Bayesian estimates of CMB gravitational lensing

Ethan Anderes[*1] **Benjamin Wandelt**[2] **and Guilhem Lavaux**[2]

[1] *Department of Statistics, University of California, Davis CA 95616, USA.*

[2] *Institut d' Astrophysique de Paris.*

Abstract: 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.

Keywords and phrases: CMB, gravitational lensing, Bayesian, Gibbs sampler.

Over the past few years the data from ground based telecopse (ACT, SPT and Bicep2) and the Plank satellite and have resulted in unprecedicented detection of weak lensing of the CMB (see Das et al. (2011) and van Engelen et al. (2012), Plank). These observations are constraining cosmological models of gravity waves, dark matter and dark energy. The state-of-the-art estimator of CMB lensing, 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.

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. The Bayesian goal in this case 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. 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, denoted T, 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.

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.

[*] Research supported by: NSF DMS-1007480, NSF CAREER DMS-1252795

[†] Code available at url: https://github.com/EthanAnderes/BayesianCmbLensing.git

1. Weak lensing primer and a Bayesian challenge

This section describes the basics of CMB lensing and Bayesian estimates of lensing. 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 $T(\boldsymbol{x} + \nabla\phi(x))$ where $T(x)$ denotes the unlensed CMB temperature fluctuations projected to the observable sky. The goal of weak lensing surveys is to use the lensed observations $T(x + \nabla\phi(x))$ (with additional noise) to estimate ϕ or the spectral density of ϕ. In the full sky, x represents an observational direction on the unit sphere. However, we will be focusing on the small angle limit so that x can be modeled as a variable in $\mathbb{R}^2$. The Einstein principle along with properties of quantum mechanics predicts that $T(x)$ 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 $T^{\mathrm{obs}}(x)$ denote the observed CMB field with noise (denoted $n(x)$) so that

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

The quadratic estimator is based on a first order Taylor approximation in $\nabla\phi$ on the lensed CMB field: $T(x + \nabla\phi()) = T(x) + \nabla\phi(x)\cdot\nabla T(x) + O(\phi^2)$. 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.

There has been active interest in devloping a Bayesian estimator of the lensing potential ϕ jointly with the unlensed CMB T given the data. A very natural approach to generating posterior samples is to In the ancillary Gibbs chain proceeds in the usual way:

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

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

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 exacerbated in the situation when noise level is small. For example, in the second conditional, if T^{i+1} is known and fixed, the extent of the 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 allowed to vary. This suggests a highly dependent posterior $P(\phi, T|\mathrm{data})$. This was also noticed by [Cite Lewis and Challanore] for the first conditional.

2. Two parameter analogy

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

$$\mathrm{data} = t + \varphi + n$$

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

$$P(t,\varphi|\text{data}) \propto P(\text{data}|t,\varphi)P(t,\varphi) \tag{3}$$

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

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

A useful heuristic for determining the convergence rate of a Gibbs chain is the extent to which the two parameters t and φ are dependent in $P(\varphi,t|\text{data})$. A highly dependent posterior $P(t,\varphi|\text{data})$ 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 t 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 t and φ are highly negatively correlated in $P(t,\varphi|\text{data})$. This motivates re-parameterizing (t,φ) to $(\widetilde{t},\varphi)$ where $\widetilde{t}\equiv t+\varphi$ so that

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

In the statistics literature, (t,φ) is commonly referred to as an **ancillary parameterization** whereas $(\widetilde{t},\varphi)$ 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 (t,φ), along with 40 steps of a Gibbs sampler. Conversely, the right plot shows the posterior density contours for the sufficient chain $(\widetilde{t},\varphi)$ 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.

