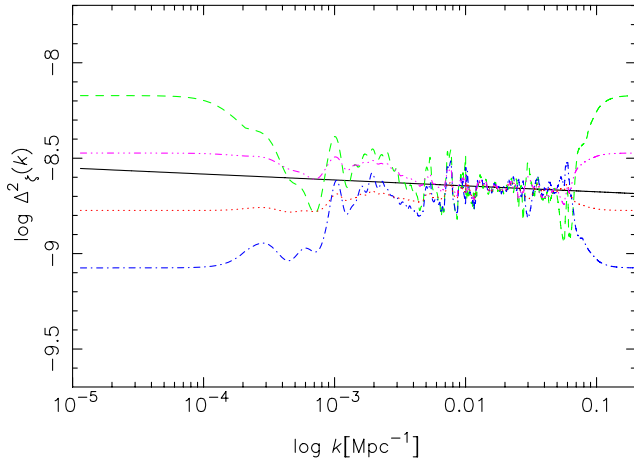


FIG. 5 (color online). Recovery of an injected featureless tilted red sPPS (the solid line) using simulated binned data. The different curves correspond to different values of  $A$ . The dashed curve corresponds to  $A = 18000$  which is equivalent to assuming the prior on  $A_S$  to be 3.365 in units of  $2 \times 10^{-9}$ . The corresponding values for other curves are: dotted curve (0.84), dash-dot curve (0.42) and dash-dot-dot-dot (1.68).

answer which at scales at which the noise is large, tends to the value of the default (i.e.,  $A$ ) while at scales at which the noise is relatively low, the recovery tends to fluctuate around the featureless injected signal. For a fixed value of  $A$ , the recovery at scales at which the noise is relatively lower shall be different if we consider different realizations of the noise. This is illustrated in Fig. 6: here the recovery shall be the same at scales with no data and shall be different at scales with data. The key question is whether we can recover features in the sPPS by this method. The

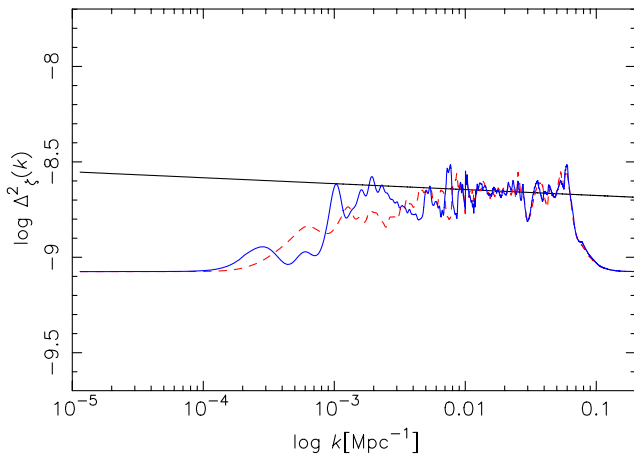


FIG. 6 (color online). The recovery at scales at which the data has lesser noise is not given by  $f_k = A$  but is dependent on the specific realization of the noise added. The solid straight line is the injected signal while the two curves are the recoveries for two different realizations of the noise. The chosen prior on  $A_S$  is 0.42 (in units of  $2 \times 10^{-9}$ ).

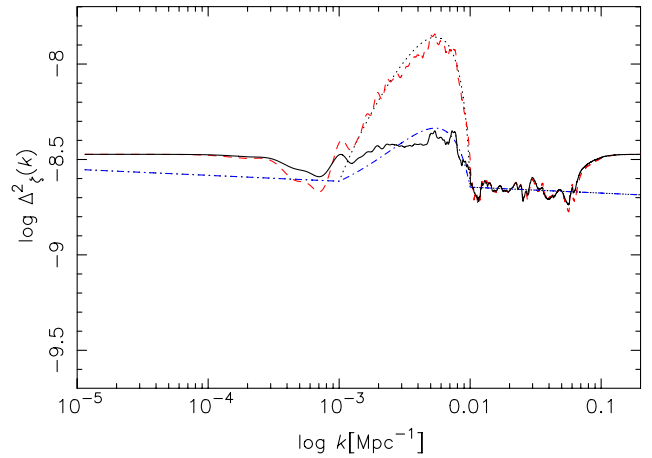


FIG. 7 (color online). Recovery of spectrum with bumpy features. The prior on  $A_S$  is set to 1.683 (in units of  $2 \times 10^{-9}$ ). The dotted curve and the dash-dot curve are the injected spectra while the solid and dashed curves are the recoveries. Had we introduced a feature at scales where the recovery goes back to the global maximum of entropy ( $A$ ) we could not have recovered it.

