

Efficient Optimal Reconstruction of Linear Fields and Band-powers from Cosmological Data

B. Horowitz^{a,b} U. Seljak^{a,b} G. Aslanyan^a

^aBerkeley Center for Cosmological Physics and Department of Physics, University of California, Berkeley, CA 94720

^bLawrence Berkeley National Laboratory, Berkeley, CA 94720

E-mail: bhorowitz@berkeley.edu, useljak@berkeley.edu

Abstract. We present an efficient implementation of Wiener filtering of real-space linear field and optimal quadratic estimator of its power spectrum Band-powers. We first recast the field reconstruction into an optimization problem, which we solve using quasi-Newton optimization. We then recast the power spectrum estimation into the field marginalization problem, from which we obtain an expression that depends on the field reconstruction solution and a determinant term. We develop a novel simulation based method for the latter. We extend the simulations formalism to provide the covariance matrix for the power spectrum. We develop a flexible framework that can be used on a variety of cosmological fields and present results for a variety of test cases, using simulated examples of projected density fields, projected shear maps from galaxy lensing, and observed Cosmic Microwave Background (CMB) temperature anisotropies, with a wide range of map incompleteness and variable noise. For smaller cases where direct numerical inversion is possible, we show that our solution matches that created by direct Wiener Filtering at a fraction of the overall computation cost. Even more significant reduction of computational is achieved by this implementation of optimal quadratic estimator due to the fast evaluation of the Hessian matrix. This technique allows for accurate map and power spectrum reconstruction with complex masks and nontrivial noise properties.

Keywords: large scale structure – power spectrum, large scale structure – gravitational lensing, large scale structure – galaxy clusters

Contents

1	Introduction	1
2	Background	2
2.1	Bandpowers posterior	3
2.2	MAP Field Reconstruction	4
2.3	Minimum Variance Estimation of the Power Spectrum	4
2.4	Estimation of the Noise Bias and the Hessian	6
2.5	Procedure Summary	7
3	Example Cases	7
3.1	Projected Density Field	7
3.2	Wiener Filtering vs. MAP Projected Density Example	10
3.3	Cosmic Microwave Background Temperature	12
3.4	Cosmic Shear	14
4	Discussion and Conclusion	16
A	Wiener Filter Review	18
B	Convergence Criteria of CMB Reconstruction	19
C	Joint E & B Cosmic Shear Reconstruction	21

1 Introduction

A ubiquitous problem of modern astrophysics is the reconstruction of the underlying signal from observed, noisy, and incomplete data. For linear fields the Wiener filter [1, 2] is the gold standard for reconstructing the underlying signal, as it is “optimal” in the sense that it minimizes the variance. It has been used as the basis of cosmological analysis for both large scale structure [3, 4] and CMB [5, 6].

However, the Wiener filter requires one to take the inverse of the overall covariance matrix, which has a noise and a signal based components. Noise is typically diagonal in observed space, while signal is diagonal in harmonic (or Fourier) space. In general, the covariance matrix will not be diagonalizable in either basis and it will be computationally difficult to invert the matrix numerically for a realistic sized survey map. While it is possible to make simplifying assumptions, like homogeneous and isotropic noise (as in, for example [7]), it is possible to efficiently implement Wiener filter by using the well studied property that the Wiener filter is mathematically equivalent to maximum a posteriori (MAP), which in turn can be solved with fast linear algebra methods or optimization [8].

This optimization can be performed using a variety of numerical techniques. Recent work used a messenger [9, 10] (or dual messenger [11]) field which can be diagonalized in either basis in which to run an approximation scheme. These methods have been argued to perform well versus other approximation schemes, but there are situations where the messenger field is zero (such as zero noise field with mask) and the method fails. In addition, it has been argued that a suitably chosen preconditioned conjugate gradient technique might be faster in some instances

[12]. Preconditioned conjugate gradient techniques have performed well in the context of CMB map reconstruction [13, 14], but require careful selection of the preconditioning scheme to achieve fast convergence. We therefore want to explore implementations that are both general and computationally efficient.

For cosmological analysis, it is not only the field that is of interest but also the band-powers (such as power spectrum amplitudes) and their covariance matrix. Work towards estimating these quantities jointly with the underlying field has been done in the cosmic shear context [10, 15], but it required sampling from the joint probability distribution which is computationally involved. In this work we assume flat prior on band-powers and hence examine maximum likelihood estimation (MLE) method, after marginalization over the field, to quickly compute these quantities for observations with complex noise and mask properties as well.

In particular, we examine three cases of cosmological interest: linear density reconstruction, cosmic shear (E mode estimation alone as well as joint E/B), and primary CMB temperature anisotropy reconstruction. The outline of the paper is as follows: we discuss our MAP/MLE for field and power spectrum estimation in §2. In §3, we use these tools in a number of contexts, and compare the optimization approach with a numerically exact case in Subsection 3.2. In §4 we discuss our results and possible extensions of the work to analyze real data.

In Appendix A we review the exact Weiner Filter approach which relies on inversion of the full covariance matrix. In Appendix B we discuss the convergence criteria in the case of primary CMB reconstruction. In Appendix C we discuss the joint estimation of E and B fields in the context of cosmic shear (although the technique transfers directly to CMB E and B polarization reconstruction as well).

2 Background

Here we summarize the optimization technique and standardize the notation. For a more thorough description, see [8, 16–18]. We measure quantities $d(\mathbf{r}_i)$ at select positions, such as a reconstructed projected density, forming a full data vector, $\mathbf{d}$. In general, this data vector will depend on a combination of underlying information about the field (“field coefficients”) we wish to estimate, $\mathbf{s}$, and that which is pure noise, $\mathbf{n}$, i.e.

$$\mathbf{d} = \mathbf{R}\mathbf{s} + \mathbf{n}, \quad (2.1)$$

where the $\mathbf{R}$ is the response matrix expressing how our measurement changes with the underlying information. We express the underlying two point function information in terms of covariance matrices, $\mathbf{S} = \langle \mathbf{s}\mathbf{s}^\dagger \rangle$, and $\mathbf{N} = \langle \mathbf{n}\mathbf{n}^\dagger \rangle$, for the estimated signal and noise components, respectively. We assume that these are uncorrelated with each other, i.e. $\langle \mathbf{n}\mathbf{s}^\dagger \rangle = 0$; changes to this assumption are tractable but would require a redefinition of our underlying likelihood function and complicate the analysis since the noise would carry signal information. The correlation matrix of the data is therefore,

$$\begin{aligned} \langle \mathbf{d}\mathbf{d}^\dagger \rangle &\equiv \mathbf{C} = \langle (\mathbf{R}\mathbf{s} + \mathbf{n})(\mathbf{R}\mathbf{s} + \mathbf{n})^\dagger \rangle \\ &= \langle (\mathbf{R}\mathbf{s}(\mathbf{R}\mathbf{s})^\dagger + \mathbf{n}\mathbf{n}^\dagger + \text{Cross Terms}) \rangle \\ &= \mathbf{R}\mathbf{S}\mathbf{R}^\dagger + \mathbf{N}. \end{aligned} \quad (2.2)$$

Our signal covariance, $\mathbf{S}$, takes a diagonal form in Fourier space. The noise covariance, $\mathbf{N}$, is often approximately diagonal in data space, as there usually are no noise correlations between

elements of the data vector. We can convert the covariance matrix into Fourier basis where signal covariance is diagonal, but this will lead to very strong off-diagonal terms of the noise matrix, in the presence of masked sky regions and/or variable noise properties. Therefore, our overall data correlation matrix cannot be diagonalized easily in either basis.

We can re-express our covariance in terms of underlying band-powers, Θ , labeling each modes of interest to estimate as $\{1, \dots, l, \dots, l_{\max}\}$, and the rest as $\{l_{\max}+1, \dots, \infty\}$, and then expressing our correlation matrix as

$$\mathbf{C} = \sum_l \Theta_l \mathbf{Q}_l + \mathbf{N}. \quad (2.3)$$

This new $\mathbf{Q}_l = \mathbf{\Pi}_l \mathbf{R} \mathbf{R}^\dagger \mathbf{\Pi}_l$ basis is the projection (with projection operators $\mathbf{\Pi}_l$) of the response matrix $\mathbf{R}$ for each mode. The band-power can correspond to averaging over spherical harmonic basis coefficients at a constant l , adding up $2l+1$ m modes, but we can also average over more than one l .

