Gravitational wave and lensing inference from the CMB polarization

Ethan Anderes† (UC Davis Statistics Department)

Joint with Marius Millea (UC Berkeley) and Ben Wandelt (Institut d'Astrophysique de Paris)

† Research supported by: IHES-IHP CARMIN Fellowship, NSF DMS-1252795 and DMS-1812199

Back story

Gravitational waves and the CMB

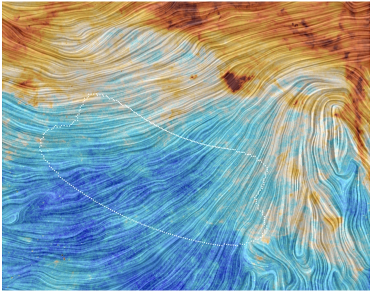
Gravitational Waves

- September 14, 2015 LIGO detected a gravitational wave as it passed by earth
- Big result for physics $\longrightarrow$ 2017 Nobel Prize
- Confirmed a prediction from Einstein's theory of relativity....
- ... also marked the beginning of gravitational wave astronomy,

  i.e. the probing of the universe through propagating distortions of space-time rather than just electromagnetic waves

Gravitational Waves

- At around the same time (2014) the BICEP team from the South Pole Telescope announced a detection of gravitational waves in the Cosmic Microwave Background (CMB)...
- ...which are predicted by a theory called cosmic inflation and imprint a specific signature on the polarization of the CMB photons



Gravitational Waves

- However the BICEP results was a false detection

> NATURE | NEWS
>
> **Gravitational waves discovery now officially dead**
>
> **Combined data from South Pole experiment BICEP2 and Planck probe point to Galactic dust as confounding signal.**

- The problem was insufficient statistical quantification of the emission from interstellar dust grains spinning in galactic magnetic fields

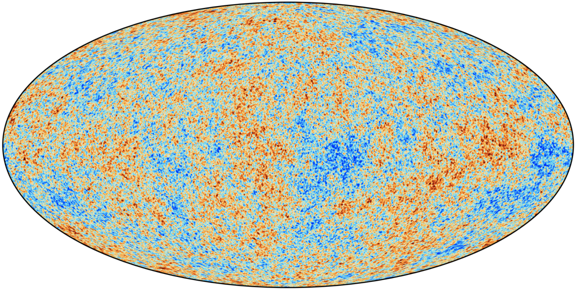
- So, the hunt is still on for the gravitational wave signatures in the CMB...
  ... is a major goal of the next generation Stage IV CMB experiments (planning underway, a projected $400M effort)

Gravitational waves and lensing of the CMB

... the basics

Cosmic Microwave Background

- The cosmic microwave background is a light that, for the most part, last interacted with matter only a few hundred thousand years after the big bang.
- Measuring the intensity of the CMB light as a function of position gives this (Planck 2015):



Cosmic Microwave Background

To give you a sense of the special nature of these observations ...

- It is basically the boundary of our observable universe
- We have highly accurate physical models from linear theory since it was generated so near the big bang
- Probes large relativistic scales and small quantum scales simultaneously
- Already it has been used to:
  - map the projected dark matter density fluctuations in our sky
  - determine that the mean curvature of space is much larger than the radius of the observable universe

- To get a handle on the problem of primordial gravitational wave detection, lets talk about a simplified flat-sky model of the CMB and the data
- In this setting the CMB polarization is characterized by a 2-d vector field

$$x \mapsto (Q(x), U(x))$$

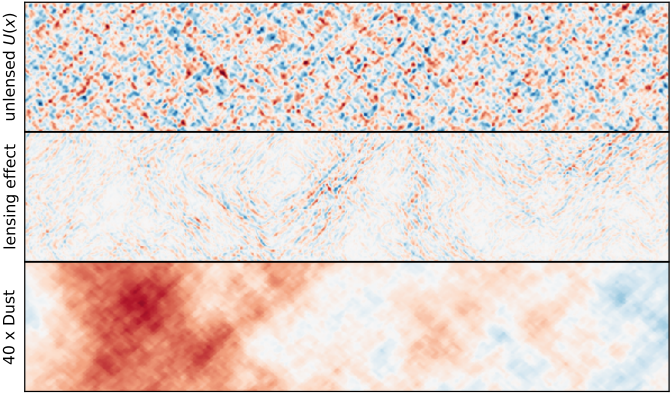
where x ranges over a compact region of $\mathbb{R}^2$
- $(Q(x), U(x))$ is a headless vector field, called spin 2

Simplified flat-sky data model for CMB polarization

$$
\begin{aligned}
d_q(x) &= Q(x + \nabla\phi(x)) \;+\; F_q(x) \;+\; N_q(x) \\
d_u(x) &= \underbrace{U(x + \nabla\phi(x))}_{\text{lensed polarization}} + \underbrace{F_u(x)}_{\text{foregrounds}} + \underbrace{N_u(x)}_{\text{noise}}
\end{aligned}
$$

- $N_q(x)$ and $N_u(x)$ denote instrumental noise
- $F_q(x)$ and $F_u(x)$ denote foreground emission from our own galaxy. E.g. emission from interstellar dust grains spinning in galactic magnetic fields
- $\phi(x)$ models the slight distortion of the CMB due to the gravitational influence of intervening matter (most of which is "dark matter") on the CMB light,
- This distortion is called "gravitational lensing"

Simulated $U(x)$ on a $\sim 0.3\%$ patch of the sky. The middle plot shows the lensing effect $U(x + \nabla\phi(x)) - U(x)$. The last plot shows a simulation of the foreground thermal emission from galactic dust.



Note: the dust emission is multiplied by a factor of 40 to make it visible on the same color scale.

The smoking gun of inflation

First consider a particular unitary linear transformation of (Q, U):

$$\begin{bmatrix} Q(x) \\ U(x) \end{bmatrix} \xrightarrow{FT} \begin{bmatrix} Q_k \\ U_k \end{bmatrix} \longrightarrow \begin{bmatrix} \cos(2\varphi_k) & -\sin(2\varphi_k) \\ \sin(2\varphi_k) & \cos(2\varphi_k) \end{bmatrix} \begin{bmatrix} Q_k \\ U_k \end{bmatrix} \xrightarrow{IFT} \begin{bmatrix} E(x) \\ B(x) \end{bmatrix}$$

- Analogous to divergence and curl of a vector field, but accounting for spin 2
- φ_k denotes the phase angle of frequency vector $k \in \mathbb{R}^2$
- The simplest models of inflation and the standard cosmological model predict that $E(x)$ and $B(x)$ are *isotropic Gaussian random fields*

The smoking gun of inflation

$$\begin{bmatrix} Q(x) \\ U(x) \end{bmatrix} \xrightarrow{FT} \begin{bmatrix} Q_k \\ U_k \end{bmatrix} \longrightarrow \begin{bmatrix} \cos(2\varphi_k) & -\sin(2\varphi_k) \\ \sin(2\varphi_k) & \cos(2\varphi_k) \end{bmatrix} \begin{bmatrix} Q_k \\ U_k \end{bmatrix} \xrightarrow{IFT} \begin{bmatrix} E(x) \\ B(x) \end{bmatrix}$$

- If cosmic inflation did not occur, and no primordial gravitational waves were produced, then $B(x)$ is predicted to be zero.