fact that this can be done is illustrated in Fig. 7: we just introduce a bumpy feature between the scales  $10^{-3} \text{ Mpc}^{-1}$  to  $10^{-2} \text{ Mpc}^{-1}$  and vary its height and see that unless the height of the bump is too small, the algorithm can recover it. Of course if we introduce a feature at a scale at which the data is not good or at which the kernel takes up negligible values, the feature shall not be recovered. Moreover, it is not surprising that the recovery is much better if the feature is more prominent.

### C. Actual binned CMB data

In this sub-section we apply the algorithm to actual CMB data. We use WMAP 7-year and 9-year binned  $TT$  data set and use it to recover the sPPS. The result for WMAP 7-year data is shown in Fig. 8. The details of the recovery of course depend on the chosen value of the parameter  $A$ . In the present context, the value of  $A$  represents our *a priori* knowledge (without using any data) of how much we think should be the scalar fluctuation in the metric in the early universe.

It may appear that if the conclusion depends on such *a priori* knowledge, we may not get anything worthwhile. But the following fact is worth noting: it is seen that at scales at which the noise is lesser, even though the recovered  $P(k)$  depends on the value of  $A$  chosen, this dependence is quite weak and quite predictable (as we increase  $A$  a lot or decrease it a lot, the recovery just “stretches” in the  $P(k)$  direction in the  $\ln P - \ln k$  plane).

One can keep on decreasing the value of  $A$  and see what happens. In this context, the case of  $A = 1$  is very interesting since this corresponds to using another familiar definition of entropy, the recovery for this case is illustrated in

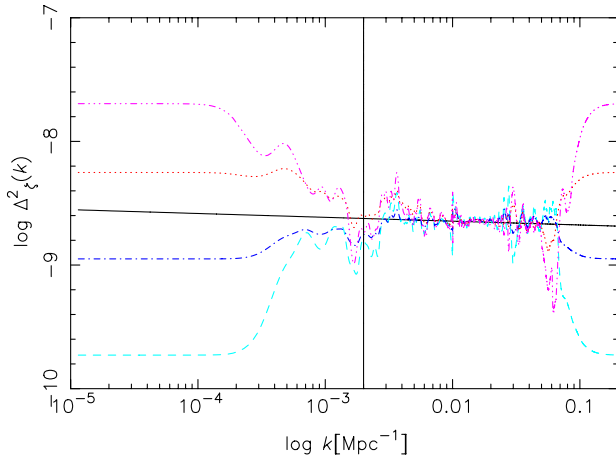


FIG. 8 (color online). The result of applying the algorithm to binned WMAP 7-year  $TT$  data. The solid vertical straight line corresponds to the pivot scale of  $k_0 = 0.002$  ( $\text{Mpc}^{-1}$ ). The solid line corresponds to the maximum likelihood sPPS that one gets if one assumes the sPPS to be a power law. The curves correspond to the following values of priors on  $A_S$  (in units of  $2 \times 10^{-9}$ ): dashed (0.09), dash-dot (0.56), dotted (2.8), and dash-dot-dot (10).

Fig. 9. What is interesting is that if we choose  $A$  to be too small, we begin to get an IR cutoff not very different from the one reported in the literature previously (see [17]), but, we also get an apparent UV cutoff. Moreover, such a small value of  $A$  causes the artificial features to get stretched so much that we may not consider the reconstruction to be trustworthy in this case.