2.1 Bandpowers posterior

We now want to find the most probable set of bandpowers for a given set of measurements $\mathbf{d}$, assuming flat prior on band-powers. We thus parametrize the power spectrum as a function of these bandpowers $\mathcal{S}(\Theta)$. If we assume our modes are Gaussian we can express the likelihood function in the familiar form, i.e.

$$L(\mathbf{d}|\Theta) = (2\pi)^{-N/2} \det(\mathbf{C})^{-1/2} \exp\left(-\frac{1}{2} \mathbf{d}^\dagger \mathbf{C}^{-1} \mathbf{d}\right). \quad (2.4)$$

Associated with the likelihood function and a parameter set $\hat{\Theta}$ which maximizes it, is the Hessian matrix [19],

$$F_{ll'} = -\frac{\partial^2 \ln L}{\partial \Theta_l \partial \Theta_{l'}}. \quad (2.5)$$

The inverse of the Hessian matrix can be interpreted as a local estimate of the covariance matrix of the parameters, i.e.

$$\mathbf{F}^{-1} = \langle \Theta \Theta^\dagger \rangle - \langle \Theta \rangle \langle \Theta \rangle^\dagger. \quad (2.6)$$

We now have the bandpower posterior in the Gaussian form, given by the mean $\hat{\Theta}$ and the covariance matrix $\mathbf{F}^{-1}$. To obtain the solution for the mean it is easiest to use Newton's second order method, which gives a quadratic estimator of the form [20]

$$\Theta_l = \frac{1}{2} \sum_{l'} F_{ll'}^{-1} (\mathbf{d}^\dagger \mathbf{C}^{-1} \mathbf{Q}_{l'} \mathbf{C}^{-1} \mathbf{d} - b_{l'}), \quad (2.7)$$

where b_l is a noise bias term that can be found by computing the ensemble average of the first term assuming $\theta_l = 0$ for all modes probed (i.e. $l < l_{\max}$),

$$b_l = \text{tr} \left[\mathbf{N} + \sum_{l_{\max}+1}^{\infty} (\Theta_l \mathbf{Q}_l) \mathbf{C}^{-1} \mathbf{Q}_l \mathbf{C}^{-1} \right]. \quad (2.8)$$

This is an implicit equation since $\mathbf{C}$ depends on Θ_l , and needs iterations, as discussed further below. We refer below the power spectrum inside $\mathbf{C}$ as $\mathcal{S}_{\text{fid}}$.

2.2 MAP Field Reconstruction

In practice this analytical calculation requires the inversion of a large matrix, $\mathbf{C}$, which does not necessarily have properties that make inversion efficient (i.e. block diagonal or sparse) and will in general require $O(n^3)$ time for an $n \times n$ matrix. In the case of reconstructing the underlying density field for astronomical large surveys with n pixels, this would be prohibitively computationally expensive for the foreseeable future. Instead, we will approach this as an optimization problem [18]. We will not use 2.4, and instead of expressing the likelihood of data given bandpowers we will work in terms of latent variables, writing the joint distribution of $\mathbf{s}$ and $\mathbf{d}$,

$$p(\mathbf{s}, \mathbf{d} | \mathbf{S}) = (2\pi)^{-(N+M)/2} \det(\mathbf{S}\mathbf{N})^{-1/2} \exp\left(-\frac{1}{2} \mathbf{s}^\dagger \mathbf{S}^{-1} \mathbf{s} + (\mathbf{d} - \mathbf{R}\mathbf{s})^\dagger \mathbf{N}^{-1} (\mathbf{d} - \mathbf{R}\mathbf{s})\right), \quad (2.9)$$

and note that the minimum variance solution for the modes can be found by minimizing the loss function χ^2 ,

$$\chi^2 = -2 \ln p(\mathbf{s}, \mathbf{d} | \mathbf{S}) + c = \mathbf{s}^\dagger \mathbf{S}^{-1} \mathbf{s} + (\mathbf{d} - \mathbf{R}\mathbf{s})^\dagger \mathbf{N}^{-1} (\mathbf{d} - \mathbf{R}\mathbf{s}), \quad (2.10)$$

with respect to $\mathbf{s}$. Taylor expanding around $\mathbf{s}_m$ to second order, we have

$$\chi^2 = \chi^2(\mathbf{s}_m) + 2\mathbf{g}(\mathbf{s} - \mathbf{s}_m) + (\mathbf{s} - \mathbf{s}_m)^\dagger \mathbf{D}(\mathbf{s} - \mathbf{s}_m), \quad (2.11)$$

with gradient function in terms of the derivative of the response function, $\mathbf{R}$, given as

$$\mathbf{g} = \frac{1}{2} \frac{\partial \chi^2}{\partial \mathbf{s}} = \mathbf{S}^{-1} \mathbf{s}_m - \mathbf{R}^\dagger \mathbf{N}^{-1} (\mathbf{d} - \mathbf{R}\mathbf{s}_m). \quad (2.12)$$

For the linear problems studied in this work, this derivative can be calculated analytically, but in other more involved cases (such as nonlinear structure formation) might be computationally involved as it would require intensive back-propagation. The solution where $\mathbf{g} = 0$, and therefore a local extremum is found, will be denoted $\hat{\mathbf{s}}$, and is the maximum a posteriori solution (MAP). For linear problems it is the best possible solution in the sense to minimizing the variance.

The curvature matrix $\mathbf{D}$ has the form

$$\mathbf{D} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \mathbf{s} \partial \mathbf{s}} = \mathbf{S}^{-1} + \mathbf{R}^\dagger \mathbf{N}^{-1} \mathbf{R}. \quad (2.13)$$

However, in this work we will not explicitly evaluate it, as it is too large. Instead, we will use low rank approximation as performed by L-BFGS quasi-Newton optimization method. We will use L-BFGS as the optimization method in this paper.

The starting point for the optimization algorithm does not play a significant role for linear problems as the posterior surface is convex and the true global minimum can always be found. In practice, for the examples in this work, we found no noticeable effects of the starting point in terms of convergence properties, i.e. required number of iterations.

2.3 Minimum Variance Estimation of the Power Spectrum

The result of the above optimization procedure is $\hat{\mathbf{s}}$, and is useful for creating maps, but has more information than needed for cosmological analysis. If our goal is to determine a set of summary statistics/band-power measurement, Θ , such as a power spectrum bandpowers,

we need to marginalize over the latent variables, the modes $\mathbf{s}$. To do so we need to define a projection matrix $\mathbf{\Pi}_l$ around a fiducial power-spectrum $\mathbf{S}_{\text{fid}}$ with associated band-powers Θ_{fid} , defined as

$$\left[\frac{\partial \mathbf{S}}{\partial \Theta_l} \right]_{\mathbf{S}_{\text{fid}}} = \mathbf{\Pi}_l. \quad (2.14)$$

This fiducial power spectrum is a regularized version of the measured power spectrum, and is thus iterated upon: we start with some fiducial prior, which we then update if the data require us to do so. This process is regularized, i.e. we use a smooth version of the measured power spectrum, for example a power spectrum predicted by the cosmological parameters we are determining from these data.

The true covariance can be written in terms of the projection operators:

$$\mathbf{S} = \sum_l \Theta_l \mathbf{\Pi}_l, \quad (2.15)$$

where $\Delta \Theta_l$ is the difference of the band-powers to those of the fiducial model. For the cases studied in this work, the dependence of $\mathbf{S}$ on Θ is linear so we can take

$$\mathbf{\Pi}_l = \frac{\mathbf{S}_{\text{fid}}}{\Theta_l}, \quad (2.16)$$

i.e. the projection matrix takes the power spectrum per bin, Θ_l , to the full power spectrum, $\mathbf{S}$. Note that the choice of the fiducial model is important in that if it is sufficiently far away from the true model the result could be biased, but iteratively recalculating $\mathbf{S}_{\text{fid}}$ with the solved new band-powers Θ will provide an asymptotically more accurate reconstruction. In the cases of interest in this work, a single iteration was sufficient to provide an accurate reconstruction. In practice for examples in this work, we used the true power-spectrum with each power re-scaled by a random value between 0.01 and 1.0; however we tested various other schemes which all provide accurate reconstructions as long as no band-power was set identically to zero.