The smoking gun of inflation

$$\begin{bmatrix} Q(x) \\ U(x) \end{bmatrix} \xrightarrow{FT} \begin{bmatrix} Q_k \\ U_k \end{bmatrix} \longrightarrow \begin{bmatrix} \cos(2\varphi_k) & -\sin(2\varphi_k) \\ \sin(2\varphi_k) & \cos(2\varphi_k) \end{bmatrix} \begin{bmatrix} Q_k \\ U_k \end{bmatrix} \xrightarrow{IFT} \begin{bmatrix} E(x) \\ B(x) \end{bmatrix}$$

- If primordial gravitational waves were present, they distort space in such a way that induces non-zero $B(x)$ fluctuations
- Quantified by a single parameter: *tensor-to-scalar ratio r*
- Showing $r > 0$, i.e. $B(x)$ has non-zero fluctuations, is often termed *the smoking gun for inflation*

The smoking gun of inflation

Simplified flat-sky data model for CMB polarization

$$\begin{aligned} d_q(x) &= Q(x + \nabla\phi(x)) \;+\; F_q(x) \;+\; N_q(x) \\ d_u(x) &= \underbrace{U(x + \nabla\phi(x))}_{\text{lensed polarization}} \;+\; \underbrace{F_u(x)}_{\text{foregrounds}} \;+\; \underbrace{N_u(x)}_{\text{noise}} \end{aligned}$$

- The difficulty, to see this in the data, is that both lensing and **foregrounds** generate non-zero B fluctuations.

$$\begin{bmatrix} F_q(x) \\ F_u(x) \end{bmatrix} \xrightarrow{FT} \begin{bmatrix} F_{q,k} \\ F_{u,k} \end{bmatrix} \longrightarrow \begin{bmatrix} \cos(2\varphi_k) & -\sin(2\varphi_k) \\ \sin(2\varphi_k) & \cos(2\varphi_k) \end{bmatrix} \begin{bmatrix} F_{q,k} \\ F_{u,k} \end{bmatrix} \xrightarrow{IFT} \begin{bmatrix} * \\ {\color{red} B > 0} \end{bmatrix}$$

The smoking gun of inflation

Simplified flat-sky data model for CMB polarization

$$
\begin{aligned}
d_q(x) &= Q(x + \nabla\phi(x)) \;+\; F_q(x) \;+\; N_q(x)\\
d_u(x) &= \underbrace{U(x + \nabla\phi(x))}_{\text{lensed polarization}} \;+\; \underbrace{F_u(x)}_{\text{foregrounds}} \;+\; \underbrace{N_u(x)}_{\text{noise}}
\end{aligned}
$$

- The difficulty, to see this in the data, is that both **lensing** and foreground generate non-zero B fluctuations.

$$
\begin{bmatrix} \widetilde{Q}(x) \\ \widetilde{U}(x) \end{bmatrix} \xrightarrow{FT} \begin{bmatrix} \widetilde{Q}_k \\ \widetilde{U}_k \end{bmatrix} \longrightarrow \begin{bmatrix} \cos(2\varphi_k) & -\sin(2\varphi_k) \\ \sin(2\varphi_k) & \cos(2\varphi_k) \end{bmatrix} \begin{bmatrix} \widetilde{Q}_k \\ \widetilde{U}_k \end{bmatrix} \xrightarrow{IFT} \begin{bmatrix} * \\ B > 0 \end{bmatrix}
$$

where $\widetilde{Q}(x) = Q(x + \nabla\phi(x))$ and $\widetilde{U}(x) = U(x + \nabla\phi(x))$

- Even when (Q, U) has zero B fluctuations.

Field operator description of the data (no foregrounds)

$$d = \mathbb{A}\,\mathbb{L}(\phi)\,f + n$$

- Unlensed polarization field $f \sim GRF\big(0, \mathbb{C}^{ff}(r)\big)$ with covariance operator $\mathbb{C}^{ff}(r)$ which depends on the tensor-to-scalar ratio r
- Lensing potential $\phi \sim GRF\big(0, \mathbb{C}^{\phi\phi}\big)$ which operates on f in the QU basis via

$$\mathbb{L}(\phi)f(x) = f(x + \nabla\phi(x))$$

- Experimental noise $n \sim GRF\big(0, \mathbb{C}^{nn}\big)$
- Operator $\mathbb{A} = \mathbb{K}\,\mathbb{M}\,\mathbb{B}$ for beam $\mathbb{B}$, pixel space mask $\mathbb{M}$ and frequency cut $\mathbb{K}$

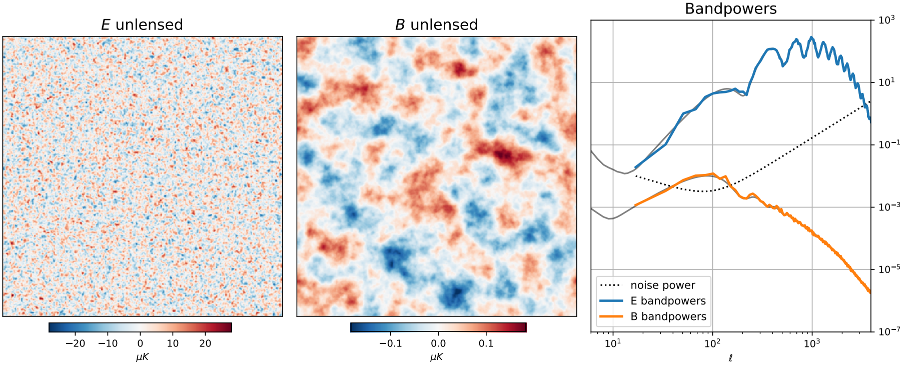


Figure: Unlensed polarization on 455 deg^2 patch of sky with $r = 0.025$. Note: r determines the amplitude of the unlensed B fluctuations. Dashed line on the right corresponds to $\sqrt{2}\ \mu$Karcmin QU noise with a knee at $\ell = 100$