### 1. WMAP 9-year data

The corresponding recovery results for WMAP 9-year data are shown in Fig. 10. The only difference between the WMAP 7- and 9-year binned  $TT$  data is that the error bars

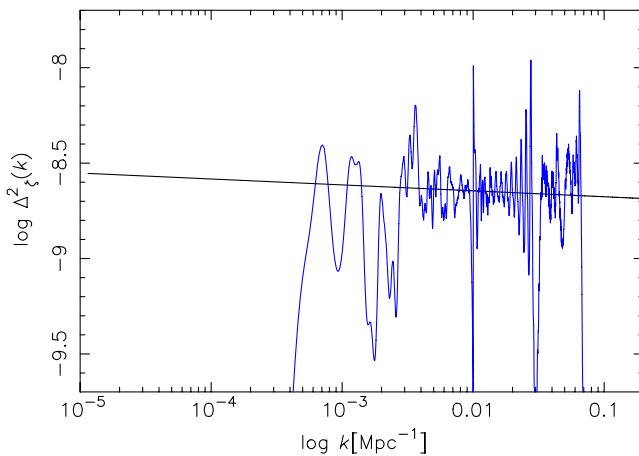


FIG. 9 (color online). Choosing  $A = 1.0$  causes the features to get overly “stretched,” and we find apparent IR and UV cutoffs in power.

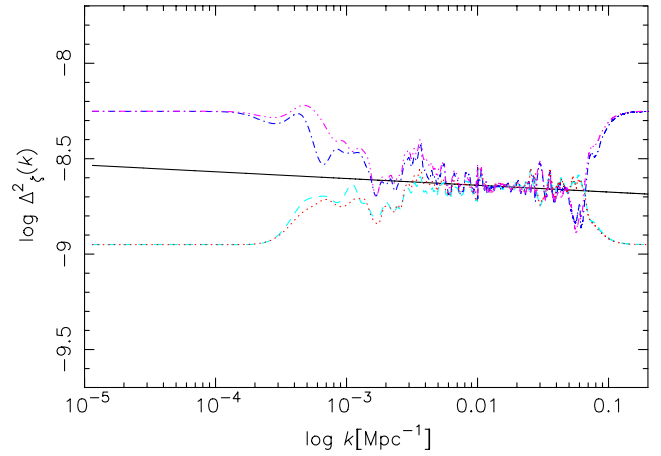


FIG. 10 (color online). The recovery of PPS using MEM with WMAP 9-year data (the difference between the two datasets is that the corresponding error bars are slightly smaller for WMAP 9-year data) for two values of  $A$ . The dotted curve with  $A = 3000$  (corresponding prior on  $A_S = 5.61$  in units of  $2 \times 10^{-9}$ ) is obtained using WMAP 9-year data while the dashed curve is for WMAP 7. For the other value of  $A$  (corresponding prior on  $A_S = 2.8$  in units of  $2 \times 10^{-9}$ ), dash-dot curve is for WMAP 9-year data while dash-dot-dot curve is for WMAP 7-year data. Notice that on very small and very large scales, the improved data does not change the recovery.

are slightly smaller for WMAP 9-year data. We thus see how the recovery shall change for smaller error bars.

### 2. A note on Planck data

The Planck Collaboration [2] also released its results and made the data publicly available recently. The great quality of data (with much smaller error bars, particularly at high  $\ell$ s) ensures that for high  $\ell$  values ( $\ell > 50$ ), the conventional  $\Lambda$ CDM model fits the data very well i.e., power law PPS is a very good fit to the data at these  $\ell$  values. The Planck Collaboration also reconstructed the PPS from their data (see [20]) for these high  $\ell$  values using a regularization method which leads to a linear problem (unlike ours). Though they report a small feature (at  $\ell \approx 1800$ ), they do not rule out that this happens due to unknown systematic effects. This lead to the conclusion that there are no significant features in the PPS from the data at large  $\ell$  values. For low  $\ell$  values (from  $\ell = 2$  to 49), the Planck Collaboration has provided the unbinned  $C_\ell$ s. For these low  $\ell$  values, the fact that the estimated  $C_\ell$ s are random variables with  $\chi^2$  distribution shows up in the form of asymmetric error bars. It may be very interesting to use our deconvolution method to the low  $\ell$  Planck data and see if there are any indications for features in the PPS. Since the noise at low  $\ell$ s can not be approximated as a Gaussian random variable, to use the Planck data robustly, we need to make a few changes in the formalism presented here. We thus leave the analysis of Planck data to look for features as a problem to be pursued in a future publication.