We are assuming flat prior for the bandpowers, so to compute the posterior distribution of band-powers we can write their (marginalized over $\mathbf{s}$) likelihood function to maximize as a second order expansion around the fiducial model

$$\ln L(\Theta_{\text{fid}} + \Delta \Theta) = \ln L(\Theta_{\text{fid}}) + \sum_l \left[\frac{\partial \ln L(\Theta)}{\partial \Theta_l} \right]_{\Theta_{\text{fid}}} \Delta \Theta_l + \frac{1}{2} \sum_{ll'} \left[\frac{\partial^2 \ln L(\Theta)}{\partial \Theta_l \partial \Theta_{l'}} \right]_{\Theta_{\text{fid}}} \Delta \Theta_l \Delta \Theta_{l'}; \quad (2.17)$$

where we assume a flat prior on the band-powers.

We define

$$E_l(\mathbf{S}_{\text{fid}}, \hat{\mathbf{s}}) = \frac{1}{2} \hat{\mathbf{s}}^\dagger \mathbf{S}_{\text{fid}}^{-1} \mathbf{\Pi}_l \mathbf{S}_{\text{fid}}^{-1} \hat{\mathbf{s}} = \frac{1}{2} \sum_{k_l} \frac{\hat{s}_{k_l}^2}{\Theta_{\text{fid},l} S_{\text{fid},k_l}}, \quad (2.18)$$

where in the last expression we define the sum over k_l as the sum over all modes which contribute to band-power, Θ_l , and in the last equality we made use of the diagonal property of the projection operators and fiducial power spectrum. Putting this together we find that the derivative of the likelihood can be expressed as [18]

$$\frac{\partial \ln L(\Theta)}{\partial \Theta_l} = E_l - b_l, \quad (2.19)$$

where we defined

$$b_l = \frac{1}{2} \text{tr} \left[\frac{\partial \det \ln(\mathbf{S}\mathbf{N})}{\partial \Theta_l} \right]_{\mathbf{S}_{\text{fid}}}. \quad (2.20)$$

For the linear cases studied in this work, this term is often called the noise bias term. However, it is worth remembering that this term's origin is the derivative of the log determinant of the product of the Hessian and the signal covariance matrices in equation 2.20 (since noise covariance derivative is zero). To find MLE we need to find the zero of Eq. 2.19, which we solve using Newton's method. To do this we define the Hessian matrix,

$$F_{ll'} = -\frac{\partial^2 \ln L(\boldsymbol{\Theta})}{\partial \Theta_l \partial \Theta_{l'}}, \quad (2.21)$$

which for linear models defines the Gaussian posterior assuming sufficient modes have been averaged over so that by central limit theorem we can describe the posterior as a multivariate gaussian. The peak of the likelihood function can be found by setting the derivative of equation 2.17 with respect to $\Delta \boldsymbol{\Theta}$ to zero, which upon inserting equation 2.19 yields

$$(\mathbf{F} \Delta \hat{\boldsymbol{\Theta}})_l = E_l - b_l. \quad (2.22)$$

2.4 Estimation of the Noise Bias and the Hessian

While the noise bias, b_l , and the Hessian matrix, $\mathbf{F}$, from Equation 2.22 could be calculated exactly, this will involve inversion of large matrices, which is what we are trying to avoid by deriving the MAP via optimization techniques. Instead, we will perform a simulation based analysis motivated by the underlying definition of each of these terms.

In general, the maximum likelihood field, $\hat{\mathbf{s}}$, attained with the procedure described in Sec 2.2, will have bias due to the presence of noise: when the noise is high the minimum variance estimator drives $\mathbf{s}$ to zero. In the case of cosmological density fields which have red power spectra (less power on small scales compared to white noise), this will result in washing out the small scale power. See the figures in Sec 3 for explicit examples.

To correct for this bias we need to understand how our reconstruction responds to the presence of noise. For this we perform a simulation analysis wherein we generate a data realization generated from a fiducial power spectrum, inject the noise and mask, perform the optimization and see how the presence of noise affects the reconstruction. Let us call the new data and noise realization data $\mathbf{d}_{s+n}$, with associated maximum likelihood reconstruction $\hat{\mathbf{s}}_{s+n}$. The gradient of equation 2.19 has to vanish if evaluated at the fiducial model. The noise bias in this case can be found directly as

$$b_l = E_l(\boldsymbol{\Theta}_{\text{fid}}, \hat{\mathbf{s}}_{s+n}). \quad (2.23)$$

This quantity should be averaged over many realizations, but for the linear signal-dominated cases studied in this work we found even one realization was sufficient for an accurate reconstruction.

To calculate the Hessian matrix, we evaluate the gradient of equation 2.19 at two different fiducial model values, and use finite differentiation [18],

$$F_{ll'} \Delta \Theta_{l'} = E_l(\boldsymbol{\Theta}_{\text{fid}} + \Delta \Theta_{l'}) - E_l(\boldsymbol{\Theta}_{\text{fid}}). \quad (2.24)$$

Its inverse is the covariance matrix for the band-powers. This is in contrast to directly using linear algebra techniques to calculate the Hessian matrix (see Equation A.5) which would be

numerically intractable for a realistic survey size. Using Equation 2.24 in the linear case, one can calculate it at the cost of additional optimization step. Since this is a linear problem Hessian matrix equals Fisher information matrix and thus gives the smallest attainable errors on the parameters (Cramér-Rao theorem).[21]

2.5 Procedure Summary

1. Initialize a Gaussian random field (the true signal field) with some underlying power-spectrum.
2. Apply the response operator to this field, and additional noise and masking terms. The output of this is the input data vector, $\mathbf{d}$.
3. An estimate of the underlying signal field is created through optimization as described in Subsection 2.2 to yield $\hat{\mathbf{s}}$.
4. An initial estimate of the band-powers is generated by taking the power spectra of the reconstructed map and binning.
5. To this band-power estimate, we apply the noise bias correction (estimated using Eq 2.23) and Hessian matrix (estimated using Eq 2.24) to provide an optimal reconstructed value of the band-powers given in Eq 2.22.

3 Example Cases

Here we implement the above scheme in a number of simulated cosmological contexts to demonstrate its versatility and efficiency. For these cases, we set our convergence criteria to be $\epsilon \equiv \delta\chi^2 = 10^{-1}$; i.e. the optimization ends when the difference of the absolute chi-square values between iterations is 10^{-1} (typical value of χ^2 is of order 2×10^5 for the dimensionality used here). We note that this is the largest which would be recommended to use to be suitably similar to the exact solution. In Appendix B we discuss the choice of this criteria in the context of CMB reconstruction, but we have found it to be sufficient for all the example cases.

3.1 Projected Density Field

As our first case, we will look at reconstructing the density field from a noisy measurement of the density field. The starting measurement could come from a variety of observations such as galaxy clustering or Lyman alpha forest tomography. For a complete analysis one would specify the response matrix $\mathbf{R}$ to go from the density to the observables which would include a model for the various biases present in the observations. In this case we assume the bias model is known and deal directly with the matter density field.

We generate a Gaussian random field with a power-spectrum given by

$$P(k) \propto \frac{k}{1 + k^3} \quad (3.1)$$

over a 2D, $L = 1380$ Mpc/h side-length box. This formula is chosen to provide a blue spectrum (significant small scale power) in order to compare later with the red Cosmic Microwave Background spectra. We introduce an anisotropic white noise over the field to simulate either irregularities in depth of a given survey or theoretical uncertainties in the underlying bias