$\mathbb{L}(\phi)\, f$

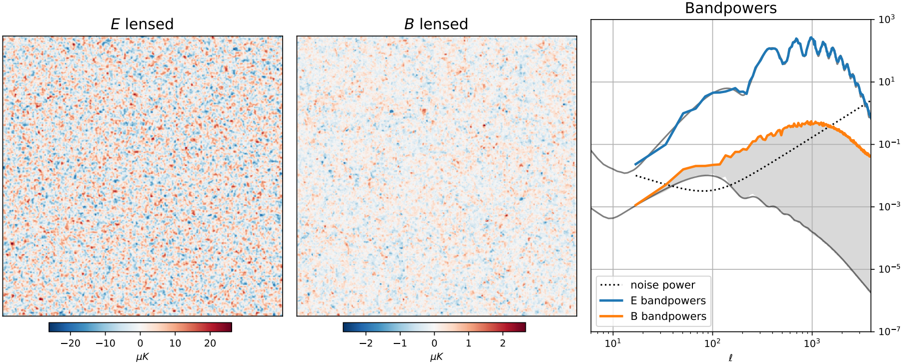


Figure: Lensed polarization. Qualitatively given by a phase distortion of E and a high frequency additive foreground corruption of B due to E fluctuations leaking into B fluctuations.

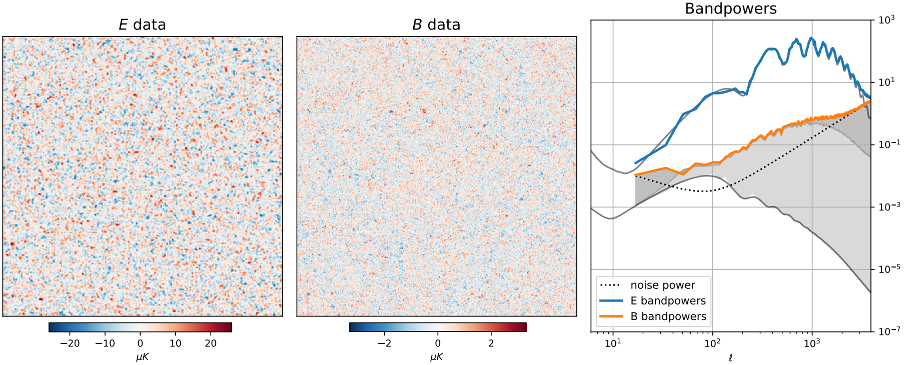


Figure: Here is what the data looks like *without* beam, masking or foreground emission. B is buried under lensing and noise corruption. However, since the main contribution of the lensing to B is from E leakage it seems possible one can estimate and remove a some of the lensing “noise”in B, a process called *delensing*.

Sampling the Bayesian Posterior

... on r, ϕ and f given d

The goal is to compute the posterior $r \mapsto P(r \,|\, d)$

- Formally
$$P(r|d) \propto P(d|r)\, P(r)$$
- Unfortunately, this form is basically intractable
- Requires computing $P(d|r) = P\big(\mathbb{A}\,\mathbb{L}(\phi)f + n \,\big|\, r\big)$ each field n, ϕ and f as random
- In this case $\mathbb{L}(\phi)f$ is isotropic but *non-Gaussian*
- Techniques for characterizing and working with non-Gaussian fields is currently extremely limited.
- A technique to get around this is to *disintegrate* by adding additional parameters to the posterior, then integrating them out.

The goal is to compute the posterior $r \mapsto P(r \,|\, d)$

- E.g. one can add ϕ as an unknown parameter.
- Trades non-Gaussianity in $\mathbb{L}(\phi)f$ for non-stationarity

$$P(r|d) \propto \int \underbrace{P(d|r,\phi)\, P(r)\, P(\phi)}_{P(r,\phi\,|\,d)/c}\, \mathrm{d}\phi$$

- $P(d|r,\phi)$ is much easier than $P(d|r)$

$$-2\log P(d|r,\phi) = \big\|d\big\|^2_{\mathbb{C}^{dd}(r,\phi)} + \log\det\left|\mathbb{C}^{dd}(r,\phi)\right|$$

- Determinant is the only tricky part

$$\mathbb{C}^{dd}(r,\phi) = \mathbb{A}\ \mathbb{L}(\phi)\ \mathbb{C}^{ff}(r)\ \mathbb{L}(\phi)^T \mathbb{A}^T + \mathbb{C}^{nn}$$

- See Hirata and Seljak (2003), Carron and Lewis (2017), Carron (2018)

The goal is to compute the posterior $r \mapsto P(r \,|\, d)$

- Additionally adding f as an unknown gives

$$P(r|d) \propto \iint \underbrace{P(d|r,\phi,f)\, P(\phi) P(f|r)\, P(r)}_{P(r,\phi,f\,|\,d)/c}\, \mathrm{d}\phi\, \mathrm{d}f$$

- Now just need $P(d|r,\phi,f)$ which is straight forward compared to $P(d|r,\phi)$ or $P(d|r)$

$$\begin{aligned}
-2\log P(d|r,\phi,f) &= \left\| d - \mathbb{A}\,\mathbb{L}(\phi) f \right\|^2_{\mathbb{C}^{nn}} \\
-2\log P(f|r) &= \left\| f \right\|^2_{\mathbb{C}^{ff}(r)} + \log\det\left|\mathbb{C}^{ff}(r)\right| \\
-2\log P(\phi) &= \left\| \phi \right\|^2_{\mathbb{C}^{\phi\phi}}
\end{aligned}$$

- No difficult determinants, covariance operators are easier
- Pushes all the difficulty into $\iint$ via sampling

$$\ldots, (r_i, \phi_i, f_i), \ldots \sim P(r,\phi,f|d)$$

Naive Gibbs sampler

Gibbs sampler

Initialize $f_0 = 0$, $\phi_0 = 0$, and r_0 to some upper bound

For $i = 1 \ldots n$

$$f_i \sim P(f \,|\, \phi_{i-1}, r_{i-1}, d) \quad \text{solved via CG}$$

$$\phi_i \sim P(\phi \,|\, r_{i-1}, f_i, d) \quad \text{HMC accept/reject}$$

$$r_i \sim P(r \,|\, f_i, \phi_i, d) \quad \text{evaluated on a grid}$$

- Easy to implement...
- ...but doesn't work!
- Theoretically it should, but the mixing time is **incredibly** slow

Naive Gibbs sampler

Gibbs sampler

Initialize $f_0 = 0$, $\phi_0 = 0$, and r_0 to some upper bound

For $i = 1 \ldots n$

$$f_i \sim P(f \,|\, \phi_{i-1}, r_{i-1}, d) \quad \text{solved via CG}$$

$$\phi_i \sim P(\phi \,|\, r_{i-1}, f_i, d) \quad \text{HMC accept/reject}$$

$$r_i \sim P(r \,|\, f_i, \phi_i, d) \quad \text{evaluated on a grid}$$

- The slow mixing is due to f and ϕ being highly correlated in the posterior

  An overdensity in $d(x)$ can explained by an overdensity in $f(x)$ with no lensing, or the lensing of a nearby overdensity.