## V. SUMMARY AND DISCUSSION

In this work, we attempted to probe the amplitude and shape of scalar primordial power spectrum (sPPS) using the CMB data. We fixed the values of various cosmological parameters (apart from the ones specifying the sPPS itself) and formulated the problem as an inverse problem. To solve the inverse problem, we use the maximum entropy method which is a nonlinear regularization method. There exist many possible ways to employ the maximum entropy regularization, we use a particular definition of entropy and a particular algorithm to solve the corresponding constrained nonlinear optimization problem in a very large dimensional parameter space.

The way we have formulated the problem, there exists a parameter (which we called  $A$ ) whose value decides the location of global maximum of entropy in the space of all primordial power spectra. In the absence of any data, the algorithm shall just send every initial guess to the global maximum of entropy. Even in the presence of data, the following is worth noting

- (1) at scales where
  - (a) we have noisy data (thus, little or no information), or
  - (b) the kernel (to be inverted) takes up negligible values (again too large or too small  $k$  values),
 the  $\Delta_{\zeta}^2(k)$  recovered by MEM depends on the value of  $A$  chosen (as  $\Delta_{\zeta}^2(k) = A$  is the maximum entropy solution), while at the scales where the data is good, we recover something which has comparatively lesser dependence on what  $A$  we choose.

- (2) at scales at which the data is good, the  $\Delta_{\zeta}^2(k)$  recovered by MEM is consistent with a power law primordial power spectrum (with any possibly small deviations which we can not say anything about at this stage). This can be seen by comparing Fig. 8 with Figs. 5 and 7. While the existence of any small deviations from power law behavior can not be completely ruled out, this analysis reinforces our belief that any such possible deviations must be small.

This is by no means the last word on the existence of features in sPPS; this is not even the last word on the use of MEM for this purpose. The implementation of our algorithm to this problem until now does not seem to give any reason to believe that there are any serious deviations from the power law. We would like to mention that this is not completely unexpected, even in the light of existing papers such as [17] because the error bars at scales at which the features were recovered in those works are very large: the maximum entropy method cannot claim any features at scales where the error bars are so large.<sup>6</sup> This analysis

<sup>6</sup>This “rules out” (or at least renders them untrustworthy) many models of inflation considered in the literature in recent times.

shows that at scales at which the CMB data is trustworthy, the primordial power spectrum of scalar metric perturbations is, to a very good approximation, a power law.

In the future, one can look at the following prospects. We should be able to solve this problem of possible existence of features in sPPS without assuming the values of other cosmological parameters (i.e., without formulating this problem as a simple inversion problem). Even in the present formulation, there may be ways of combining results from different values of  $A$  to get a better recovery. One may wish to use the actual WMAP likelihood (or rather, the corresponding  $\chi_{\text{eff}}^2$ ) as a measure of misfit, but this is not easy in the way we have attempted to solve the problem (we need to know the  $\chi^2$  and its first two derivatives). Also, we have lost a lot of information in the process of binning the kernel and working with the binned, uncorrelated data. We would like to use all that lost information. Similarly, we have only used the  $TT$  angular power spectrum of CMB, we would also like to use the polarization spectra to probe the sPPS. We may also need to postprocess the recovered sPPS to get more useful information. Another interesting possibility worth exploring is the connection of maximum entropy deconvolution with other ways of deconvolution (e.g., Richard Lucy deconvolution).

## ACKNOWLEDGMENTS

The authors acknowledge the use of WMAP data and the use of codes such as CMBFAST, GTFast, and CAMB. The authors would also like to thank Tarun Souradeep (IUCAA, Pune) for reading through the manuscript and giving useful comments. G.G. thanks Rajaram Nityananda (NCRA, Pune), Mihir Arjunwadkar (CMS, Pune university, Pune), Abhilash Mishra (CALTECH, Pasadena), and Ranjeev Misra (IUCAA, Pune) for discussions at various stages of the work. G.G. also thanks the Council of Scientific and Industrial Research (CSIR), India, for the Research Grant Award No. 10-2(5)/2006(ii)-EU II. J.P. acknowledges support from the Swarnajayanti Fellowship, DST, India (awarded to Professor Tarun Souradeep, IUCAA, Pune, India).

