Power spectrum estimators for large CMB datasets

Julian Borrill
Center for Particle Astrophysics, University of California, Berkeley, California 94720
and National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720

(Received 9 December 1997; published 8 December 1998)

Forthcoming high-resolution observations of the cosmic microwave background radiation will generate data sets many orders of magnitude larger than have been obtained to date. The size and complexity of such data sets present a very serious challenge to analyzing them with existing or anticipated computers. Here we present an investigation of the currently favored algorithm for obtaining the power spectrum from a sky-temperature map—the quadratic estimator. We show that, while improving on a direct evaluation of the likelihood function, current implementations still inherently scale as the equivalent of $O(N_p^3)$ in the number of pixels or worse, and demonstrate the critical importance of choosing the right implementation for a particular data set. [S0556-2821(98)03024-0]

PACS number(s): 98.70.Vc, 98.80.Es

I. INTRODUCTION

Over the next ten years a number of ground-based, balloon-borne and satellite observations of the cosmic microwave background (CMB) are planned with sufficient resolution to determine the CMB power spectrum up to multipoles $l \sim 1000$ or more (for a general review of forthcoming observations see [1]). According to current theory this will provide us with the locations, amplitudes, and shapes of the Doppler peaks, and hence the values of the fundamental cosmological parameters to unprecedented accuracy. The CMB will then have lived up to its promise of being an extremely powerful discriminant between cosmological models [2].

In preparation for these data sets considerable effort is being put into developing ways of extracting the information they contain. Typically the raw data are cleaned and converted into a time-ordered data set. This is then turned into a sky temperature map, and the map analyzed to find its power spectrum. Having obtained the power spectrum of the data set we can compare it with the predictions of any class of cosmological models to determine the most likely values of the parameters associated with that class. While it would also be possible to estimate such cosmological parameters directly from the data, this would require the assumption of a class of models during the data analysis. We therefore choose to provide the more generic result of the power spectrum.

We consider the analysis of an N_p pixel map from a simple pointing experiment for multipoles $2 \leqslant l \leqslant N_l$ in bins $1 \leqslant b \leqslant N_b$—i.e. we determine the location of and the curvature about the peak of the maximum likelihood function of the binned power spectrum coefficients C_b. Traditionally this has been done by directly evaluating the likelihood function $\mathcal{L}(C)$ over the bin parameter space to locate its maximum (for example for the COBE data [3], with the additional refinement of using a complete set of cut-sky basis functions in place of the incomplete spherical harmonics). The fastest general solution uses Cholesky decomposition of the data covariance matrix, costing $O(N_p^2)$ in size and $O(N_p^3)$ in time for a single point in the bin parameter space. Searching this space—for example by maximum gradient ascent—typically requires $O(N_b)$ likelihood evaluations at each of many steps. Moreover, calculating the curvature of the likelihood function at its maximum by discrete differencing requires further $O(N_b^2)$ likelihood evaluations. Overall such calculations scale as $O(N_p^2)$ in size and $O(N_b^2 N_p^3)$ in time, and become hopelessly intractable for any of the anticipated data sets. However, it is worth noting that where it is possible, exhaustive direct evaluation does have the advantage of giving the full probability distribution rather than just the maximum and its variance.

There have been a number of attempts to improve on this scaling—for example by using quadratic estimation [4,5], by transforming to the signal-to-noise eigenbasis [5], by using approximations for the determinant [6], or by assuming azimuthally symmetric noise [as is expected from the Microwave Anisotropy Probe (MAP) satellite] [7]. However, none has yet provided a way to search a high dimensional multipole bin parameter space under an arbitrarily complex data set in fewer than $O(N_p^3)$ operations. In this work we discuss the two proposed implementations of quadratic estimation, clarifying the full scaling behavior of each and hence demonstrating the circumstances under which each should be used.

