Compressed WMAP 3-Year Low-ℓ TT Likelihood Evaluation

J. L. Sievers[1] and M. Nolta[1]

1. Introduction

We present a compressed version of the WMAP low-ℓ TT likelihood code, resulting in an order of magnitude speed improvement in the TT evaluation. The compression is lossless, keeping the 957 modes corresponding to $\ell \leq 30$, after removing the monopole and dipole.

2. Likelihood Evaluation and Optimal Compression

The standard form of the likelihood for Gaussian random variables, up to an irrelevant additive constant, is

$$-2\log(\mathcal{L}) = \Delta^T C^{-1}\Delta + log(|C|)$$

for data vector Δ, and covariance C, defined by $c_{ij} = \langle \delta_i \delta_j \rangle$. The covariance is usually made up of separate signal and noise contributions, S and N. The noise part may contain terms other than pure instrumental noise. In particular, one can add projection terms to the noise.

It is not infrequent that the signal matrix spans only a subspace of the data, for instance due to oversampling the data. Calculating the exact likelihood requires one to perform an n^3 operation on the covariance matrix, which is computationally expensive if the covariance matrix is large. In this case, one can benefit greatly from only considering the subspace spanned by possible signal matrices - *i.e.,* throw out "noise-only" combinations of the data. To do this requires a rotation into the space in which both the noise and the signal are diagonal. For a fixed S this is always possible. First, rotate and re-scale into a space in which the noise is the identity matrix. Define $N^{-1/2}$ to be $V_N \Lambda_N^{-1/2} V_N^T$, where Λ_N are the eigenvalues of N, and V_N the corresponding eigenvectors. Then the following holds (since $N^{-1/2}N^{1/2}$ equals the identity matrix):

$$-2\log(\mathcal{L}) = \Delta^T N^{-1/2} N^{1/2} (S+N)^{-1} N^{1/2} N^{-1/2} + \log(|N^{1/2} N^{-1/2} (S+N) N^{-1/2} N^{1/2}|)$$

Define $\Delta_* \equiv N^{-1/2}\Delta$ and $S_* \equiv N^{-1/2} S N^{-1/2}$ and recall that the determinant of the product of matrices is the product of their determinants. Then the likelihood can be rewritten

$$-2\log(\mathcal{L}) = \Delta_*^T (S_* + I)^{-1} \Delta_* + \log(|S_* + I|) + \log(|N|)$$

[1]Canadian Institute for Theoretical Astrophysics, University of Toronto, ON M5S 3H8, Canada

The log of the determinant of N is a constant, which can be safely ignored. One can then rotate (without re-scaling) into a space in which S_* is diagonal using V_{S_*}, the eigenvectors of S_*. At this point, the noise remains the identity, and the signal matrix is diagonal. The diagonal elements of S are the signal-to-noise ratio (squared) of their corresponding noise. A chop in SNR is then the optimal compression of the data, under the assumtion that the true signal matrix is in some sense similar to S. Furthermore, if S spans only a subspace for all possible models, then the compression is not only optimal but *exact* if enough compressed modes are kept to span S. This is the case for the WMAP low-ℓ TT, where S is confined to $\ell \leq 30$ for a total of 957 modes after removing the monopole and dipole ($\sum_{\ell=2,30} 2\ell+1$), while the HEALPIX res-4 maps are 2680 pixels. We therefore expect that we can losslessly compress from matrices of size 2680 to matrices of size 957, roughly an order of magnitude increase in speed.

If we define $V_{S_*,big}$ to be the eigenvectors corresponding to the "large" eigenvalues of S_*, then the (rectangular) compression matrix is $Z \equiv N^{-1/2} V_{S_*,big}$. Any uncompressed matrix A can be compressed by $A^\dagger = Z^T A Z$, and vectors can be compressed by $\Delta^\dagger = Z^T \Delta$.