## APPENDIX: TESTING THE METHOD

The main text described the material necessary to employ the maximum entropy inversion in any circumstance. The following points need to be noted (these are just tried and tested facts about the algorithm, many of which are illustrated here for the case of a toy problem shown in Fig. 11, whose solution is given in Fig. 12):

- (i) If we did not have any data available, the optimization problem would have involved maximizing entropy subject to no constraints. In such a scenario, the solution we should get must be  $f_i = A$  as that is where the global maximum of entropy is.

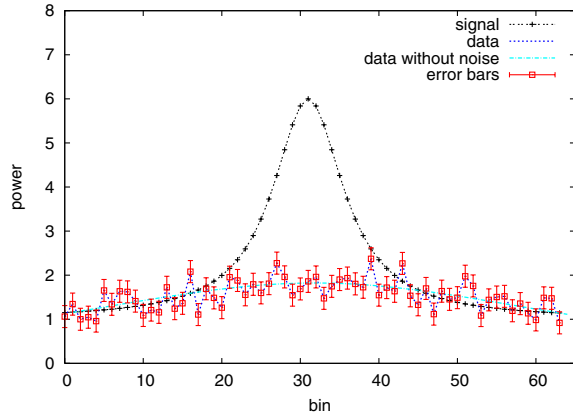


FIG. 11 (color online). A toy problem to test the algorithm. The signal, which has a bump, gets completely smoothed after the application of the kernel (chosen to be a Lorentzian profile), a known amount of random noise is then added giving the final data. Figure 12 illustrates the recovery with two distinct initial guesses.

- (ii) If the value of  $A$  is such that the  $\chi^2$  of the global maximum of entropy is smaller than  $C_{\text{aim}}$ , then  $f_i = A$  is itself the desired solution since “the data are too noisy for any information to be extracted” (see the last paragraph of p. 113 of [21]).
- (iii) It is not a surprise at all that choosing too small value of  $A$  should lead to negative value for entropy (the fact that depending upon the choice of  $A$ , sometimes we could be at locations in the parameter space with negative value of  $S$  has no impact on the solution of the problem) (see Fig. 13).
- (iv) In Eq. (48),  $\alpha = \infty$  corresponds to the unconstrained maximization of  $\tilde{S}$  irrespective of  $\tilde{C}$ . If we are too close to the global maximum of entropy

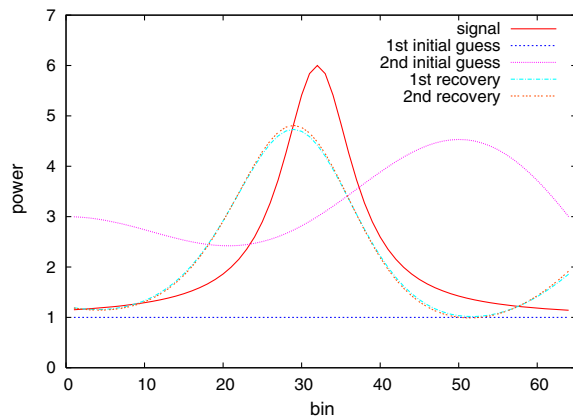


FIG. 12 (color online). An illustration of the fact that even completely different initial guesses lead to the same final recovery (done for the toy problem of Fig. 11). Notice that the location of the recovered bump and its amplitude are not exactly right: the quality of the recovery depends on many factors including the form of the kernel matrix itself.

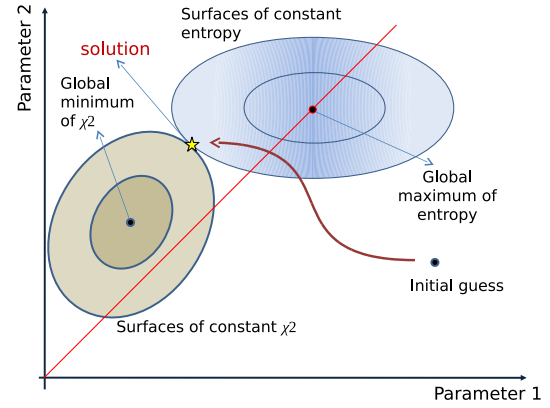


FIG. 13 (color online). The straight line in this solution to  $f_k = A$  for various values of  $A$ . Changing the value of  $A$  shall change the recovery because the point on the  $\chi^2 = \text{constant}$  surface with maximum entropy changes in the process.

