Bayesian CMB lensing

Gravitational lensing studies have become a powerful probe of dark matter, the invisibility of which provides the main obstacle for detection. In particular, the ground based Atacama Cosmology Telescope (ACT) and the South Pole Telescope (SPT) have mapped the cosmic microwave background (CMB) at such unprecedented resolution as too allow a detection of weak gravitational lensing from the CMB alone. The CMB shows a picture of radiation fluctuations frozen at the instant the universe became transparent. Estimating the gravitational lensing of the CMB is important for two reasons. First, if the CMB is mapped at a sufficient resolution one can use weak lensing estimates to construct a map of dark matter in the sky. Second, weak lensing estimates can be used, in principle, to un-distort the observed lensed CMB and construct the original unlensed CMB radiation fluctuations. Both of these maps, the unlensed CMB radiation field and the dark matter field, are deep probes into the nature of cosmology and cosmic structure.

Over the past year, the data from two ground based telescopes, ACT and SPT, have resulted in the first direct detection of weak lensing from the CMB alone, see Das et al. (2011) and van Engelen et al. (2012). In the coming years, the data from the Plank satellite and new instrumentation being installed on these telescopes (ACTpol and SPTpol) will begin probing this lensing a much greater scale. This will eventually help cosmologists gain an understanding of gravity waves, dark matter and dark energy. Unfortunately, the estimators which have been developed for weak lensing are still not fully understood. The state-of-the-art estimator, the quadratic estimator developed by Hu and Okomoto (2001, 2002), works in part through a delicate cancelation of terms in an infinite Taylor expansion of the lensing effect on the CMB. The effect of this cancelation is particularly sensitive to foreground contaminants and sky masking, which if not fully accounted for, limit the statistical inference obtainable from these studies. We propose a new Bayesian estimator which has the potential to revolutionize the way weak lensing studies are done.

Possibly the most promising alternative to the quadratic estimator, is Bayesian lensing. Indeed, Bayesian techniques applied to the lensed CMB observations have the potential for drastically changing the way lensing is estimated and used for inference. From the geometry of weak lensing, most of the lensing power comes from matter at a redshift $z \approx 2$. At these distances the matter distribution is well approximated by linear theory which predicts the matter density fluctuations are nearly Gaussian. Moreover, the unlensed CMB Θ is also extremely close to Gaussian. From a statistical perspective, this is a perfect scenario for Bayesian methods, in that both the observations and the unknown lensing potential are *physically predicted* to be Gaussian random fields. Moreover, the structure of the lensing operation gives a compelling case for MCMC posterior sampling techniques. To see an example, we show the feasibility of Gibbs sampling for generating posterior samples of the unlensed CMB $\Theta(x)$ and the lensing potential $\phi(x)$ from observations in the form $\Theta^{\text{obs}}(x) = \Theta(x + \nabla\phi(x)) + N(x)$. The Bayesian solution to the estimation of CMB lensing is to generate random draws from the posterior distribution, given the observed data, on the lensing potential and the unlensed (noiseless) CMB field. This approach is attractive for statistical inference since posterior draws are easy to use and to interpret for scientific inference. Moreover, posterior distributions can often be sequentially updated to incorporate additional information from additional data or other experiments.

History of the Bayesian lensing problem.

A full solution to this problem would handle non-stationary noise, non-stationary beam, cut sky or masking, In this paper we ... one of the main obstacles for the Bayesian lensing problem is ...

There are two components to this solution. The first.

1 Weak Lensing

This section describes the basics of CMB lensing and The effect of weak lensing is to simply remap the CMB, preserving surface brightness. Up to leading order, the remapping displacements are given by $\nabla\phi$, where ϕ denotes a lensing potential and is the planer projection of a three dimensional gravitational potential (see Dodelson, S. (2003), for example). Therefore the lensed CMB can be written $T(\boldsymbol{x} + \nabla\phi(\boldsymbol{x}))$ where $T(\boldsymbol{x})$ denotes the unlensed CMB temperature fluctuations projected to the observable sky. The goal of weak lensing surveys is to use the lensed observations $T(\boldsymbol{x}+\nabla\phi(\boldsymbol{x}))$ (with additional noise) to estimate ϕ or the spectral density of ϕ. In the full sky, $\boldsymbol{x}$ represents an observational direction on the unit sphere. However, we will be focusing on the small angle limit so that $\boldsymbol{x}$ can be modeled as a variable in $\mathbb{R}^2$. The Einstein principle along with properties of quantum mechanics predicts that $T(\boldsymbol{x})$ is a Gaussian isotropic random field. These properties translate to the independence of the Fourier transform of T across different frequencies. However, for a fixed lensing potential ϕ, the lensed CMB becomes non-isotropic, which leads to a correlation in the Fourier transform across different frequencies. The quadratic estimator takes advantage of this correlation and uses weighted sums of Fourier cross products to unbiasedly (up to leading order) estimate the lensing potential. The quadratic estimator is derived under the assumption that the observed lensed CMB field is contaminated by additive noise and an instrumental beam. Throughout this proposal we let $T^{\text{obs}}(x)$ denote the observed CMB field with noise (denoted $N(\boldsymbol{x})$) so that