II. MAXIMUM LIKELIHOOD ANALYSIS

Any observation of the CMB contains both signal and noise

$$\Delta_i = s_i + n_i \tag{1}$$

at each pixel. For independent noise and zero-mean signal the covariance matrix of the data,

$$M \equiv \langle \Delta \Delta^T \rangle = \langle s s^T \rangle + \langle n n^T \rangle, \tag{2}$$

is symmetric, positive definite and dense. Given any binned power spectrum C_b and a shape parameter C_l^s within each bin such that

$$C_l = C_l^s C_b, \quad l \in b, \tag{3}$$

we can construct the signal covariance matrix; for a simple pointing experiment this is

$$S_{ii'} \equiv \langle s_i s_{i'} \rangle = \sum_{b=0}^{N_b} \frac{2l+1}{4\pi} C_l B_l^2 P_l(\cos\,\theta_{ii'})$$
$$= \sum_{b=0}^{N_b} C_b \sum_{l \in b} \frac{2l+1}{4\pi} C_l^s B_l^2 P_l(\cos\,\theta_{ii'}) \tag{4}$$

where B_l is the multipole beam map and $\theta_{ii'}$ is the angular separation of pixels i,i'. Taking the CMB fluctuations to be Gaussian, consistent with inflationary cosmologies, the probability of the observed data set given the assumed power spectrum is then

$$\mathcal{L}(C) \equiv P(\Delta|C) = \frac{e^{-\Delta^T M^{-1} \Delta/2}}{(2\pi)^{N_p/2} |M|^{1/2}}. \tag{5}$$

Assuming a uniform prior, so that $P(C|\Delta) \propto P(\Delta|C)$, the most likely power spectrum will be that which maximizes $\mathcal{L}(C)$, with covariance matrix $\mathcal{Q}$ where

$$[\mathcal{Q}^{-1}]_{bb'} \equiv - \left. \frac{\partial^2 \mathcal{L}}{\partial C_b \partial C_{b'}} \right|_{C=C_{max}}. \tag{6}$$

III. QUADRATIC ESTIMATORS

We review the derivation of the quadratic estimator given by Bond, Jaffe and Knox [5]. For an alternative derivation, see Tegmark [4]. Since we are primarily interested in finding the maximum of $\mathcal{L}$, and evaluating its curvature matrix at this maximum, we solve

$$\frac{\partial \ln \mathcal{L}}{\partial C} = 0 \tag{7}$$

iteratively by the Newton-Raphson method. Starting from some (sufficiently good) target power spectrum C the correction

$$\delta C = - \left[\frac{\partial^2 \ln \mathcal{L}}{\partial C^2} \right]^{-1} \frac{\partial \ln \mathcal{L}}{\partial C} \tag{8}$$

gives rapid convergence to the maximum of $\mathcal{L}$.

Taking the logarithm and repeatedly differentiating Eq. (5),

$$\ln \mathcal{L} = -\frac{1}{2} (\Delta^T M^{-1} \Delta + \mathrm{Tr}[\ln M] + N_p \ln 2\pi)$$

$$\frac{\partial \ln \mathcal{L}}{\partial C_b} = \frac{1}{2} \left(\Delta^T M^{-1} \frac{\partial S}{\partial C_b} M^{-1} \Delta - \mathrm{Tr}\left[M^{-1} \frac{\partial S}{\partial C_b} \right] \right)$$

TABLE I. Scaling in the calculation of the Fisher matrix F for the two quadratic estimator algorithms A1 (first two rows), A2 (last three rows).

Term		Memory	Operations
$X_b = M^{-1} \frac{\partial S}{\partial C_b}$	$\forall b$	$O(N_b N_p^2)$	$O(N_b N_p^3)$
$\mathrm{Tr}[X_b X_{b'}]$	$\forall b, b'$	$O(N_p^2)$	$O(N_b^2 N_p^2)$
$X_l = M^{-1} Y_l$	$\forall l$	$O(N_l^2 N_p)$	$O(N_l^2 N_p^2)$
$Z_{ll'} = Y_l^T X_{l'}$	$\forall l, l'$	$O(N_l^4)$	$O(N_l^4 N_p)$
$\mathrm{Tr}[Z_{ll'} Z_{ll'}^T]$	$\forall l, l'$	$O(N_l^2)$	$O(N_l^4)$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial C_b \partial C_{b'}} = \frac{1}{2} \Bigg(\Delta^T \Bigg[M^{-1} \frac{\partial^2 S}{\partial C_b \partial C_{b'}} M^{-1} - 2 M^{-1} \frac{\partial S}{\partial C_b} M^{-1} \frac{\partial S}{\partial C_{b'}} M^{-1} \Bigg] \Delta - \mathrm{Tr}\Bigg[M^{-1} \frac{\partial^2 S}{\partial C_b \partial C_{b'}} M^{-1} - M^{-1} \frac{\partial S}{\partial C_b} M^{-1} \frac{\partial S}{\partial C_{b'}} \Bigg] \Bigg). \tag{9}$$