3. Application to WMAP

We carry out this compression for the WMAP data. The noise consists of a small ($1\mu K$) term along the diagonal to regularize the noise, and projection terms for the monopole, dipole, $\ell > 30$, and a foreground template. Note that if the noise changes, then in general the compression changes and must be recalculated. We find that with the WMAP data, there is a drop in the signal-to-noise of the modes at $n = 957$ of several orders of magnitude (see Figure 1). This is as expected since S ought only span 957 modes. We calculate the rotated signal matrices for each ℓ - $S_\ell^\dagger = Z^T S_\ell Z$, and the rotated data vector $Z^T \Delta$. The upper triangles of the $S_\ell^\dagger$ are saved to a FITS file, and $\Delta^\dagger$ is saved as a text file. We calculate the likelihood offset $|N| + \sum \Delta_{small}^2$, where the Δ_{small} are the rotated data values thrown out by the compression, so that likelihood values between different versions of the code can be directly compared. The low-ℓ only part of the TT code runs a factor of 12 faster on the CITA McKenzie cluster nodes, taking $\sim$0.2 seconds per evaluation. With polarization included, the overall likelihood evaluation takes 0.6 second, down from 2.6-5.0 seconds (depending on approximation level used in the uncompressed likelihood code), for an overall increase in speed of a factor of 4-8.

4. Conclusions

We present a lossless compression of the WMAP 3-year low-ℓ TT data. Uncompressed, this dominates the execution time of the WMAP likelihood code. The compression matrix

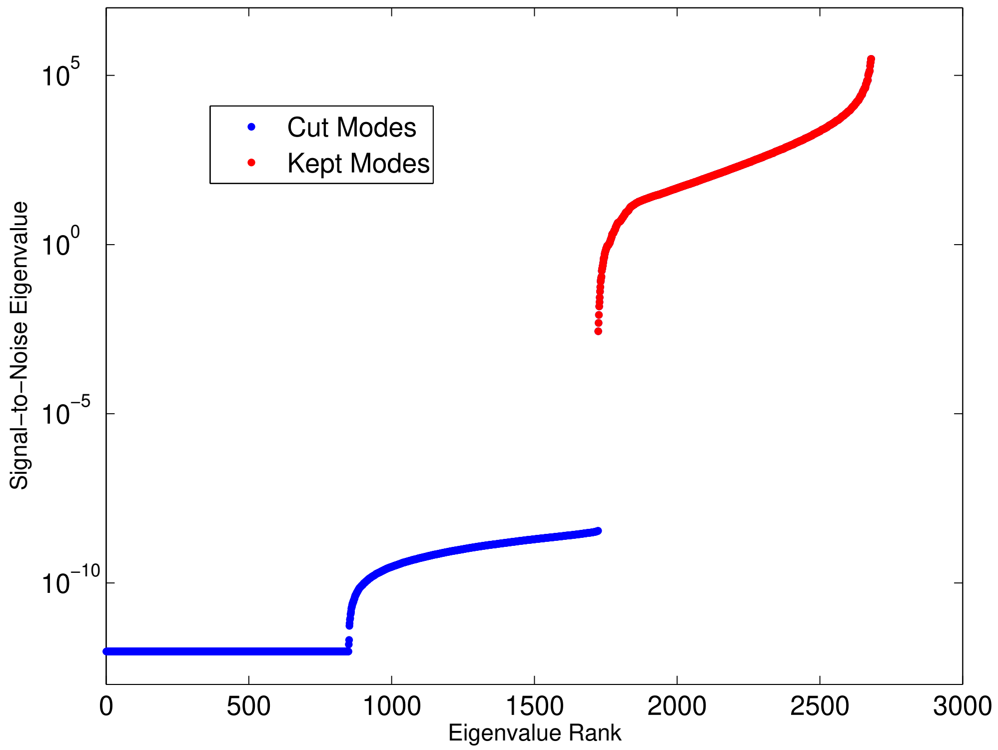

Fig. 1.— Signal-to-Noise eigenvalues of the WMAP 3-year low-ℓ TT ($\ell \leq 30$) for a typical ΛCDM cosmology. It is expected that, after removing the monopole and the dipole, 957 modes should have CMB signal. The large drop in the eigenvalues is indeed after the 957^{th} largest one, and is six orders of magnitude. The particular details of the eigenvalues depend on the assumed cosmology, but the large jump in eigenvalues and the subspace spanned by the eigenvectors of the 957 largest eigenvalues are independent of the spectrum.

depends on the noise, and so if the noise matrix (including changes to foreground marginalization) is changed, the compressed data and matrices must be recalculated. With the compression, we speed up the overall execution of the likelihood by at least a factor of 4. For many current cosmological parameter searches (especially flat, unlensed models), the overall execution time is dominated by the WMAP likelihood evaluation, so parameter chains can converge substantially faster by adopting this compression.