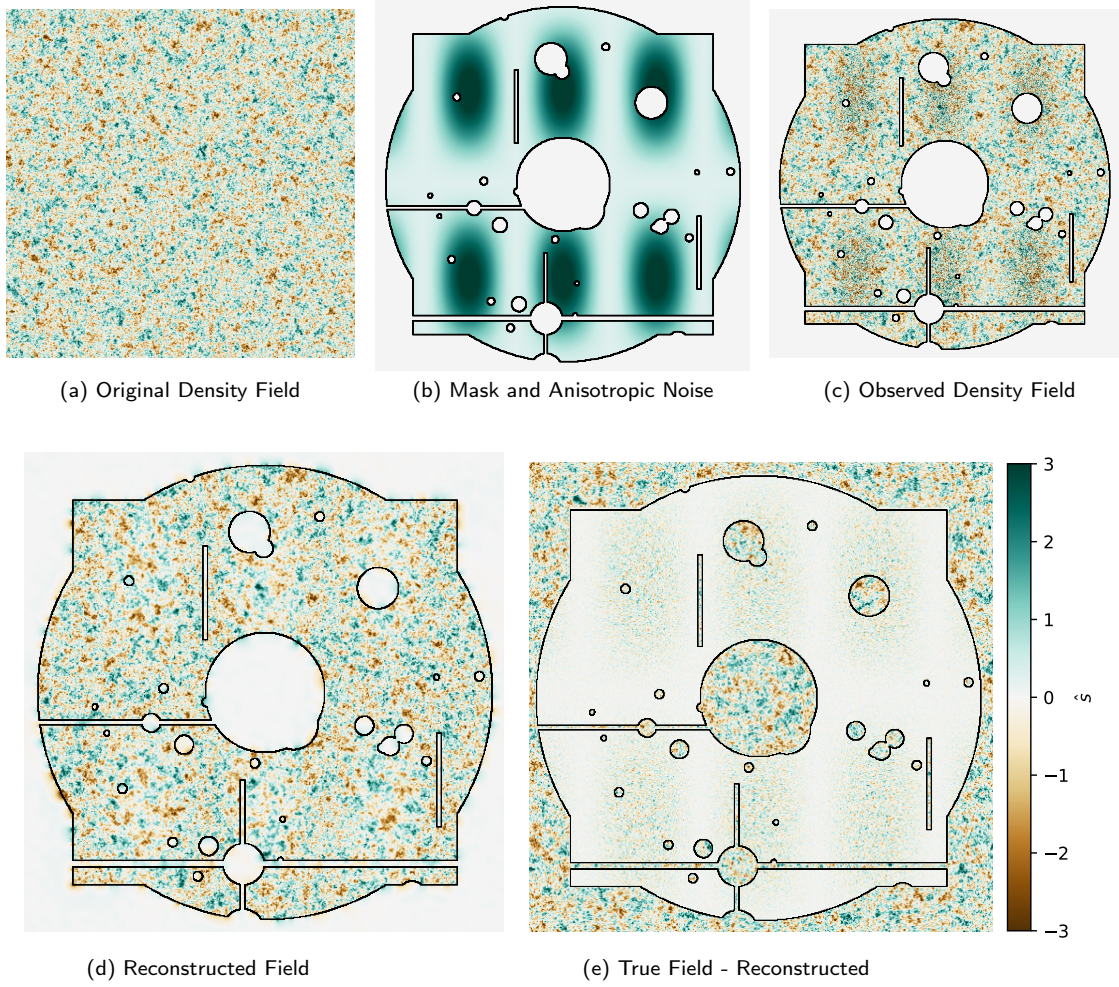


Figure 1: MAP density reconstruction for the 512x512 pixel case. Note that images (a), (c), (d), and (e) have same absolute color scale, while (b) shows the spatial variance of the noise properties. Color scale is normalized to show standard deviations away from mean.

model. In Appendix A we demonstrate the validity of the L-BFGS optimization method in a small test case (64×64 pixel) where it is also tractable to exactly invert the full covariance matrix numerically thereby providing validation of our maximum likelihood technique, while here we examine a more realistic 512×512 pixel map. We also use a realistic mask which includes foreground stars and other potential image defects.

Using the input power spectrum of Eq. 3.1, we generate a density field shown in Fig 1(a), apply a mask and anisotropic noise shown in Fig 1(b), which results in a mock observation in Fig 1(c). We perform the minimization routine outlined in Sec. 2.2, with the optimized map shown in Fig. 1(d) with residuals shown in Fig 1(e). Qualitatively the field is accurately reconstructed within the mask in the low-noise regions and is even able to reconstruct the larger scale modes right on the border within the masked region. However, as expected, the small scale modes within the high noise regions within the mask are poorly reconstructed since it is impossible to differentiate those modes in real space with the noise. In addition,

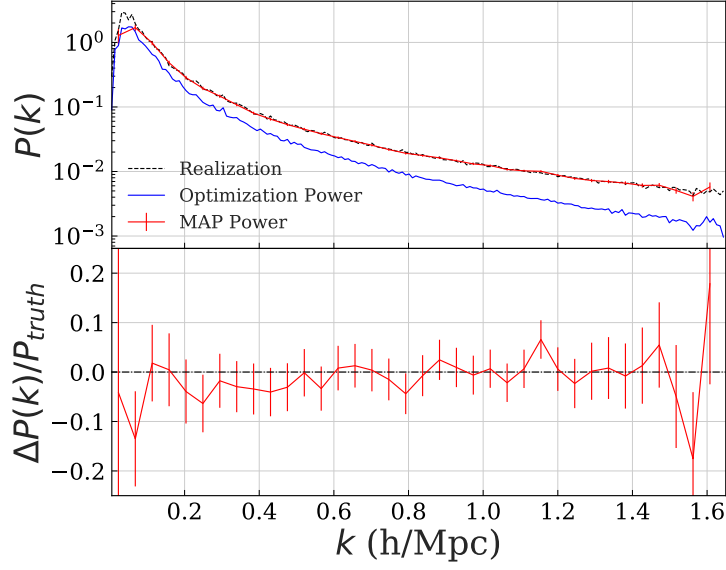


Figure 2: Comparison of the maximum likelihood power spectrum attained from optimization versus the true power-spectrum of the region for the density field. Also shown is the importance of the noise bias correction (or, equivalently, the importance of the Hessian determinant).

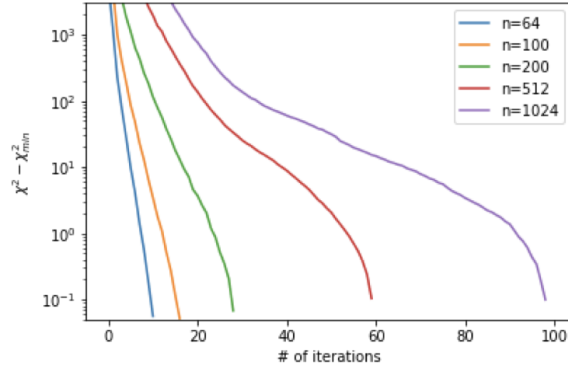


Figure 3: Convergence properties of the L-BFGS estimator as a function of the box-side dimension n .

small masked regions have very low residual error as there are sufficient, well sampled, nearby large scale modes to infer the regions value.

In Fig 2 we show the comparison of the optimized result with the true power-spectra of the entire field. We also show the effect of the noise bias correction, which in this case is substantial as small scale power is washed out in the high noise regions as well as due to the masked regions. However, this power is recoverable using the analysis described in Section 2.