Re-parameterization

- One way to overcome this difficulty is to re-parameterize the model

Original parameters: (f, ϕ)

Data written as

$$d = \mathbb{A}\,\mathbb{L}(\phi)\,f + n$$

New parameters: (f°, ϕ)

Data written as

$$d = \mathbb{A}\,\mathbb{L}(\phi)\mathbb{D}(r)^{-1}\mathbb{L}(\phi)^{-1}\,f^\circ + n$$

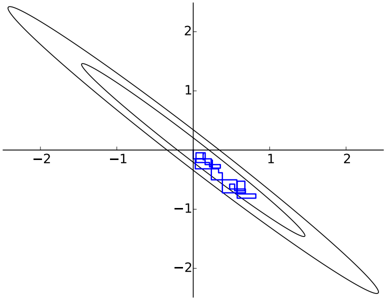
where

$$f^\circ := \mathbb{L}(\phi)\,\mathbb{D}(r)\,f$$

- This re-parameterization is basically a mix of ancillary vrs sufficient parameterization (see classic work by Gelfand, Roberts, Yu, Meng, etc...).

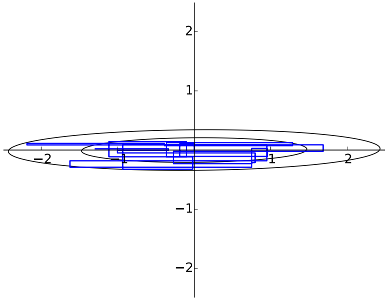
Here is the picture

$P(f, \phi \,|\, r, d)$ looks like this:



Slow mixing

$P(f^{\circ}, \phi \,|\, r, d)$ looks like this:



Fast mixing

Brief Pause

... outline: what we have done; the rest of the talk

Recap ...

- Intro to lensing, primordial gravitational wave and CMB
- Using parameter expansion and posterior marginalization to avoid non-Gaussian likelihoods and nasty determinants

$$P(r \mid d) = \int P(r, \phi \mid d)\, \mathrm{d}\phi = \iint P(r, \phi, f \mid d)\, \mathrm{d}\phi\, \mathrm{d}f$$

- Then need to re-parameterize so coordinates are more axes-aligned to make Gibbs tractable

$$f^\circ = \mathbb{L}(\phi)\mathbb{D}(r) f$$

The remainder of the talk ...

- More details for the re-parameterization

$$f^\circ = \mathbb{L}(\phi)\mathbb{D}(r)f$$

  ... need to define $\mathbb{D}(r)$ and explain why it is key that it is applied before $\mathbb{L}(\phi)$

- Why we had to invent LenseFlow, a custom dynamical systems algorithm for pixel-to-pixel lensing, to work with this re-parameterized posterior

- Finally some simulation examples

Why $\mathbb{L}(\phi)\mathbb{D}(r)$

...and why in that order

Naive parameterization

$$\phi \mapsto P(\phi \,|\, f, r, d)$$

- Want the conditioning variables as weakly informative for ϕ as possible
- Note that conditioning on f and d we are basically given a noisy version of $\mathbb{L}(\phi)f$ and f.
- Recovering ϕ is then a template matching problem:

  warp f till it looks like d

- Small range of ϕ values that matches template f to d

Lensed parameterization $\widetilde{f} = \mathbb{L}(\phi) f$

$$\phi \mapsto P(\phi \,|\, \widetilde{f}, r, \cancel{d})$$

- Warp estimation without a template
- More uncertainty for ϕ
- However, for CMB polarization the smallness of $B(x)$ means that "zero B" implicitly acts as a quasi-template:

unwarp lensed $\widetilde{Q}, \widetilde{U}$ till B is nearly zero

Mixed parameterization $f^\circ = \mathbb{L}(\phi)\mathbb{D}(r)f$

$$\phi \mapsto P(\phi \,|\, f^\circ, r, d)$$

- $\mathbb{D}(r) = \big[\widetilde{\mathbb{C}}^{ff}(r)\big]^{1/2}\big[\mathbb{C}^{ff}(r)\big]^{-1/2}$
- $\widetilde{\mathbb{C}}^{ff}(r)$ denotes the *lensed* spectrum
- $\mathbb{D}(r)$ rescales the nearly zero $B(x)$ quasi-template to have pre-lensed power that looks lensed
- Basically $\mathbb{D}(r)$ scrambles the quasi-template so its not as informative

LenseFlow

A custom algorithm for pixel-to-pixel lensing

Mixed parameterization log posterior

$$\begin{aligned}
&\log P\big(f^\circ, \phi,\, r, \,|\, d\big) + \text{constant} \\
&= \quad -\frac{1}{2}\big\|\, d - \mathbb{A}\,\mathbb{L}(\phi)\mathbb{D}(r)^{-1}\mathbb{L}(\phi)^{-1} f^\circ \,\big\|^2_{\mathbb{C}^{nn}} \\
&\qquad -\frac{1}{2}\big\|\, \mathbb{D}(r)^{-1}\mathbb{L}(\phi)^{-1} f^\circ \,\big\|^2_{\mathbb{C}^{ff}(r)} - \frac{1}{2}\log\big|\,\mathbb{D}(r)^2\mathbb{C}^{ff}(r)\big| \\
&\qquad + \log(P(r))
\end{aligned}$$

- Using transformation of variables formula for densities
- There should be an extra $\log\det(\mathbb{L}(\phi))$ term...
- ...these logdet terms can be very difficult to work with so if $\log\det(\mathbb{L}(\phi)) \neq 0$ it would present problems

Where is the missing $\log\det(\mathbb{L}(\phi))$

- Millea, EA, Wandelt (2017) developed a ODE characterization of pixel-to-pixel lensing, called LenseFlow where $\log\det(\mathbb{L}(\phi)) = 0$ is provably true
- Operators such as $\mathbb{L}(\phi)^{\dagger}$, $\mathbb{L}(\phi)^{-\dagger}$ and $\left[\frac{\partial}{\partial\phi}\mathbb{L}(\phi)^{-1}g\right]^{\dagger}$ are derived analytically via the ODE dynamics
- ... and also gives fast and exact delensing $\tilde{f} \mapsto \mathbb{L}(\phi)^{-1}\tilde{f}$ by running the ODE in reverse (still using forward lensing potential ϕ)
- The ODE for the posterior gradients turn out to be a special case of back-propagation

Defining LenseFlow

- Simply introduce an artificial time variable to the CMB field that connects $\tilde{f}$ at "time" 1 with f at "time" 0.