- ( $f_i = A$ ), the value of  $\alpha$  required to solve the constrained optimization problem in the subspace [for  $\tilde{Q}$  defined by Eq. (48)] shall become too large. In this situation, it may be difficult to numerically find any solution for  $\alpha$ .
- (v) As long as we do not stay too close to the global maximum of entropy (so that numerical problems such as those stated in the previous point above do not turn up), the choice of the initial guess for running the algorithm is immaterial. That is, all the initial guesses shall lead to the same answer (see Fig. 12).
- (vi) All the above problems can be easily avoided if we just choose a value of  $A$  s.t. the  $\chi^2$  of  $f_i = A$  configuration is much higher than the  $\chi^2$  of initial guess (which better be more than  $C_{\text{aim}}$ ). Notice that this is not a requirement, just a trick. Also, this does not help us in finding any unique preferable value of  $A$ .
- (vii) For many kernels the exact value of  $C_{\text{aim}}$  chosen does not matter as far as the recovered  $f$  is concerned, as long as the final value of  $\theta$  becomes sufficiently small compared to a unit radian, all recoveries with different final  $\chi^2$  are almost the same. The  $\chi^2$  of signal (for a given realization) shall just fluctuate around (roughly)  $n_d$ , we have tested that if  $C_{\text{aim}}$  is set equal to  $C_{\text{signal}}$ , the recovery does not change. This happens to be true, e.g., for the case of CMB kernel, the case of our interest.
- (viii) The exact details of the shape of the final recovered solution does depend upon the actual value of  $A$  chosen: the  $\chi^2 = C_{\text{aim}}$  surface can be thought of as a closed ellipsoidal surface in the  $n_s$  dimensional  $f$ -space while as we change  $A$ , we define the line  $f_i = A$  as being the location of global maximum of entropy for these different values

of  $A$ . This will of course mean that as we change  $A$ , the place where the entropy is maximum on the  $\chi^2 = C_{\text{aim}}$  surface shall also change. Thus, as we continuously change  $A$ , we shall get a family of recoveries (see Fig. 13). So, the details of the recovered answer depends on the chosen value of this free (or adjustable) parameter. But since the global maximum of entropy is at  $f_i = A$ , the

value of  $A$  represents the “background” (i.e., *a priori*) knowledge of how much the power in various bins is, without using any knowledge of data at all.

- (ix) Whether the recovery is good or bad, depends on the details of the kernel. For the case of CMB kernel, we have tested that the recovery is often quite good.