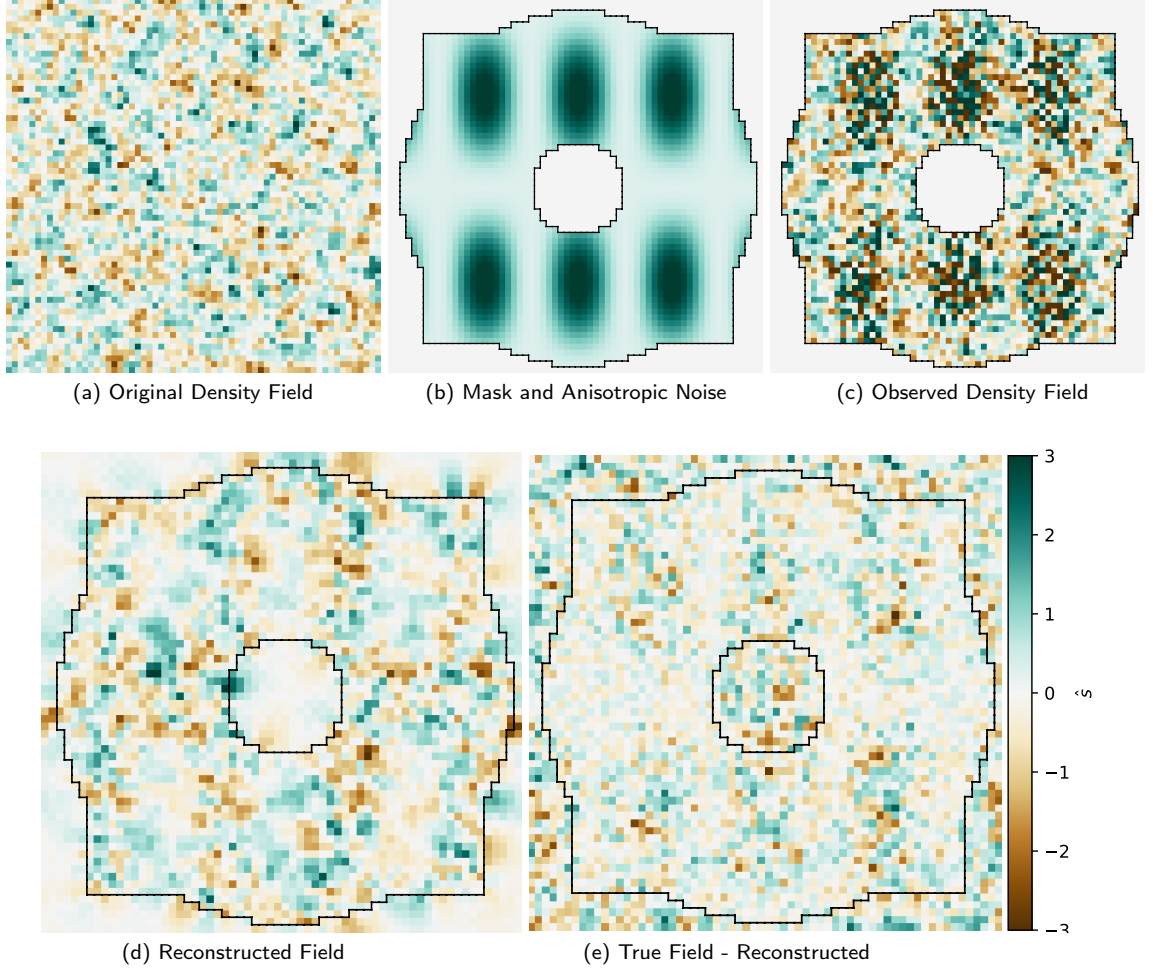


Figure 4: MAP density reconstruction for the 64x64 case.

In Fig 3 we show how the number of iterations of the reconstruction algorithm scales with the box size, holding the noise per unit angle fixed. Each iteration requires a calculation of the loss function and the derivatives of the field which can be found analytically. Notice that the iteration number increases with the box size, but only a factor of a few when going from 64^2 to 1024^2 . This does not include estimation of the noise bias and the Hessian matrix, which will depend on the number of bandpowers. As each row of the Hessian matrix requires an additional optimization, the true number of iteration will scale linearly with the number of bandpowers. In practice, since the Hessian matrix is very smooth and nearly translationally invariant (in this case, it is peaked on the diagonal and monotonically decreasing away from the diagonal), one simply needs to sample the matrix along a small number of rows and interpolate between them.

3.2 Wiener Filtering vs. MAP Projected Density Example

Direct numerical evaluation of Wiener filtering is computationally expensive as it requires the direct inversion of a matrix with the square of the number of pixels in the survey (see App A), so we specialize our direct comparison to a small 64×64 pixel image.

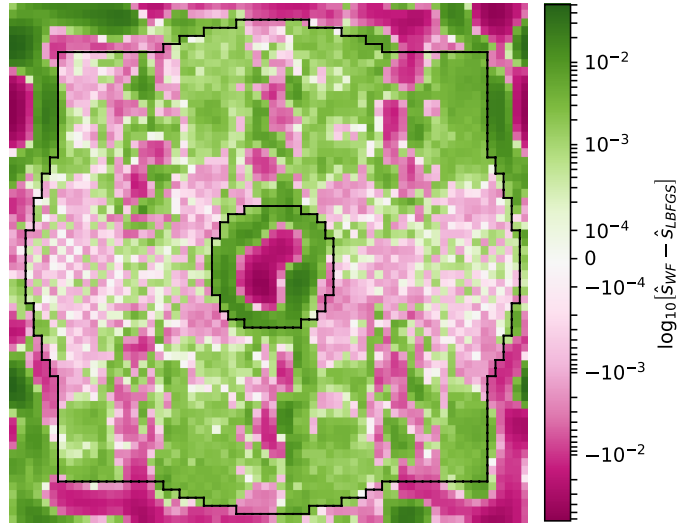


Figure 5: The log absolute magnitude difference of the direct matrix inversion Wiener filter solution and that attained via a MAP method. Note that differences are extremely small throughout the map and are particularly small in the unmasked region.

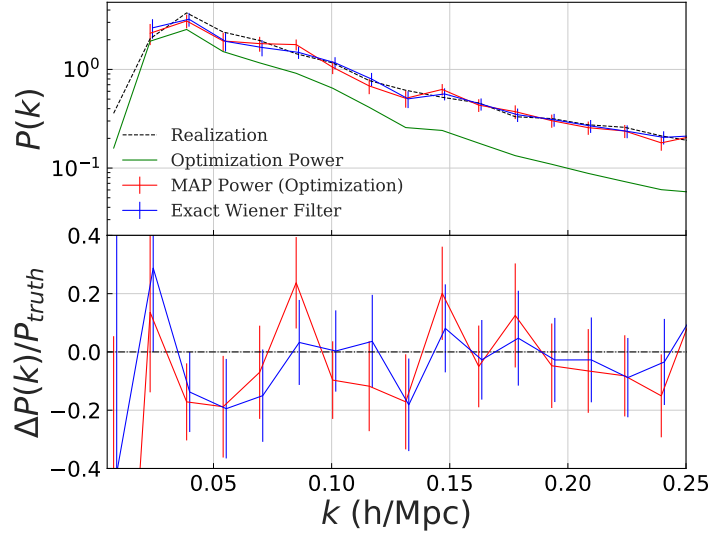


Figure 6: Comparison of the maximum likelihood power spectrum attained from optimization versus that for brute force matrix inversion Wiener filter. Errors are visually identical, and all points k bins are within one standard deviation of each other.

Using the input power spectrum of Eq. 3.1 we generate a density field shown in Fig 4(a), apply a mask and anisotropic noise shown in Fig 4(b), which results in a mock observation in Fig 4(c). We perform the minimization routine outlined in Sec. 2.2, with the optimized map shown in Fig. 4(d) with residuals shown in Fig 4(e). There is qualitative agreement between the truth and the reconstructed field within the mask in the low-noise regions and is even able to reconstruct the larger scale modes right on the border within the masked region. However, as expected, the small scale modes within the high noise regions within the mask are poorly reconstructed since it is impossible to differentiate those modes in real space with the noise.

For this small test case we can compare the optimization result directly against a numerical inversion Wiener filter solution, which we show on the map level in Fig 5. The results match outside the masked region within 10^{-4} , while in the masked region there is a slightly large difference due to the imposition of a convergence criteria in our L-BFGS scheme. As we increase the required precision of the L-BFGS in terms of $\epsilon \equiv \Delta\chi^2$, we asymptotically approach the Wiener filter solution.

Using the formulation in Sec 2.3, we can look at the performance of the technique as a function of scale. In Fourier space we can account for the reduction of small scale power caused by noise and also estimate the Hessian matrix (thereby giving error estimates). We show the power spectrum and error estimates from the optimization technique versus the direct Wiener filtering in Fig 6. Note that the full reconstruction relies on both calculation of noise bias and Hessian matrix. We have compared each of these terms from the optimization method to those calculated via direct matrix inversion to confirm they are equal within the error of the required optimization precision. Also note that we only used one noise realization to estimate the noise bias. In general, the number of noise realizations necessary to appropriately estimate the number of underlying band-powers will depend on both the underlying noise model and the band-powers of interest. In this particular case we found the improvements from including multiple noise realizations minimal as the effect on the overall power-spectrum were sub 1%. As with the map-level reconstruction, we find that decreasing the ϵ criteria leads to asymptotic convergence to the Wiener filtered solution.

We also explored the noise dominated regime more explicitly in Figure 7, where we apply a uniform high noise level over the entire field with variance 1.5 times the average variance in density. In this regime the optimized power is significantly suppressed across all scales, but the modes are still recoverable.

3.3 Cosmic Microwave Background Temperature

The question of optimal reconstruction of CMB maps given irregular sky coverage and variable noise and foreground subtraction is a common issue for existing CMB surveys. So far, true maximum likelihood power spectra estimators have only been applied to data from WMAP [6] and for the largest angular scales in Planck data [22], but these techniques are difficult to scale to the entire Planck dataset due to the significant increase in computational cost.