- In particular for $t \in [0, 1]$ let

$$f_t(x) = f(x + t\nabla\phi(x))$$

so that $f_0(x) = f(x)$ and $f_1(x) = \tilde{f}(x)$.

- Taking a time derivative and a spatial derivative gives

$$\begin{aligned} \frac{df_t(x)}{dt} &= \nabla^i f(x + t\nabla\phi(x)) \left[\nabla\phi(x)\right]^i \\ \nabla^i f_t(x) &= \nabla^j f(x + t\nabla\phi(x))[M_t(x)]^{ji} \end{aligned}$$

where $M_t(x) := \left[\delta^{ij} + t\nabla^i\nabla^j\phi(x)\right]$

- Re-arranging gives a ODE for the field f_t

Defining LenseFlow

LenseFlow

$$\dot{f}_t = (\nabla^j \phi)\,(M_t^{-1})^{ji}\,\nabla^i f_t$$

with initial conditions $f_0(x) \equiv f(x)$.

- Flowing the ODE forward gives $f \equiv f_0 \xrightarrow{t=0\to 1} f_1 = \mathbb{L}(\phi) f$.
- Flowing the ODE backward gives $\tilde{f} \equiv f_1 \xrightarrow{t=1\to 0} f_0 = \mathbb{L}(\phi)^{-1}\tilde{f}$
- **Note:** only ϕ is needed for backward flow ... i.e. one doesn't need to compute the inverse displacement for $\mathbb{L}(\phi)^{-1}$.
- **Note:** ∇^i is the only non-diagonal (in pixel space) operation needed.

LenseFlow uses an alternative expansion of the lensing effect

Comparing TayLense vrs LenseFlow expansions

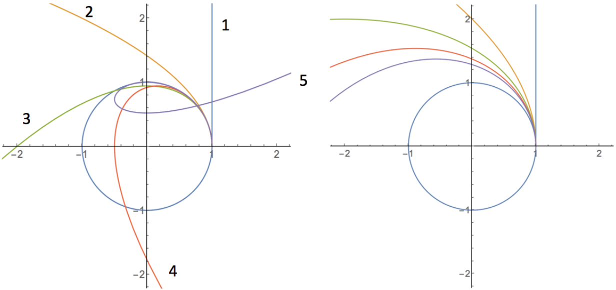
$$
\begin{aligned}
f(\boldsymbol{x} + \nabla\phi(\boldsymbol{x})) &\approx \left[\sum_{n=0}^{N} \frac{1}{n!}[\nabla\phi(\boldsymbol{x})]^n \nabla^n\right] f(\boldsymbol{x}) \\
&\approx \left[\prod_{n=0}^{N} \left(1 + \tfrac{1}{N} p_{n/N} \cdot \nabla\right)\right] f(\boldsymbol{x})
\end{aligned}
$$

- Both give exact results on finite pixels as $N \to \infty$ when ∇ is the true gradient (this is where sub-grid scale fluctuations come in)
- TayLense: N corresponds to Taylor order
- LenseFlow: $1/N$ corresponds to an ODE time step size
- For discrete pixels $\det(\mathit{LenseFlow}) \to 1$ as $1/N \to 0$ for any numerical ∇ such that $\nabla^\dagger = -\nabla$

LenseFlow uses an alternative expansion of the lensing effect

Comparing TayLense vrs LenseFlow expansions

$$
\begin{aligned}
f(\boldsymbol{x} + \nabla\phi(\boldsymbol{x})) &\approx \left[\sum_{n=0}^{N} \frac{1}{n!}[\nabla\phi(\boldsymbol{x})]^n \nabla^n\right] f(\boldsymbol{x}) \\
&\approx \left[\prod_{n=0}^{N} \left(1 + \tfrac{1}{N} p_{n/N} \cdot \nabla\right)\right] f(\boldsymbol{x})
\end{aligned}
$$



$\log\det(\mathbb{L}(\phi)) = 0$ with LenseFlow

- The LenseFlow ODE decomposes $\mathbb{L}(\phi)$ into infinitesimally small (local) linear operations

$$f_1 = \underbrace{\left[I + \varepsilon\; p_{t_n} \cdot \nabla\right] \cdots \left[I + \varepsilon\; p_{t_0} \cdot \nabla\right]}_{\xrightarrow{\epsilon \to 0} \mathbb{L}(\phi)} f_0$$

where $[p_t(x)]^j = \nabla^i \phi(x)\, [M_t^{-1}(x)]^{ij}$, $t_{i+1} = t_i + \varepsilon$ and $\epsilon = \frac{1}{n}$

- Since

$$\log\det\left[I + \varepsilon\; p_t \cdot \nabla\right] = \varepsilon\, \underbrace{\mathrm{Tr}\left[\, p_t \cdot \nabla\,\right]}_{=^* 0} + \mathcal{O}(\epsilon^2),$$

we have $\log\det(\mathbb{L}(\phi)) = 0$.

$*$ ... by the Hermitian anti-symmetry of ∇, i.e.

$$\mathrm{Tr}\left[diag(p^i)\nabla^i\right] = \mathrm{Tr}\left[(diag(p^i)\nabla^i)^\dagger\right] = \mathrm{Tr}\left[(\nabla^i)^\dagger diag(p^i)\right] = -\mathrm{Tr}\left[diag(p^i)\nabla^i\right].$$

Transpose lensing with LenseFlow

- Transpose (or adjoint) lensing $\mathbb{L}(\phi)^\dagger$ can be characterized with a ODE flow.
- Start by writing

$$\mathbb{L}(\phi)f = [I + \varepsilon\ p_{t_n} \cdot \nabla] \cdots [I + \varepsilon\ p_{t_0} \cdot \nabla] f$$

- Taking a formal adjoint

$$\begin{aligned}\mathbb{L}(\phi)^\dagger f &= [I + \varepsilon\ p_{t_0} \cdot \nabla]^\dagger \cdots [I + \varepsilon\ p_{t_n} \cdot \nabla]^\dagger f \\ &= \big[I - \varepsilon \nabla^i (p^i_{t_0} \bullet)\big] \cdots \big[I - \varepsilon \nabla^i (p^i_{t_n} \bullet)\big] f\end{aligned}$$

- Therefore $\mathbb{L}(\phi)^\dagger f = f_0$ where f_t satisfies the ODE

$$\dot{f}_t = \nabla^i (p^i_t\, f_t)$$

with initial conditions $f_1 = f$.