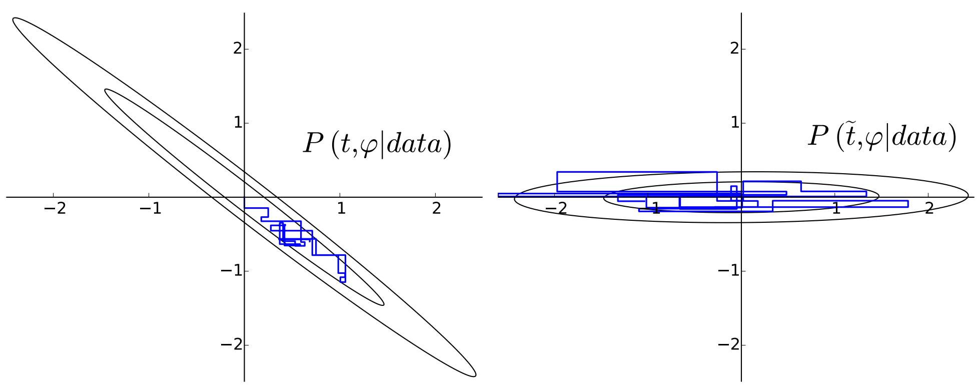

FIG 1. Left: *density contours of the* **ancillary** *chain* $P(t,\varphi|data)$ *with 40 steps of a Gibbs sampler.* Right: *density contours of the* **sufficient** *chain* $P(\widetilde{t},\varphi|data)$ *with 40 steps of a Gibbs sampler.*

3. Ancillary and sufficient parameters for the lensed CMB

The analog to the Ancillary toy problem is related to the lensed CMB problem as follows

$$\text{data}(x) = T(x+\nabla\phi(x)) + n(x) \quad \textbf{\textit{analogous to}} \quad \text{data} = t+\varphi+n$$

where the unlensed CMB temperature field T and the lensing potential ϕ are the two unknown parameters. Since we have seen in Section 1 that this Ancillary chain is exceedingly slow this clearly modivates attempting to quantify a sufficient parameterization. The analog to the sufficient toy problem, data $= \widetilde{t} + n$, in Section 2 to the lensed CMB is

$$\text{data}(x) = \widetilde{T}(x) + n(x) \quad \textbf{\textit{analogous to}} \quad \text{data} = \widetilde{t} + n$$

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

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

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

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

3.1. Anti-lensing approximation

Notice in the analog toy model the relation between the suffient parameter $\widetilde{t}$ and the ancillary parameter t is through the equation $\widetilde{t} - \varphi = t$. A similar type of relation hold in the lensed CMB problem which becomes key for sampling from both conditionals distributions (4) and (5). The analogy is what we are referring to as the *anti-lensing approximation*

$$\widetilde{T}(x - \nabla\phi(x)) \approx T(x) \quad \textbf{\textit{analogous to}} \quad \widetilde{t} - \varphi = t. \tag{6}$$

There are two points to make here. First, anti-lensing is a good approximation to the inverse lensing operation $\widetilde{T}(x) \to T(x)$. Second, in the methodology that we describe below, the anti-lensing approximation can be made exact with the introdution of a field rotation.

Regarding the accuracy of the anti-lensing approximation... and is subdomanty to the esimtation uncertainty of the displacement field. This is do too the fact that the expected size of the displacement $\nabla\phi$ is much smaller than the correlation length scale of $\nabla\phi$. Indeed, the typoical displacement size $\nabla\phi(x)$ is less than 3 arcmin whereas the correlation length scale of ϕ is on the order of degrees.

Regarding the field rotation, notice one can introduce a field rotation parameter, call it ψ, which makes antilensing exact: $\widetilde{T}\big(x - \nabla\phi^a(x) + \nabla^\perp\psi^a(x)\big) = T(x)$ where $\nabla^\perp \equiv \big(-\frac{\partial}{\partial y}, \frac{\partial}{\partial x}\big)$.

We can quantify...this by showing the simluated Helmholts spectrums for the inverse lensign operation and show how small it is.

4. Hamiltonian sampler for $P(\phi|\widetilde{T}, \text{data})$

Throuout this paper the Fourier transform of any function $f(x)$ is denoted f_l or f_k so that $f_l = \int_{\mathbb{R}^2} e^{-ix\cdot l} f(x) \frac{dx}{2\pi}$ and $f(x) = \int_{\mathbb{R}^2} e^{ix\cdot l} f_l \frac{dl}{2\pi}$ where $l \in \mathbb{R}^2$ is a two dimensional frequency vector and $x \in \mathbb{R}^2$ is a two dimensional spatial coordinate.