While Planck’s power spectrum measurements, and therefore cosmological parameter constraints [23], do not rely on construction of the actual full map, other spatially dependent signals do. Cross-correlations between the primary CMB and other cosmological probes, such as x-ray signal or galaxy positions, require an accurate spatial reconstruction of the CMB map. In addition, full sky CMB lensing maps are constructed by applying the quadratic estimators to the CMB map (in either temperature, polarization, or some minimum variance combination of the two) and will similarly suffer if the reconstructed temperature and/or

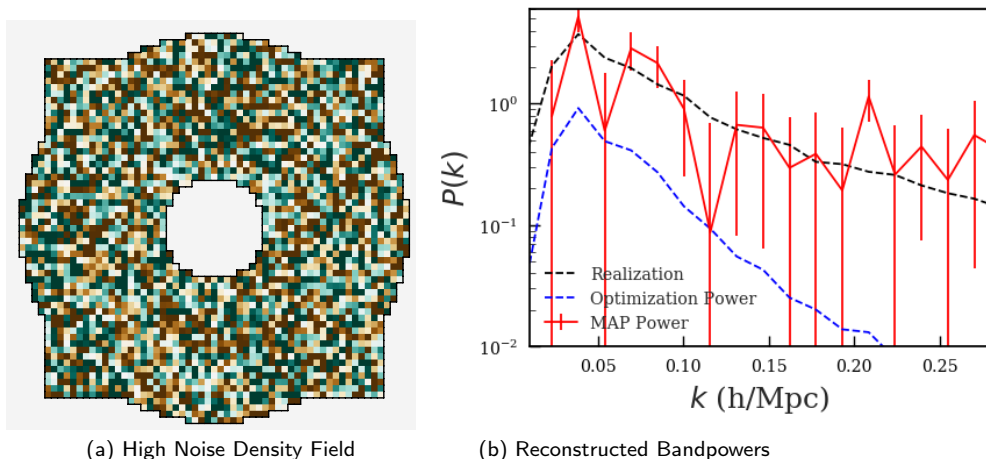


Figure 7: Observed field and associated reconstructed bandpowers for the noise dominated regime. Mask and color scale are the same as in Fig 5, and noise is uniform over the field.

polarization maps are suboptimal. In terms of map reconstruction, there are highly efficient Wiener filter programs available [14] which perform a multi-scale analysis with preconditioned optimization routines, which could replace our L-BFGS, but since the number of iterations needed is already low we did not explore this further.

In Figure 8 we show results for temperature, without polarization. We mention that polarization would be analogous to the example of joint optimization of E and B fields, which we do in the case of cosmic shear in Appendix C.

The case of CMB reconstruction is analogous to that of the density field, but with a significantly redder spectrum. The condition number of the covariance matrix is thus significantly larger, and BFGS needs more iterations to converge. While we could have used conjugate gradient with a preconditioner (as in [14]) to improve the convergence of the optimization step, we chose not to do so here since the computational cost was not significantly higher than that for the simple density case even for this case (see Figure 11). For the implementation of our algorithm it is important to recognize that the increase in power on larger scales in the CMB case makes the masked region the most computationally expensive region to reconstruct, and a redder spectrum will allow more mode reconstruction within this region. For more discussion, see Appendix B.

We generate a mock primary CMB full-sky field using HEALPIX [24] based on power spectrum generated from CLASS using the Planck 2015 cosmological parameters [23]. We then extract a 10×10 degree patch which we then mask a central region and introduce a white noise of $6 \mu K$ -arcmin.

The reconstruction is quite good in the observed region as the presence of noise mostly affects small scales where there is very little power. In the masked region the optimization is able to reproduce some clear long-scale modes. In Appendix B we show that with a more accurate convergence criteria we reproduce more of these large scale modes in the masked region but in terms of power spectrum estimation this is unnecessary, as we are already cosmic variance limited even with $\epsilon = 0.1$.

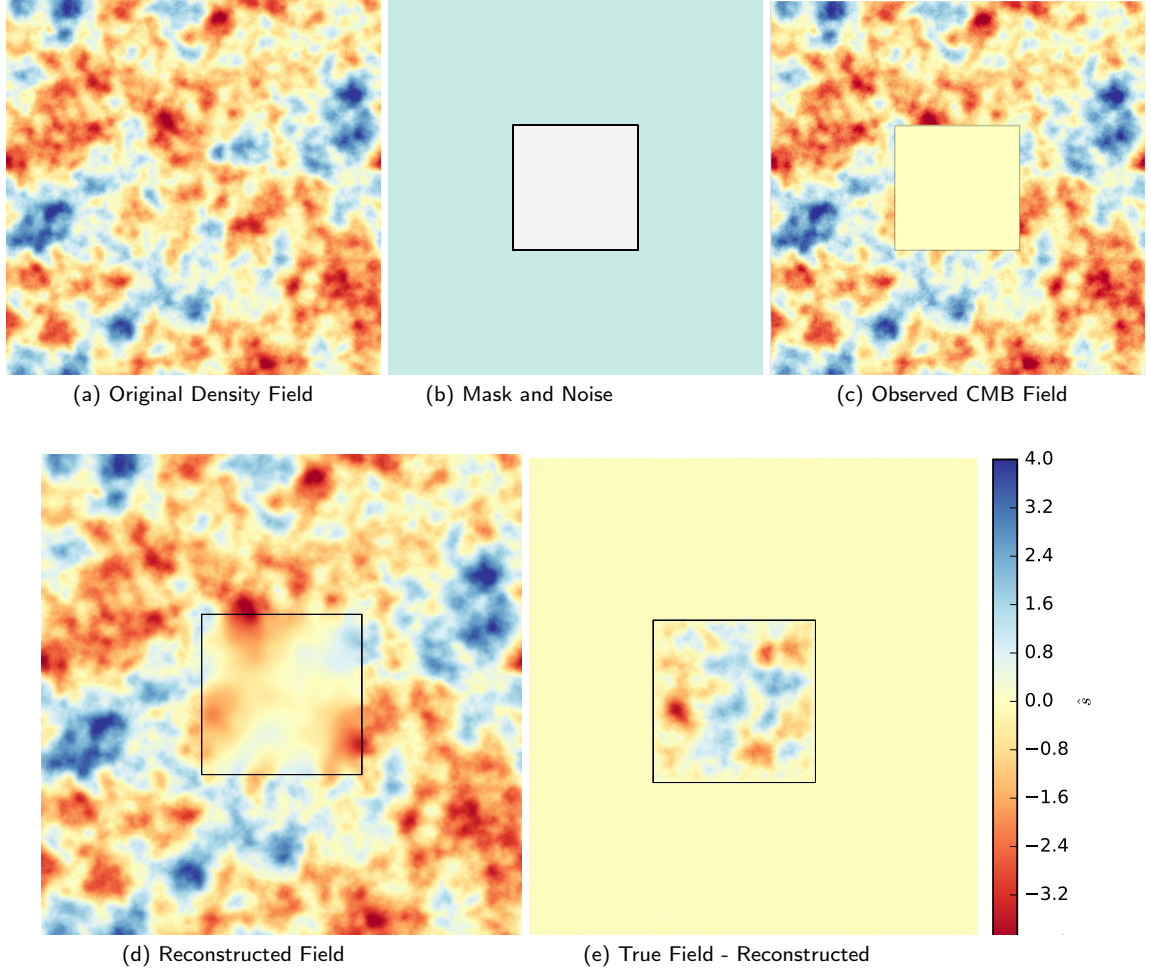


Figure 8: MAP CMB reconstruction for 512x512 pixel map. Note that images (a), (c), (d), and (e) have the same absolute color scale, while (b) shows the spatial variance of the noise properties. Color scale is normalized to show standard deviations away from mean.

3.4 Cosmic Shear

In this section we specialize to only fitting a curl-free E component; we discuss the more general case including a curl component in Appendix C. For details on the cosmic shear formalism, see [25–27].

To apply our method, we perform optimization over the underlying density field and at each step of the optimization compute the corresponding shear maps, γ_1 and γ_2 , to compare with the mock observed shear maps. In principle, instead of working with the shear maps which require binning and/or interpolation between galaxies, one could work directly with the catalog of galaxy shapes and compute the likelihood of the observed ellipticity instead of the given averaged shear maps (as in [28, 29]).