Lensing and inverse lensing accuracy

- 7 Runge-Kutta (4^{th} order) time steps produces accurate lensing (and inverse lensing).

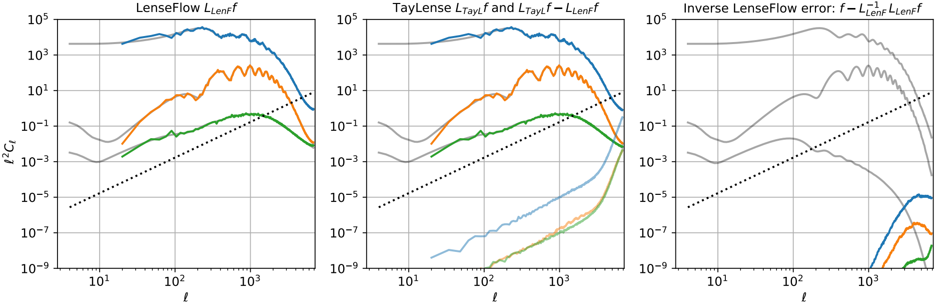


Figure: Simulation on a 1024x1024 flat sky periodic grid with pixel side length 1.5 arcmin. Timing $\sim$ 900 ms for a single lensing operation of a IQU field.

Fast posterior gradients

- Posterior gradients in re-parameterized coordinates given by

$$\nabla_{f^\circ,\phi} \log P = \begin{bmatrix} \mathbb{L}(\phi) & \frac{\partial}{\partial\phi}\mathbb{L}(\phi)\mathbb{D}(r)f \\ 0 & \mathbb{I} \end{bmatrix}^{-T} \begin{bmatrix} \mathbb{D}(r)^{-T} & 0 \\ 0 & \mathbb{I} \end{bmatrix} \underbrace{\nabla_{f,\phi} \log P}_{\text{orig parameters}}$$

- Everything can solved via the *adjoint ODE* of

$$\begin{bmatrix} \dot{\delta f_t} \\ \dot{\delta \phi_t} \end{bmatrix} = \begin{bmatrix} p_t^i \nabla^i & v_t^i \nabla^i - tW_t^{ij}\nabla^i\nabla^j \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \delta f_t \\ \delta \phi_t \end{bmatrix}$$

- p_t, v_t, and W_t can be pre-computed from an initial LenseFlow
- Note: these gradients account for the numerical implementation of ∇ used in LenseFlow

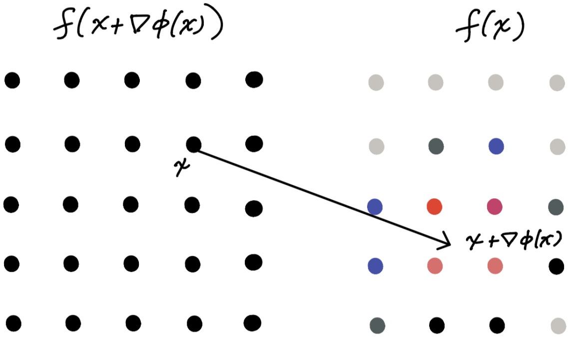
Intuition why

$$\mathbb{L}_{LenF}(\phi)f \approx \mathbb{L}_{TayL}(\phi)f$$

but...

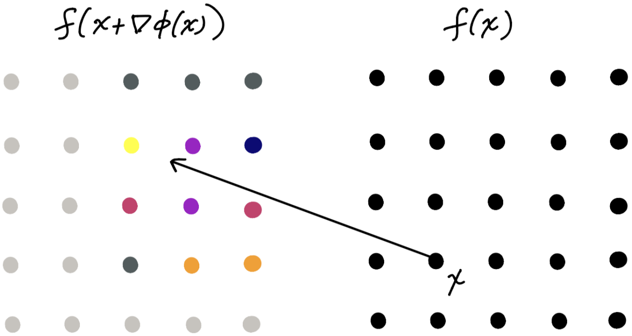
$$\log\det \mathbb{L}_{LenF}(\phi) \not\approx \log\det \mathbb{L}_{TayL}(\phi)$$

- Any pixel-to-pixel lensing will need to do some type of weighted averaging to account for sub-gridscale variability ...



- ... but there is flexibility in how one chooses these weights.

- These weights should be invertible and *perhaps* mostly local (for both forward and inverse lensing)



- For LenseFlow the forward lensing weights are local. This follows since

$$\mathbb{L}(\phi)f = [\,I + \varepsilon\; p_{t_n}\cdot \nabla\,]\cdots[\,I + \varepsilon\; p_{t_0}\cdot \nabla\,]\,f$$

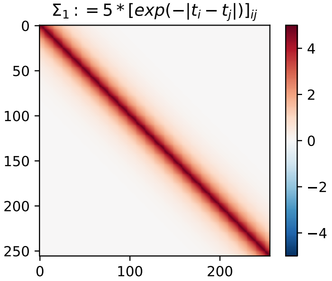
  and the operators $I + \varepsilon\; p_{t_i}\cdot \nabla$ are all localized.
- Inverse lensing weights are *also* local via the ODE time reversal

$$\mathbb{L}(\phi)^{-1}f = [\,I - \varepsilon\; p_{t_0}\cdot \nabla\,]\cdots[\,I - \varepsilon\; p_{t_n}\cdot \nabla\,]\,f$$

- For other alternatives, locality of both forward and inverse weights is not guaranteed.

Toy example: banded matrix with banded inverse

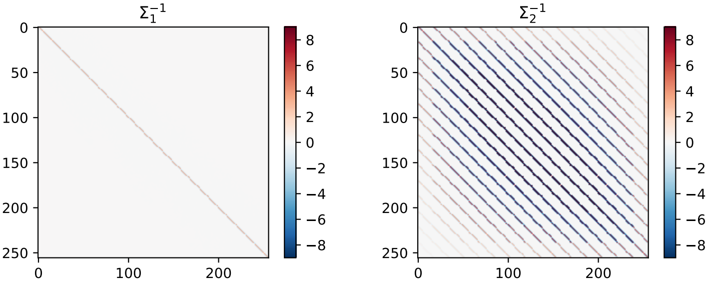
- The following two matrices have the same local linear behavior near the diagonal



- The left has exponential decay whereas the right truncates to zero

Toy example: banded matrix with banded inverse

- Here is the matrix inverse:



- The left is only non-zero on the diagonal and the near off-diagonal.... but the right has non-trivial weights spread across each row

First simulation example

CMB-S4 experimental conditions in the flat sky.