Claim 1. *Under the anti-lensing approximation (6) for any nonzero frequecy vector* $l \equiv (l_1, l_2) \in \mathbb{R}^2$

$$\frac{\partial}{\partial \phi_l} \log P(\phi|\widetilde{T}, \text{data}) \propto -\frac{\phi_l}{C_l^{\phi\phi}} - \sum_{q=1,2} i l_q \int_{\mathbb{R}^2} e^{-ix\cdot l} A^q(x) B(x) \frac{dx}{2\pi}$$

where $\phi_l = \text{re}\phi_l + i\,\text{im}\phi_l$, $\frac{\partial}{\partial\phi_l} \equiv \frac{\partial}{\partial\,\text{re}\phi_l} + i\frac{\partial}{\partial\,\text{im}\phi_l}$ *and*

$$B_l \equiv \frac{1}{C_l^{TT}} \int e^{-ix\cdot l} \widetilde{T}(x - \nabla\phi(x)) \frac{dx}{2\pi} \tag{7}$$

$$A^q(x) \equiv \frac{\partial \widetilde{T}}{\partial x_q}\big(x - \nabla\phi(x)\big). \tag{8}$$

The main advantage of this claim is that the gradient can be computed by iterating Fourier transofrms.

Hamiltonian sampler algorithm...

Remark: We also remark that the gradient is also an un-normalized quadratic destimate when the noise is zero and ...

5. Iterative message passing algorithm for $P(\widetilde{T}|\phi, \text{data})$

6. Simulation examples

7. Concluding remarks

What we have accomplished

What needs to be done.

- High resolution embedding does not scale to Planck (the native resolution is ∼10 million pixels)
- Approximate Gaussian conditional sampling on Planck data resolution (using previous ϕ^i to seed a conj gradient?)
- Need fast anti-lensing operations in frequency space to compute $A^q(x)$ and $B(x)$ that does not requires high res $\tilde{T}$
- Incorporate spectral density uncertainty
- Incorporate polarization

Appendix A

Before we proceed to the proofs we say a few words regarding notation. Firstly, we do not differentiate, notationally, from the case of smooth random field with periodic boundary conditions defined on $(-L/2, L/2]^2$ and the case where $L \to \infty$ so that the Fourier series $\sum_{l \in \frac{2\pi}{L}\mathbb{Z}} e^{ix\cdot l} f_l \frac{2\pi/L}{2\pi}$ converges to the continuous Fourier transform $\int_{\mathbb{R}^2} e^{ix\cdot l} f_l \frac{dl}{2\pi}$. For example, at times we will refer to an infinitesimal area element dl or dk in Fourier space, which simply equals $\frac{2\pi}{L}$ for large L. In this case δ_l denotes a discrete dirac delta function which we equate with $1/dk$ when $l = 0$ and zero otherwise. Secondly, for any function $f(x)$ let $f^\phi(x) = f(x - \nabla\phi(x))$ denote anti-lensing of f and f^ϕ_l denote the Fourier transform of $f^\phi(x)$.

Proof of Claim 1. Since $\tilde{T}$ is sufficient for the unknown ϕ we have that

$$P(\phi|\widetilde{T}, \text{data}) = P(\phi|\widetilde{T}) \propto P(\widetilde{T}|\phi)P(\phi).$$