This method is in contrast to the standard Kaiser-Squires (KS) [30] technique which has proven quite successful so far in cosmic shear analysis and cluster mass estimation. However, KS has some notable downsides in the presence of anisotropic noise or a mask as it is not able to self-consistently down-weight the high noise areas and masked regions, resulting in defects

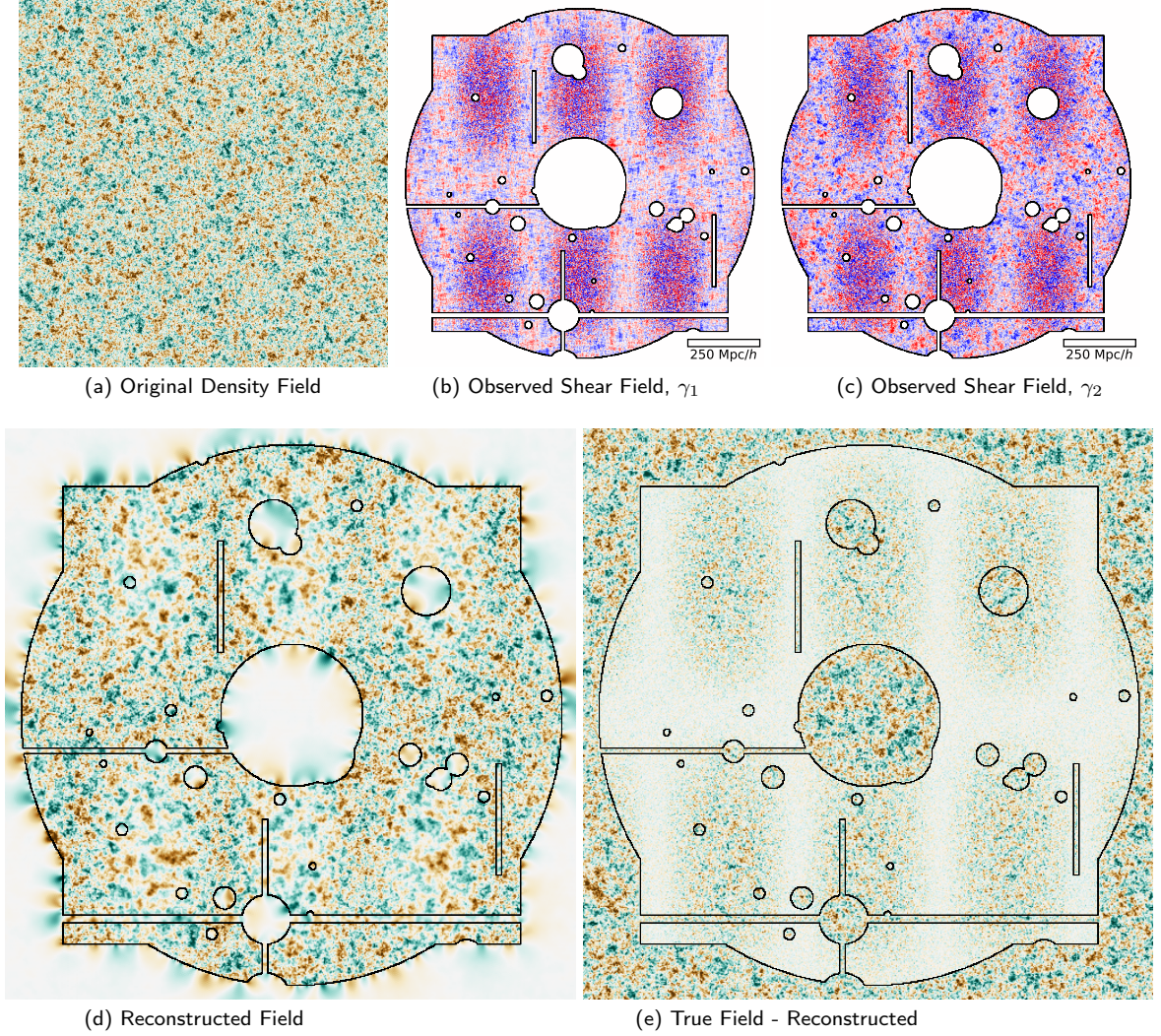


Figure 9: Maximum likelihood shear reconstruction for a 512x512 pixel map. Here we use the same mask and noise properties, as well as color scaling, as shown in Figure 1.

on boundaries. In addition, the noise inherent in these measurements propagates onto the final mass-maps, resulting in an inaccurate small scale power measurement. This has been shown to be particularly detrimental to peak statistics measurements [31].

The results in position space are shown in Figure 9. In Figure 9(a) we show the initial density field and in Figure 9(b,c) we show the observed shear fields including mask and noise properties. Our reconstructed maximum likelihood map is shown in Figure 9(d) and the difference between the original field and reconstructed in Figure 9(e). As in the case of the density field, the optimization technique is able to exactly reconstruct the density in the low noise, unmasked regions, but only recovers large scale power in the higher noise unmasked regions.

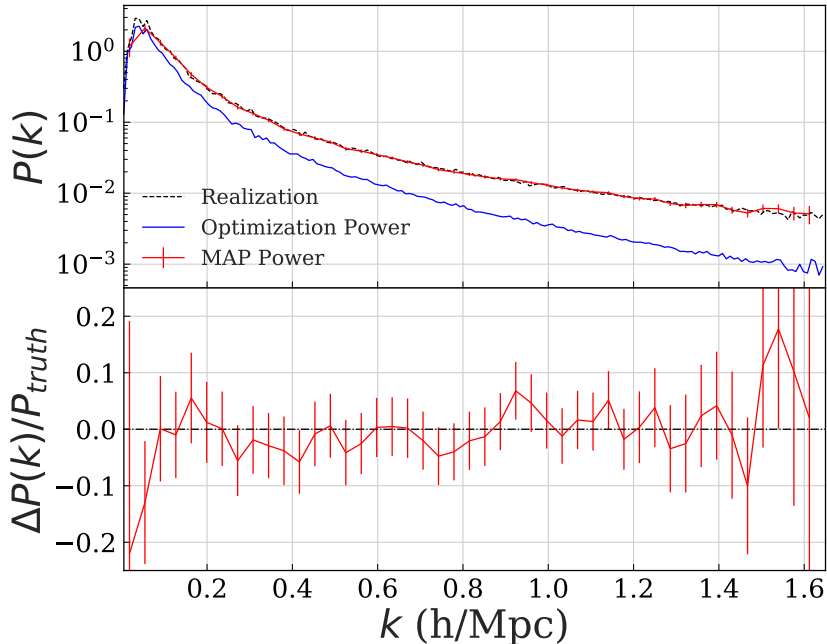


Figure 10: Comparison of the maximum likelihood power spectrum attained from optimization versus the true power-spectrum of the region for the shear-only reconstruction. Also shown is the importance of the noise bias correction (or, equivalently, the importance of the Hessian determinant).

4 Discussion and Conclusion

In this work we have demonstrated that it is possible to efficiently reconstruct the MAP signal field and the maximum likelihood power spectrum for linear fields for realistic survey sizes. This technique is equivalent to the Wiener filter solution for small enough convergence criteria and has been applied to a number of cosmological fields (density, CMB, and cosmic shear maps). We are able to reconstruct the initial density field and the overall power spectrum, accounting for noise bias and window-function effects due to masking. We first recast the field reconstruction into an optimization problem, which we solve using quasi-Newton optimization. We then recast the power spectrum estimation into the field marginalization problem, from which we obtain an expression that depends on the field reconstruction solution and a determinant term. We develop a novel simulation based method for the latter. We extend the simulations formalism to provide the covariance matrix for the power spectrum.

This technique outperforms the brute force Wiener filter technique in terms of computational time and memory requirements. True Wiener filter requires an inversion of the full pixel covariance matrix, $\mathbf{C}$, which for realistic surveys would be highly non trivial. Numerical methods approximate $\mathbf{C}^{-1}\mathbf{d}$, which allows for map-level reconstruction but by itself doesn't allow calculate of the Hessian matrix for band power reconstruction. This reconstruction also requires evaluation of the determinant of the Hessian, or its derivative, trace, and where the

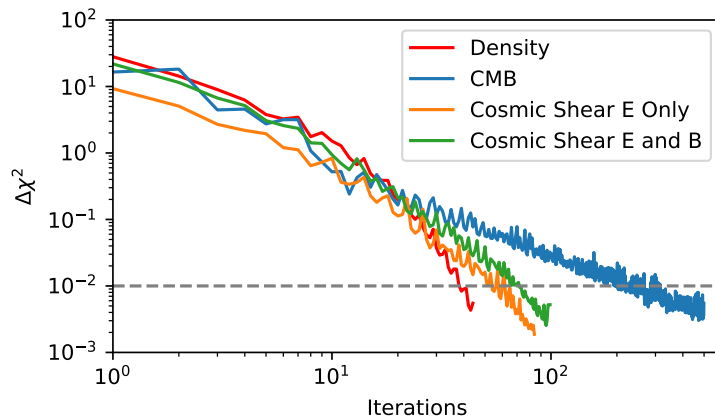


Figure 11: Comparison of the convergence properties of the various cosmological density fields studied in this work. All cases have 512^2 pixels, and comparable effective volume.

techniques provided in this work allows orders of magnitude improvement in realistic cases. We evaluate this determinant derivative using forward model realizations and additional optimization. This allows us to use off the shelf optimization codes such as L-BFGS, as well as convergence criteria to find the proper trade-off point between accuracy and computation time. We compare the two in Appendix 3.2, finding good agreement for low dimensionality problems where brute force approach is feasible.