The (simulated) data

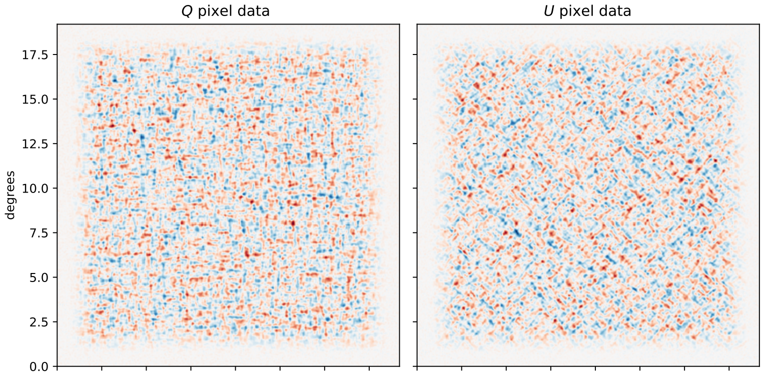


Figure: 384x384 Float32 flat sky pixels 3.0 arcmin pixels and beam (fsky 0.8936%, 368 deg^2). Noise at $\sqrt{2} - \mu$Karcmin with a knee at $\ell = 90$. $\ell_{min} = 30$ and $\ell_{max} = 2700$ which is 75% of nyquist limit.

The chain

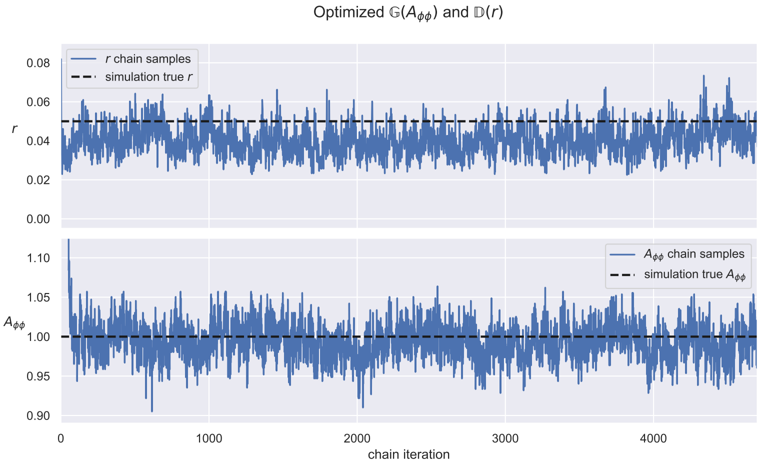


Figure: **Top**: r sample iterations. **Bottom**: $A_{\phi\phi}$ sample iterations. Chain started at fiducial $r = 0.1$ and $A_{\phi\phi} = 0.95$. Gibbs pass without $A_{\phi\phi}$ runs in 125 seconds. With $A_{\phi\phi}$ runs in 216 seconds

Marginal posterior density estimates

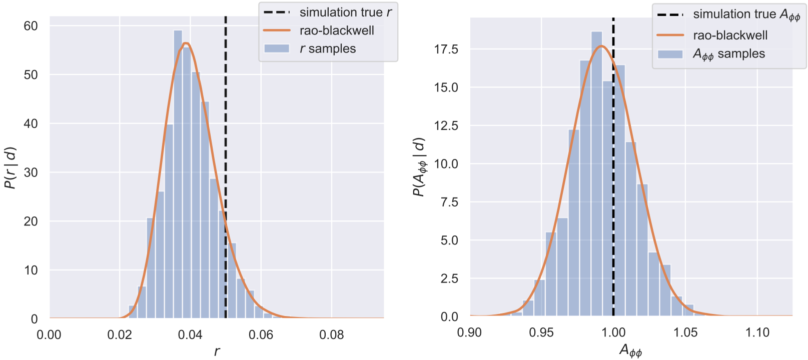


Figure: **Left**: r marginal samples. **Right**: $A_{\phi\phi}$ marginal samples

Joint posterior density estimates

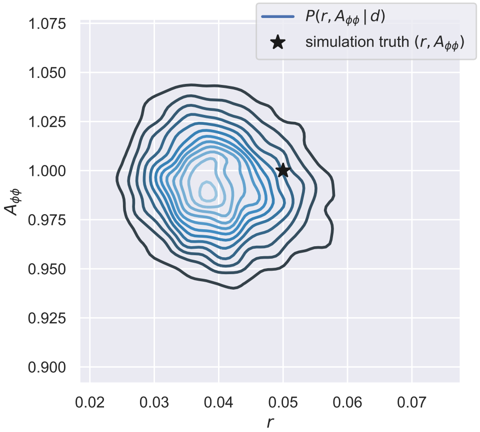


Figure: Joint $(r, A_{\phi\phi})$ samples. Suggests we can fix a fiducial $A_{\phi\phi}$ without severely effecting the r samples.

Showing the improvement gained by $\mathbb{G}(A_{\phi\phi})$

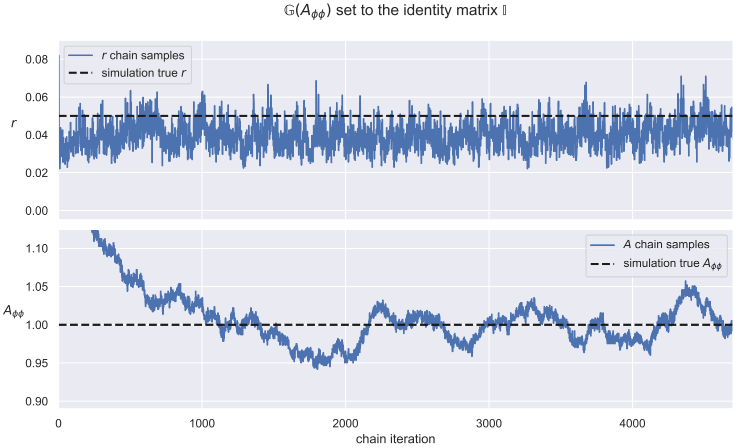
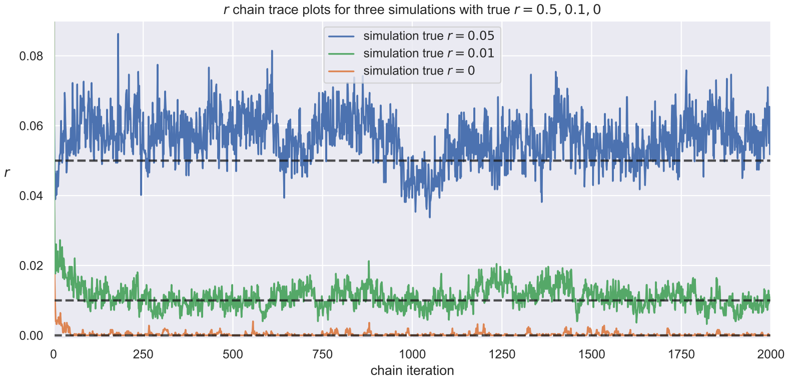