$$\text{data}(\boldsymbol{x}) = T(\boldsymbol{x} + \nabla\phi(\boldsymbol{x})) + n(\boldsymbol{x})$$

The quadratic estimator is based on a first order Taylor approximation in $\nabla\phi$ on the lensed CMB field: $T(\boldsymbol{x} + \nabla\phi(\boldsymbol{x})) = T(\boldsymbol{x}) + \nabla\phi(\boldsymbol{x}) \cdot \nabla T(\boldsymbol{x}) + O(\phi^2)$. In Anderes and Paul (2012) they showed that this estimator is essentially a generalized least square regression estimator obtained by stacking the cross product of the Fourier transform separated at a certain lag.

2 Modivation: simple two parameter toy model

To motivate our solution to the Bayesian lensing problem we start with a simple two parameter statistical problem. This system has two unknown parameters θ, ϕ with data given by

$$\text{data} = \theta + \phi + n$$

where n denotes additive noise. In the Bayesian setting, the posterior distribution is computed as

$$P(\theta, \phi|\text{data}) \propto P(\text{data}|\theta, \phi)P(\theta, \phi) \tag{1}$$

where $P(\text{data}|\theta, \phi)$ denotes the likelihood of the data given θ, ϕ and $P(\theta, \phi)$ denotes the prior on θ, ϕ. The Gibbs sampler is a widely used algorithm for generating (asymptotic) samples from $P(\theta, \phi|\text{data})$ [add citations]. The algorithm generates a Markov chain of parameter values $(\theta^1, \phi^1), (\theta^2, \phi^2), \ldots$ generated by iteratively sampling from the conditional distributions:

$$\begin{aligned}
\theta^{i+1} &\sim P(\theta|\phi^i, \text{data}) \\
\phi^{i+1} &\sim P(\phi|\theta^{i+1}, \text{data}).
\end{aligned}$$

A useful heuristic for determining the convergence rate of a Gibbs chain is the extent to which the two parameters θ and ϕ are dependent in $P(\phi, \theta|\text{data})$.

A highly dependent posterior $P(\theta, \phi|\text{data})$ leads to a slow Gibbs chain, near independence leads to a fast Gibbs chain. Indeed, exact independence gives a sample of the posterior after one Gibbs step. A technique for accelerating the convergence of a Gibbs sampler is to find a re-parameterization of θ and ϕ in a way which makes the posterior less dependent. In the remainder of this section we discuss a specific re-parameterization which, by analogy, can be applied to Bayesian lensing.

The relevant situation for Bayesian lensing is the case that θ and ϕ are highly negatively correlated in $P(\theta, \phi|\text{data})$. This motivates re-parameterizing (θ, ϕ) to $(\widetilde{\theta}, \phi)$ where $\widetilde{\theta} \equiv \theta + \phi$ so that

$$\text{data} = \widetilde{\theta} + n.$$

In the statistics literature, (θ, ϕ) is commonly referred to as an **ancillary parameterization** whereas $(\widetilde{\theta}, \phi)$ is referred to as a **sufficient parameterization** [add citations]. Figure 2 illustrates the difference between an ancillary versus sufficient posterior distribution for our simple two parameter model. The left plot shows the posterior density contours for the ancillary parameterization (θ, ϕ), along with 40 steps of a Gibbs sampler. Conversely, the right plot shows the posterior density contours for the sufficient chain $(\widetilde{\theta}, \phi)$ with 40 Gibbs steps. Notice that negative correlation in the ancillary parameterization manifests in near independence for the sufficient chain. Indeed, the slower the ancillary chain the faster the sufficient chain and vice-versa.