Now if instead of the computationally intensive full curvature matrix we settle for its much simpler ensemble average (i.e. the Fisher information matrix), we have

$$F_{bb'} = - \left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial C_b \partial C_{b'}} \right\rangle = \frac{1}{2} \mathrm{Tr}\left[M^{-1} \frac{\partial S}{\partial C_b} M^{-1} \frac{\partial S}{\partial C_{b'}} \right] \tag{10}$$

and Eq. (8) reduces to

$$\delta C = F^{-1} \frac{\partial \ln \mathcal{L}}{\partial C}. \tag{11}$$

Note that this procedure both locates the maximum and generates the (albeit approximated) covariance matrix F^{-1}.

The most computationally expensive calculation here is the evaluation of the Fisher matrix, for which two methods have been proposed. Noting that, from Eq. (4), the derivative matrix for each bin,

$$\frac{\partial S}{\partial C_b} = \sum_{l \in b} \frac{2l+1}{4\pi} C_l^s B_l^2 P_l, \tag{12}$$

is independent of iterative step, Bond, Jaffe and Knox [5] calculate them explicitly and solve

$$M X_b = \frac{\partial S}{\partial C_b} \tag{13}$$

column by column for each bin. The first two rows of Table I show the cost of evaluating the Fisher matrix this way.

Alternatively, Tegmark [4] has pointed out that each $(N_p \times N_p)$ Legendre polynomial matrix can be factorized

TABLE II. Size and time costs for the calculation of the Fisher matrix F for archetypal datasets on a SUN Ultra II for the two quadratic estimator algorithms A1, A2.

				Size		Time	
Dataset	N_p	N_b	N_l	A1 $O(N_b N_p^2)$	A2 $O(N_l^4)$	A1 $O(N_b N_p^3)$	A2 $O(N_l^4 N_p)$
COBE	10^3	30	30	240 Mb	8 Mb	15 min	1 min
MAXIMA-1/ BOOMERANG N. AMERICA	10^4	30	1000	24 Gb	8 Tb	10 days	20 yr
MAXIMA-2/ BOOMERANG ANTARCTICA	10^5	30	1000	2.4 Tb	8 Tb	30 yr	200 yr
MAP	10^6	1000	1000	8 Pb	8 Tb	1 Myr	2 kyr
PLANCK	10^7	1000	1000	800 Pb	8 Tb	1 Gyr	20 kyr

into the product of the corresponding $[N_p \times (2l+1)]$ spherical harmonic matrix and its transpose;

$$\frac{2l+1}{4\pi} P_l = Y_l Y_l^T \tag{14}$$

where

$$[Y_l]_{im} = Y_{lm}(\theta_i, \psi_i) \tag{15}$$

for the real spherical harmonic Y_{lm} in the direction of pixel i. Now

$$\frac{\partial S}{\partial C_b} = \sum_{l \in b} C_l^s B_l^2 Y_l Y_l^T \tag{16}$$

and we can use the invariance of the trace of a product of matrices under cyclic permutations to rewrite Eq. (10) as

$$F_{bb'} = \frac{1}{2} \sum_{l \in b} \sum_{l' \in b'} C_l^s C_{l'}^s B_l^2 B_{l'}^2 \times \mathrm{Tr}[(Y_{l'}^T M^{-1} Y_l)(Y_{l'}^T M^{-1} Y_l)^T] \tag{17}$$

and solve

$$M X_l = Y_l \tag{18}$$

column by column for each multipole, and

$$Z_{ll'} = Y_l^T X_{l'} \tag{19}$$

for each pair of multipoles, and hence each pair of bins. The last three rows of Table I show the cost of evaluating the Fisher matrix this way.

For CMB observations we have $N_b \ll N_p$, so that the first algorithm (A1) scales as $O(N_b N_p^2)$ in size and $O(N_b N_p^3)$ in time. Similarly $N_l^2 \gtrsim N_p$, with approximate equality for all-sky maps, so that the second algorithm (A2) scales as $O(N_l^4)$ in size and $O(N_l^4 N_p)$ in time. Table II shows the implications for a range of future experiments, scaled from implementations of each algorithm applied to an unbinned reduced Cosmic Background Explorer (COBE) data set. Note that no assumption has been made about binning in the Microwave Anisotropy Probe (MAP) and PLANCK data sets.