Figure: Using the same seed but with $\mathbb{G}(A_{\phi\phi}) = I$, the identity, so that $\phi^\circ = \phi$

Second simulation example

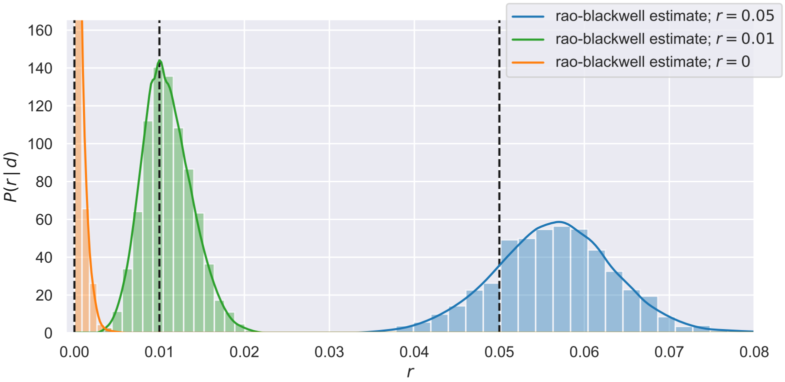
Fixed fiducial $A_{\phi\phi}$. fsky $= 1.59\%$ (655 deg^2).
Three simulation truths $r = 0.05,\ 0.01,\ 0$.

The chains

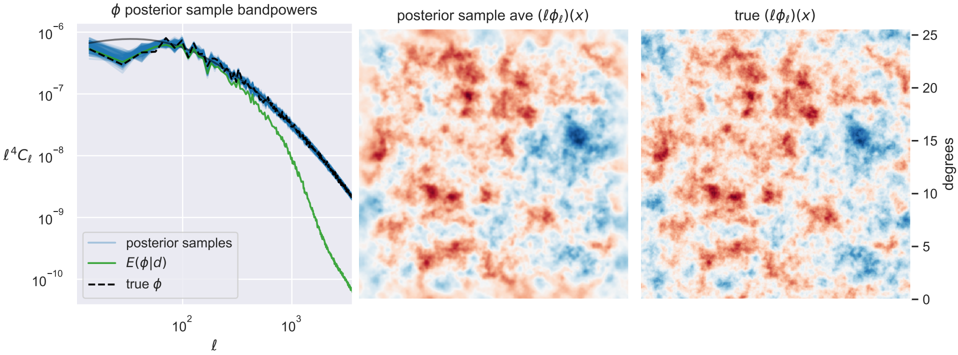


r chain trace plots for three simulations with true $r = 0.5, 0.1, 0$
simulation true $r = 0.05$
simulation true $r = 0.01$
simulation true $r = 0$
r
0.00
0.02
0.04
0.06
0.08
0
250
500
750
1000
1250
1500
1750
2000
chain iteration

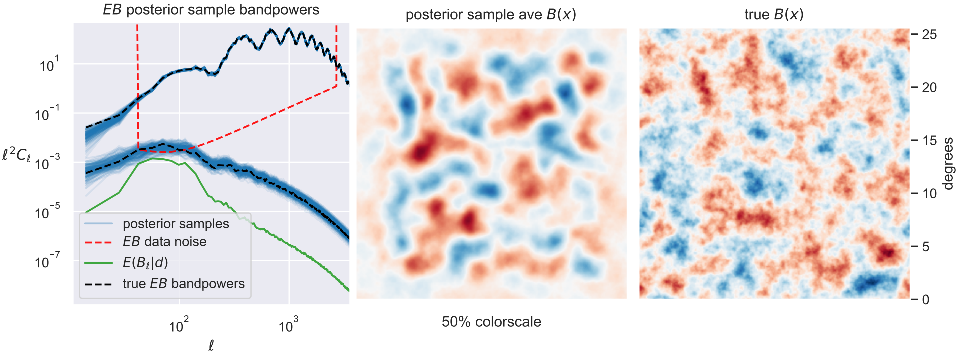
Marginal posterior density estimates



ϕ posterior samples



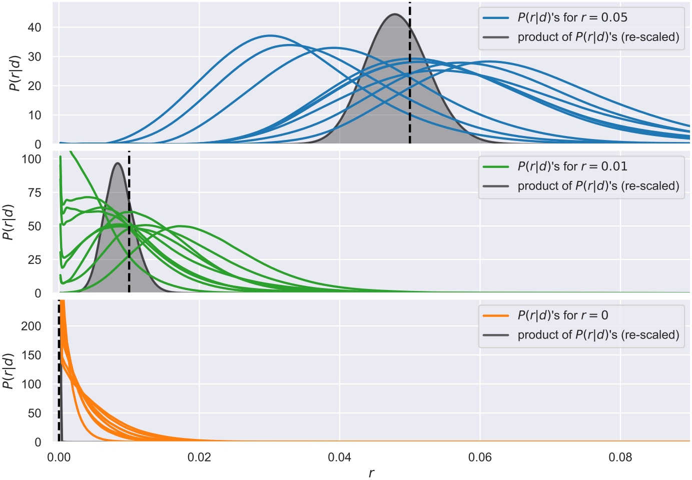
Unlensed B samples



50% colorscale

Third simulation example

Coverage sanity check



- Multiple independent samples (d_i, ϕ_i, f_i, n_i) on small 256x256 pixel grids (fsky $= 0.3972\%$, 164 deg^2) for $r = 0.05,\ 0.01,\ 0$.
- Compute the product $\Pi_i P(r \mid d_i)$. Check that it supports the simulation truth r.

Some questions

- Is there a way to make sense of the continuum version of $\log\det\mathbb{L}(\phi)$?
  - *Is it possible that continuum* $\log\det\mathbb{L}(\phi)\neq 0$ *and pixel-to-pixel* LENSEFLOW *is just adjusting near Nyquist frequency modes to get* $\log\det = 0$*?*
  - *Are there splitting methods that can guarantee the discrete ODE solvers for* LENSEFLOW *satisfy* $\log\det\mathbb{L}(\phi) = 0$ *exactly and works for strong lenses?*
- LENSEFLOW ODE formalism for the sphere and other geometries?
  - *For spin 2 vector fields the lensing displacement should parallel transport the vectors*
  - *Replacing* ∇ *with discrete pixel space differences. Needs to work on HealPix and non-uniform grids*
- Can one use Lie Group to make sense of these re-parameterizations for general Gibbs samplers ?