Since $\phi(x)$ is an isotropic random field with spectral density $C^{\phi\phi}_l$ we have that $E(\phi_l\, \phi^*_{l'}) = \delta_{l-l'} C^{\phi\phi}_l$. Therefore $E(\phi_l\, \phi^*_l) = \delta_0 C^{\phi\phi}_l$ and $E(\phi_l \phi_l) = 0$ one gets that the random variables $\text{re}\phi_l$, $\text{im}\phi_l$ are independent $\mathcal{N}(0, \frac{1}{2}\delta_0 C^{\phi\phi}_l)$ for each fixed l. Moreovoer $\phi(x)$ takes values in $\mathbb{R}$ so that $\phi_l = \phi^*_{-l}$. This implies (what exactly implies what here? clearly the moments don't tell us exactly that they are independent since $E(zw^*) = 0$ could happen when $w = z^*$, and $(\text{re}z, \text{im}z)^t \sim \mathcal{N}(0, \sigma^2 I)$. But I think this is the only case where $E(zw^*) = 0$ when we know z and w are marginally Gaussian.) that ϕ_l and are independent random variables over all l which are restricted to the a Hermitian half of the Fourier grid, denoted $\mathbb{H}$. In particular, if we exclude the zero frequency $l = 0$ we get

$$\log P(\phi) - c_1 = -\frac{1}{2}\sum_{k\in\mathbb{H}\setminus\{0\}}\left[\frac{(\mathrm{re}\phi_k)^2}{\frac{1}{2}\delta_0 C_k^{\phi\phi}} + \frac{(\mathrm{im}\phi_k)^2}{\frac{1}{2}\delta_0 C_k^{\phi\phi}}\right] = -\frac{1}{2}\int_{\mathbb{R}^2}\frac{|\phi_k|^2}{C_k^{\phi\phi}}dk \tag{9}$$

$$\log P(\widetilde{T}|\phi) - c_2 = -\frac{1}{2}\sum_{k\in\mathbb{H}\setminus\{0\}}\left[\frac{(\mathrm{re}\widetilde{T}_k^\phi)^2}{\frac{1}{2}\delta_0 C_k^{TT}} + \frac{(\mathrm{im}\widetilde{T}_k^\phi)^2}{\frac{1}{2}\delta_0 C_k^{TT}}\right] = -\frac{1}{2}\int_{\mathbb{R}^2}\frac{\big|\widetilde{T}_k^\phi\big|^2}{C_k^{TT}}dk \tag{10}$$

where c_1 and c_2 are constants and $\widetilde{T}^\phi(x) \equiv \widetilde{T}(x - \nabla\phi(x))$. Taking derivatives in (9) gives

$$\frac{\partial}{\partial\phi_l}\log P(\phi) = -2(dl)\frac{\phi_l}{C_l^{\phi\phi}}. \tag{11}$$

Taking derivatives in (10) gives

$$\frac{\partial}{\partial\,\mathrm{re}\phi_l}\log P(\widetilde{T}|\phi) = -\,\mathrm{re}\int_{\mathbb{R}^2}\frac{\partial\widetilde{T}_k^\phi}{\partial\,\mathrm{re}\phi_l}\,\frac{\widetilde{T}_k^{\phi^*}}{C_k^{TT}}\,dk \tag{12}$$

$$\frac{\partial}{\partial\,\mathrm{im}\phi_l}\log P(\widetilde{T}|\phi) = -\,\mathrm{re}\int_{\mathbb{R}^2}\frac{\partial\widetilde{T}_k^\phi}{\partial\,\mathrm{im}\phi_l}\,\frac{\widetilde{T}_k^{\phi^*}}{C_k^{TT}}\,dk. \tag{13}$$

Taking linear combinations of the two equalities in Lemma 1 we get

$$\frac{\partial\widetilde{T}_k^\phi}{\partial\,\mathrm{re}\phi_l} = \frac{1}{2}\frac{\partial\widetilde{T}_k^\phi}{\partial\phi_l} + \frac{1}{2}\frac{\partial\widetilde{T}_k^\phi}{\partial\phi_l^*} = \frac{dk}{2\pi}\sum_{q=1,2} il_q\left\{[(\nabla^q\widetilde{T})^\phi]_{k-l} - [(\nabla^q\widetilde{T})^\phi]_{k+l}\right\} \tag{14}$$

$$\frac{\partial\widetilde{T}_k^\phi}{\partial\,\mathrm{im}\phi_l} = \frac{-i}{2}\frac{\partial\widetilde{T}_k^\phi}{\partial\phi_l} + \frac{i}{2}\frac{\partial\widetilde{T}_k^\phi}{\partial\phi_l^*} = \frac{dk}{2\pi}\sum_{q=1,2} l_q\left\{-[(\nabla^q\widetilde{T})^\phi]_{k-l} - [(\nabla^q\widetilde{T})^\phi]_{k+l}\right\}. \tag{15}$$