Finally we should note that these results are for a single iteration of the quadratic estimator. Given the nature of the algorithm (i.e. Newton-Raphson, with no guaranteed convergence) and our lack of prior knowledge of the shape of the likelihood function, it is hard to make concrete statements about the number of iterations required to achieve convergence to a given accuracy. However, our experience to date with the COBE, MAXIMA and BOOMERANG data sets suggests that it is at most very weakly dependent on the size of the data set, with data sets spanning 3 orders of magnitude, all requiring 5–10 iterations to achieve convergence at the 1% level starting from a flat spectrum.

IV. CONCLUSIONS

We have implemented two algorithms using the quadratic estimator as a means of determining the maximum likelihood power spectrum and its covariance matrix from a pixelized map of the CMB. Despite some previous claims, and in line with the assertion of Bond, Jaffe and Knox [5], while each is an improvement on the direct evaluation of the likelihood function, neither scales better in time than $O(N_p^3)$ in the number of pixels in the map. Ultimately the advantage of each is in a reduction of the scaling prefactor as compared with direct evaluation.

Comparing the two algorithms it is apparent that the choice of which to use for a particular data set is critical—with timings differing by up to a factor of 1000. Broadly speaking, observations of small patches of the sky, where $N_l \gg \sqrt{N_p}$, should be analyzed using A1, while all-sky maps, with $N_l \sim \sqrt{N_p}$, should be analyzed using A2.

All timings have been scaled from a small data set analyzed on a SUN Ultra II. Two further considerations immediately apply.

(i) Moving to parallel architectures will give a significant reduction in these timings. Implementation of each algorithm on the 512 processor Cray T3E at NERSC indicates that the improvement can be up to a factor of 1000. However, this does assume that we continue to keep all the necessary matrices simultaneously in core; any reduction to vector operations, relocation to disc, or recalculation will dramatically reduce this improvement.

(ii) The data sets under consideration will be obtained incrementally over the next 10 years. We should therefore take into consideration Moore's law—that computer power doubles every 18 months—to allow for corresponding increases in available memory and speed. Current trends do not, however, suggest any significant increase in the total parallel processor time ($O(10^4)$ hours) available to us.

Taken together, we can conclude that these algorithms, judiciously applied, will be sufficient to analyze 10^4 pixel data sets immediately, the 10^5 pixel data sets expected in the next 2 years some 6 years from now, and the 10^6 pixel data sets expected in 5–10 years only 16 years from now. However, since we would like to be able to analyze not only the actual data sets as soon as they are obtained, but also simulated data sets in advance of the observations, improved algorithms are still essential.

ACKNOWLEDGMENTS

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under the U.S. Department of Energy, Contract No. DE-ACO3-76SF00098, and used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Energy Research of the U.S. Department of Energy. This work is also part of the COMBAT project supported by NASA AISRP grant NAG5-3941. The author wishes to thank Andrew Jaffe for many enlightening discussions.

[1] G. F. Smoot, in *The Cosmic Microwave Background*, edited by C. Lineweaver *et al.* (Kluwer, Dordrecht, 1997).

[2] See for example L. Knox, Phys. Rev. D **52**, 4307 (1995); C. Jungman, M. Kamiankowski, A. Kosowsky and D. Spergel, *ibid.* **54**, 1332 (1996); W. Hu, N. Sugiyama and J. Silk, Nature (London) **386**, 37 (1997); J. R. Bond, G. Efstathiou and M. Tegmark, Mon. Not. R. Astron. Soc. **291**, 33 (1997); M. Zaldarriaga, D. Spergel and U. Seljak, Astrophys. J. **488**, 1 (1997).

[3] K. M. Górski, Astrophys. J. Lett. **430**, L85 (1994); K. M. Górski, G. Hinshaw, A. J. Banday, C. L. Bennett, E. L. Wright, A. Kogut, G. F. Smoot and P. Lubin, *ibid.* **430**, L89 (1994).

[4] M. Tegmark, Phys. Rev. D **55**, 5895 (1997).

[5] J. R. Bond, A. H. Jaffe and L. Knox, Phys. Rev. D **57**, 2117 (1998).

[6] A. Kogut, presented at the INPAC/ITP Conference on ''CMB Data Analysis and Parameter Extraction,'' Santa Barbara, 1997.

[7] S. P. Oh, D. N. Spergel and G. Hinshaw, Astrophys. J. (to be published), astro-ph/9805339.