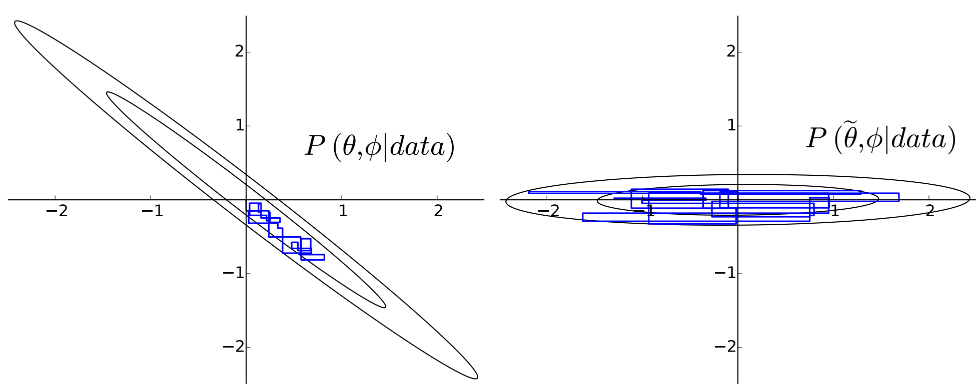

Figure 1: *Left:* density contours of the **ancillary** chain $P(\theta, \phi|\text{data})$ with 40 steps of a Gibbs sampler. *Right:* density contours of the **sufficient** chain $P(\widetilde{\theta}, \phi|\text{data})$ with 40 steps of a Gibbs sampler.

3 Ancillary and sufficient parameters for the lensed CMB

We will see that the ancillary parameterization for the lensed CMB problem is extreemely slow, whereas the sufficient parameterization is fast. Indeed, the analog to the equation, data $= \theta + \phi + n$, from Section 2 for lensed CMB is given by

$$\mathrm{data}(x) = T(x + \nabla\phi(x)) + n(x)$$

where the unlensed CMB temperature field T and the lensing potential ϕ are the two unknown parameters. In the ancillary Gibbs chain proceeds in the usual way:

$$T^{i+1} \sim P(T|\phi^i, \mathrm{data}) \tag{2}$$

$$\phi^{i+1} \sim P(\phi|T^{i+1}, \mathrm{data}). \tag{3}$$

Sampling from $P(T|\phi^i, \mathrm{data})$ is simply a Gaussian random field prediction problem since conditioning on ϕ^i models the data as

$$\mathrm{data}(x) = T(\underbrace{x + \nabla\phi^i(x)}_{\text{known obs locations}}) + n(x).$$

In otherwords, he data is a noisy version of T observed on an irregular grid. Conversly, when sampling from $P(\phi|T^{i+1}, \mathrm{data})$ the data is of the form

$$\mathrm{data}(x) = \underbrace{T^{i+1}}_{\text{known}}(x + \nabla\phi(x)) + n(x).$$

Both of these conditionals make the Gibbs very slow to converge. The case is esasterbated in the situation the noise level is small. For example, in the second conditional, if T^{i+1} is knownw and fixed, the extent of possible ϕ's which are possible under $P(\phi|T^{i+1}, \mathrm{data})$ is very small compared to the possible ϕ's in $P(\phi, T|\mathrm{data})$ when T is allowd to vary. This suggests a highly dependent posteror $P(\phi, T|\mathrm{data})$. This was also noitce by [Cite Lewis and Challanore] for the first conditnioal.

This clearly modivates attempting to quantify a sufficient parameterization. The analog to the sufficient toy problem, data $= \widetilde{\theta} + n$, in Section 2 to the lensed CMB is

$$\mathrm{data}(x) = \widetilde{T}(x) + n(x)$$

where now $\widetilde{T}$ denotes the lensed CMB tempuruature field with no noise or beam. The suffient chain then procceeds as

$$\widetilde{T}^{i+1} \sim P(\widetilde{T}|\phi^i, \mathrm{data}) \tag{4}$$

$$\phi^{i+1} \sim P(\phi|\widetilde{T}^{i+1}, \mathrm{data}). \tag{5}$$

In the following two ections we discuss these two conditional in detail.

4 $\widetilde{T}^{i+1} \sim P(\widetilde{T}|\phi^i, \mathbf{data})$

5 $\phi^{i+1} \sim P(\phi|\widetilde{T}^{i+1}, \mathbf{data})$

Theorem 1.

$$\frac{\partial}{\partial \phi_k} \log P(\phi|\tilde{T}, data) \propto -\frac{ik_q}{2\pi} \int e^{-ix\cdot k}[A^q(x)B(x)]dk + \frac{\phi_k}{C_k^{\phi\phi}}$$

where

$$A^q(x) \equiv \frac{\partial \tilde{T}}{\partial x_q}(x - \nabla\phi(x)) \tag{6}$$

$$B_k \equiv \frac{1}{C_k^{TT}} FFT\left\{\tilde{T}(x - \nabla\phi(x))\right\} \tag{7}$$