- 
- [1] E. Komatsu *et al.*, *Astrophys. J. Suppl. Ser.* **192**, 18 (2011).  
 [2] Planck Collaboration *et al.*, [arXiv:1303.5076](https://arxiv.org/abs/1303.5076).  
 [3] A. A. Starobinsky, *Phys. Lett.* **91B**, 99 (1980); D. Kazanas, *Astrophys. J. Lett.* **241**, L59 (1980); A. H. Guth, *Phys. Rev. D* **23**, 347 (1981); A. D. Linde, *Phys. Lett.* **108B**, 389 (1982); A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).  
 [4] U. Seljak and M. Zaldarriaga, *Astrophys. J.* **469**, 437 (1996).  
 [5] A. Lewis, A. Challinor, and A. Lasenby, *Astrophys. J.* **538**, 473 (2000).  
 [6] A. Lewis and S. Bridle, *Phys. Rev. D* **66**, 103511 (2002).  
 [7] A. A. Starobinskiĭ, *JETP Lett.* **30**, 682 (1979); V. F. Mukhanov and G. V. Chibisov, *ZhETF Pisma Redaktsiiu* **33**, 549 (1981); S. W. Hawking, *Phys. Lett.* **115B**, 295 (1982); A. A. Starobinsky, *Phys. Lett.* **117B**, 175 (1982); A. H. Guth and S.-Y. Pi, *Phys. Rev. Lett.* **49**, 1110 (1982).  
 [8] A. Linde, *Contemp. Concepts Phys.* **5**, 1 (1990); D. Baumann, [arXiv:0907.5424](https://arxiv.org/abs/0907.5424); D. Langlois, *Lect. Notes Phys.* **800**, 1 (2010); L. Sriramkumar, [arXiv:0904.4584](https://arxiv.org/abs/0904.4584).  
 [9] R. K. Jain, P. Chingangbam, J.-O. Gong, L. Sriramkumar, and T. Souradeep, *J. Cosmol. Astropart. Phys.* **01** (2009) 009.  
 [10] R. K. Jain, P. Chingangbam, L. Sriramkumar, and T. Souradeep, *Phys. Rev. D* **82**, 023509 (2010).  
 [11] R. Saito, J. Yokoyama, and R. Nagata, *J. Cosmol. Astropart. Phys.* **06** (2008) 024; R. K. Jain, P. Chingangbam, and L. Sriramkumar, *J. Cosmol. Astropart. Phys.* **10** (2007) 003; S. Chongchitnan and G. Efstathiou, *J. Cosmol. Astropart. Phys.* **01** (2007) 011.  
 [12] E. Bugaev and P. Klimai, *Phys. Rev. D* **78**, 063515 (2008).  
 [13] S. M. Leach and A. R. Liddle, *Phys. Rev. D* **63**, 043508 (2001).  
 [14] A. A. Starobinskiĭ, *JETP Lett.* **55**, 489 (1992).  
 [15] J. Prasad and T. Souradeep, *Phys. Rev. D* **85**, 123008 (2012); K. Ichiki and R. Nagata, *Phys. Rev. D* **80**, 083002 (2009); K. Ichiki, R. Nagata, and J. Yokoyama, *Phys. Rev. D* **81**, 083010 (2010); D. K. Hazra, A. Shafieloo, and T. Souradeep, [arXiv:1303.5336](https://arxiv.org/abs/1303.5336).  
 [16] D. K. Hazra, M. Aich, R. K. Jain, L. Sriramkumar, and T. Souradeep, *J. Cosmol. Astropart. Phys.* **10** (2010) 008.  
 [17] A. Shafieloo and T. Souradeep, *Phys. Rev. D* **70**, 043523 (2004); R. Sinha and T. Souradeep, *Phys. Rev. D* **74**, 043518 (2006); A. Shafieloo, T. Souradeep, P. Manimaran, P. K. Panigrahi, and R. Rangarajan, *Phys. Rev. D* **75**, 123502 (2007); A. Shafieloo and T. Souradeep, *Phys. Rev. D* **78**, 023511 (2008); G. Nicholson and C. R. Contaldi, *J. Cosmol. Astropart. Phys.* **07** (2009) 011; D. K. Hazra, A. Shafieloo, and T. Souradeep, [arXiv:1303.4143](https://arxiv.org/abs/1303.4143).  
 [18] D. Tocchini-Valentini, Y. Hoffman, and J. Silk, *Mon. Not. R. Astron. Soc.* **367**, 1095 (2006).  
 [19] J. Hamann, A. Shafieloo, and T. Souradeep, *J. Cosmol. Astropart. Phys.* **04** (2010) 010.  
 [20] Planck Collaboration *et al.*, [arXiv:1303.5082](https://arxiv.org/abs/1303.5082).  
 [21] J. Skilling and R. K. Bryan, *Mon. Not. R. Astron. Soc.* **211**, 111 (1984).  
 [22] E. T. Jaynes, *Phys. Rev.* **108**, 171 (1957).  
 [23] E. T. Jaynes, *Phys. Rev.* **106**, 620 (1957).  
 [24] R. Narayan and R. Nityananda, *Annu. Rev. Astron. Astrophys.* **24**, 127 (1986).  
 [25] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).  
 [26] D. J. C. MacKay, *Neural Comput.* **4**, 415 (1992).  
 [27] J. Skilling and S. F. Gull, *Maximum-Entropy and Bayesian Methods in Inverse Problems* (Springer, New York, 1985); J. Skilling, *Maximum Entropy and Bayesian Methods in Applied Statistics* (Cambridge University Press, Cambridge, 1986); *Maximum Entropy and Bayesian Methods* (Kluwer, Boston, 1989).  
 [28] L. C. F. v. V and H. Gene, *Matrix Computations* (Johns Hopkins, Baltimore, 1996).  
 [29] E. Komatsu and D. N. Spergel, *Phys. Rev. D* **63**, 063002 (2001).