Now the above two equations establish, by Lemma 2, that both integrals $\int_{\mathbb{R}^2}\frac{\partial\widetilde{T}_k^\phi}{\partial\,\mathrm{re}\phi_l}\,\frac{\widetilde{T}_k^{\phi^*}}{C_k^{TT}}\,dk$ and $\int_{\mathbb{R}^2}\frac{\partial\widetilde{T}_k^\phi}{\partial\,\mathrm{im}\phi_l}\,\frac{\widetilde{T}_k^{\phi^*}}{C_k^{TT}}\,dk$ are real which implies

$$\begin{aligned}\frac{\partial}{\partial\phi_l}\log P(\widetilde{T}|\phi) &= -\int_{\mathbb{R}^2}\frac{\partial\widetilde{T}_k^\phi}{\partial\phi_l}\,\frac{\widetilde{T}_k^{\phi^*}}{C_k^{TT}}\,dk\\ &= -\frac{dk}{\pi}\sum_{q=1,2} il_q\int_{\mathbb{R}^2}[(\nabla^q\widetilde{T})^\phi]_{k+l}\,\frac{\widetilde{T}_k^{\phi^*}}{C_k^{TT}}\,dk\\ &= -i2(dk)\sum_{q=1,2} l_q\int_{\mathbb{R}^2}[(\nabla^q\widetilde{T})^\phi]_{k+l}\,\frac{\widetilde{T}_k^{\phi^*}}{C_k^{TT}}\,\frac{dk}{2\pi}\\ &= -i2(dk)\sum_{q=1,2} l_q\int_{\mathbb{R}^2} e^{-ix\cdot l}A^q(x)\,B(x)\,\frac{dx}{2\pi}, \quad \text{by Lemma 3}\end{aligned}$$

where $A^q(x) \equiv (\nabla^q\widetilde{T})^\phi(x)$ and $B_k \equiv (\widetilde{T}_k^\phi)^*/C_k^{TT}$. □

Lemma 1.

$$\frac{\partial\widetilde{T}_k^\phi}{\partial\phi_l} = \frac{dk}{\pi}\sum_{q=1,2} -il_q[(\nabla^q\widetilde{T})^\phi]_{k+l} \tag{16}$$

$$\frac{\partial\widetilde{T}_k^\phi}{\partial\phi_l^*} = \frac{dk}{\pi}\sum_{q=1,2} il_q[(\nabla^q\widetilde{T})^\phi]_{k-l} \tag{17}$$

where $\nabla^q\widetilde{T} \equiv \frac{\partial\widetilde{T}}{\partial x_q}$.

Proof. First notice

$$\frac{\partial}{\partial\,\mathrm{re}\phi_l}\frac{\partial\phi(x)}{\partial x_q} = \int_{\mathbb{R}^2} ik_q e^{ix\cdot k}\frac{\partial\phi_k}{\partial\,\mathrm{re}\phi_l}\frac{dk}{2\pi} = \left[il_q e^{ix\cdot l} - il_q e^{-ix\cdot l}\right]\frac{dk}{2\pi} \tag{18}$$

$$\frac{\partial}{\partial\,\mathrm{im}\phi_l}\frac{\partial\phi(x)}{\partial x_q} = \int_{\mathbb{R}^2} ik_q e^{ix\cdot k}\frac{\partial\phi_k}{\partial\,\mathrm{im}\phi_l}\frac{dk}{2\pi} = \left[-l_q e^{ix\cdot l} - l_q e^{-ix\cdot l}\right]\frac{dk}{2\pi}. \tag{19}$$

This implies

$$\begin{aligned}
\frac{\partial\widetilde{T}^\phi_k}{\partial\phi_l} &= \frac{\partial}{\partial\phi_l}\int_{\mathbb{R}^2} e^{-ix\cdot k}\widetilde{T}(x+\nabla\phi(x))\frac{dx}{2\pi}\\
&= \sum_{q=1,2}\int_{\mathbb{R}^2} e^{-ix\cdot k}\nabla^q\widetilde{T}(x+\nabla\phi(x))\left[\frac{\partial}{\partial\,\mathrm{re}\phi_l}\frac{\partial\phi(x)}{\partial x_q} + i\frac{\partial}{\partial\,\mathrm{im}\phi_l}\frac{\partial\phi(x)}{\partial x_q}\right]\frac{dx}{2\pi}\\
&= \sum_{q=1,2}\frac{-il_q dk}{\pi}\int_{\mathbb{R}^2} e^{-ix\cdot(k+l)}\nabla^q\widetilde{T}(x+\nabla\phi(x))\frac{dx}{2\pi}, \quad \text{by (18) and (19)}\\
&= \sum_{q=1,2}\frac{-il_q dk}{\pi}[(\nabla^q\widetilde{T})^\phi]_{k+l}
\end{aligned} \tag{20}$$

Similarly

$$\frac{\partial\widetilde{T}^\phi_k}{\partial\phi^*_l} = \sum_{q=1,2}\frac{il_q dk}{\pi}\int_{\mathbb{R}^2} e^{-ix\cdot(k-l)}\nabla^q\widetilde{T}(x+\nabla\phi(x))\frac{dx}{\pi} \tag{21}$$

$$= \sum_{q=1,2}\frac{il_q dk}{\pi}[(\nabla^q\widetilde{T})^\phi]_{k-l}. \tag{22}$$

□

Lemma 2. *If $A(x)$ and $B(x)$ are real scalar fields then the two integrals, $\int_{\mathbb{R}^2} i\{A_{k-l} - A_{k+l}\}B^*_k dk$ and $\int_{\mathbb{R}^2}\{A_{k-l} + A_{k+l}\}B^*_k dk$, are both real numbers.*

Proof. By a simple change of variables it is clear that $\int_{\mathbb{R}^2}\left(i\{A_{k-l} - A_{k+l}\}B^*_k\right)^* dk = \int_{\mathbb{R}^2} i\{A_{k'-l} - A_{k'+l}\}B^*_{k'}dk'$ and $\int_{\mathbb{R}^2}\left(\{A_{k-l} + A_{k+l}\}B^*_k\right)^* dk = \int_{\mathbb{R}^2}\{A_{k'-l} + A_{k'+l}\}B^*_{k'}dk'$.

□

Lemma 3. *If $A(x)$ and $B(x)$ are real scalar fields then $\int_{\mathbb{R}^2} A_{k+l}B^*_k\frac{dk}{2\pi} = \int_{\mathbb{R}^2} e^{-ix\cdot l}A(x)B(x)\frac{dx}{2\pi}$.*

Proof.

$$\begin{aligned}
\int_{\mathbb{R}^2} A_{k+l}B^*_k\frac{dk}{2\pi} &= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} e^{-x\cdot(k+l)}A(x)B^*_k\frac{dx}{2\pi}\frac{dk}{2\pi}\\
&= \int_{\mathbb{R}^2} e^{-ix\cdot l}A(x)\left[\int_{\mathbb{R}^2} e^{-x\cdot k}B^*_k\frac{dk}{2\pi}\right]\frac{dx}{2\pi}\\
&= \int_{\mathbb{R}^2} e^{-ix\cdot l}A(x)\left[\int_{\mathbb{R}^2} e^{x\cdot k}B_k\frac{dk}{2\pi}\right]^*\frac{dx}{2\pi}\\
&= \int_{\mathbb{R}^2} e^{-ix\cdot l}A(x)B^*(x)\frac{dx}{2\pi}\\
&= \int_{\mathbb{R}^2} e^{-ix\cdot l}A(x)B(x)\frac{dx}{2\pi}, \quad \text{since } B(x) \text{ is real.}
\end{aligned}$$

□