In Figure 11 we show the convergence properties for the 3 cases studied in Section 3, as well as the joint E/B cosmic shear case presented in Appendix C. While all the cases have a comparable number of unmasked pixels, convergence properties differ due to the difference of the underlying fields power-spectrum. More power at larger scales (i.e. a redder spectra) requires additional iterations to reconstruct the power within the masked regions.

While we used L-BFGS due to its well established optimization properties in very high dimensional convex optimization problems, we do not make a claim of optimality in terms of the particular technique for performing the optimization. L-BFGS constructs low rank Hessian approximation to the Hessian, which makes it a quasi-Newton method: the closer this Hessian is to the true Hessian the closer we are to true second order optimization. In this limit this method will outperform any other method, including preconditioned conjugate gradient, which is only effective if preconditioning reduces the condition number of the problem. On the other hand, true second order Newton works for any condition number. We also note that since linear problems are convex, an optimizer is always guaranteed to find the global minimum. In general, we did not find particularly large performance changes when using other optimization techniques, such as conjugate gradient. Sampling based methods, like Hamiltonian Monte Carlo, are unnecessary for these linear cases, as there is no need to sample the distribution which is well approximated as a multivariate Gaussian, except for modes on order the size of the survey volume which are poorly sampled (and therefore do not follow the central limit theorem), in which case one can use approximations of inverse Wishart distribution developed in appendix A of [18].

A comparison can be made in our primary CMB example in Section 3.3 to the results of the messenger and dual messenger field found in [11] (see Figure 6 in [11]). The L-BFGS

approach requires significantly fewer iterations than the messenger field ($\sim 5x$) and dual messenger field ($\sim 2x$) for a comparable convergence criteria. It is possible that an optimized cooling scheme for the messenger/dual messenger field would yield similar convergence properties, but this choice would likely be problem-specific and introduce an additional parameter to tune in the optimization.

Going forward, it will be useful to extend this technique to other cosmological observables such as cosmic shear tomography (such as in [16]), Lyman - α tomography, and CMB lensing. Already work has been done applying this maximum likelihood approach to CMB lensing [32, 33], and further extending this work with these methods to small scales where standard quadratic estimators [34] are known to be suboptimal [35] would be promising future approach. Another avenue of particular interest is the ability for this technique to be useful for combining multiple (biased) tracers of some underlying field to create a maximum likelihood estimate of the field. One particularly promising example is jointly maximizing the underlying density likelihood function with respect to both the shear map as well as the projected galaxy density map [36, 37].

These linear methods have a limitation when applied to nonlinear fields. Recent work [31] has demonstrated that Wiener Filtering is not optimal in terms of detecting peaks in the density field and that sparsity-based reconstruction methods can yield higher signal to noise. More general reconstruction of density fields was explored in [38], where they created MAP estimates of physical clusters with associated error estimates without making assumptions about the Gaussian of the likelihood surface. However, these approaches are ad-hoc, as the loss function they minimize cannot be theoretically justified. For the case of nonlinear large scale structure, best analog of WF reconstructions in terms of minimizing the error are the nonlinear density reconstruction techniques developed e.g. in [18, 39, 40], which first give a minimal variance reconstruction of initial Gaussian density fluctuations, and then project these into the nonlinear structures using an N-body simulation.

Acknowledgments

We would like to thank Vanessa Böhm, Simone Ferraro, Stephanie Ger, Chirag Modi, and Michael Schneider for useful discussions and/or comments during the preparation of this manuscript. BH is supported by the NSF Graduate Research Fellowship, award number DGE 1106400. US acknowledges support from NSF 1814370, NSF 1839217, NASA 17-ATP17-0007 and NNX15AL17G.

This research has made use of NASA’s Astrophysics Data System. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

A Wiener Filter Review

Wiener filtering (WF) [1, 2] is a popular way to non-parametrically reconstruct cosmological data as, in the linear case, it should minimize variance. In the absence of non-Gaussian sources of signal or noise, WF is optimal in the sense that it is equal to the maximum posterior probability estimator [18]. Here we want to reconstruct the field itself, $\mathbf{s}$, given the noisy and/or poorly sampled data, $\mathbf{d}$. We define our estimated field $\hat{\mathbf{s}} = \Phi \mathbf{d}$, where Φ is a linear operator, i.e. a $N \times M$ dimensional matrix transforming from “image space” to

“field space”. This can be found by attempting to minimize the variance of the residual, i.e. $\langle (\mathbf{s} - \hat{\mathbf{s}})(\mathbf{s} - \hat{\mathbf{s}})^\dagger \rangle$, with respect to Φ . The Wiener filtered estimator is

$$\hat{\mathbf{s}} = \Phi \mathbf{d} = \langle \mathbf{s} \mathbf{d}^\dagger \rangle \langle \mathbf{d} \mathbf{d}^\dagger \rangle^{-1} \mathbf{d} = \mathbf{S} \mathbf{R}^\dagger \mathbf{C}^{-1} \mathbf{d}, \quad (\text{A.1})$$

which will result in a variance of residuals of the form

$$\langle (\mathbf{s} - \hat{\mathbf{s}})(\mathbf{s} - \hat{\mathbf{s}})^\dagger \rangle = \mathbf{S} - \mathbf{S} \mathbf{R}^\dagger \mathbf{C}^{-1} \mathbf{R} \mathbf{S}. \quad (\text{A.2})$$

Wiener filter only uses the mean and variance of the statistical distribution. If our underlying field is strongly non-Gaussian the WF may no longer be optimal. However, it will still minimize the variance, as defined in equation 2.6, just that this minimization of variance may not correspond to the notion of the best reconstruction as it only captures the two point statistics of the underlying field. It may be difficult to even define a measure to use for optimality of reconstruction in these cases, although nonlinear reconstruction methods certainly exist [18].

We now want to connect the Wiener Filter solution to the optimal power spectrum estimator. We multiply equation 2.7 by the Hessian matrix,

$$\begin{aligned} (\mathbf{F}\Theta)_l &= \frac{\mathbf{F}}{2} \sum_{l'} F_{ll'}^{-1} (\mathbf{d}^\dagger \mathbf{C}^{-1} \mathbf{Q}_{l'} \mathbf{C}^{-1} \mathbf{d} - b_{l'}), \\ &= \frac{\delta_{ll'}}{2} (\mathbf{d}^\dagger \mathbf{C}^{-1} \Pi_{l'} \mathbf{R} \mathbf{R}^\dagger \Pi_{l'} \mathbf{C}^{-1} \mathbf{d} - b_{l'}). \end{aligned} \quad (\text{A.3})$$

The noise $\mathbf{b}$ can be similarly transformed as

$$b_l = \text{tr}(\Pi_l \mathbf{R}^\dagger \mathbf{C}^{-1} (\mathbf{N}) \mathbf{C}^{-1} \mathbf{R} \Pi_l), \quad (\text{A.4})$$

and the Hessian matrix itself as

$$F_{ll'} = \text{tr}(\mathbf{C}^{-1} \mathbf{Q}_l \mathbf{C}^{-1} \mathbf{Q}_{l'}) = \frac{1}{2} |\Pi_l \mathbf{R}^\dagger \mathbf{C}^{-1} \mathbf{R} \Pi_{l'}|^2. \quad (\text{A.5})$$

Both the Wiener Filter and the optimal power spectrum estimator first weigh the data by the inverse covariance matrix, essentially down weighting modes that either have high measurement error or strong correlation with other measurements.

B Convergence Criteria of CMB Reconstruction

An important question to answer is to what the required convergence criteria are for a given algorithm/observable. In general, this will depend on what sort of scales are being probed and what other sources of error exist in the problem. In this section we will consider how changing the convergence criteria, $\epsilon \equiv \Delta\chi^2$, affects the net reconstructed map. We will specialize our analysis to that of the CMB case since it has the largest condition number, and most power on large scales, which will be particularly sensitive to reconstruction within the masked region. A similar analysis with our cosmic shear example will lead to smaller effects.

To demonstrate the convergence properties of our technique we performed a high-accuracy run demanding $\epsilon < 10^{-4}$ as our convergence criteria, as opposed to $\epsilon < 10^{-1}$ for the runs in the main body of the paper. We show these results, as well as the difference with the true field, in Figure 12. Note that very quickly we find the true solution in the unmasked region, but continue to reconstruct the large scale modes in the masked region as the optimization rerouting continues.

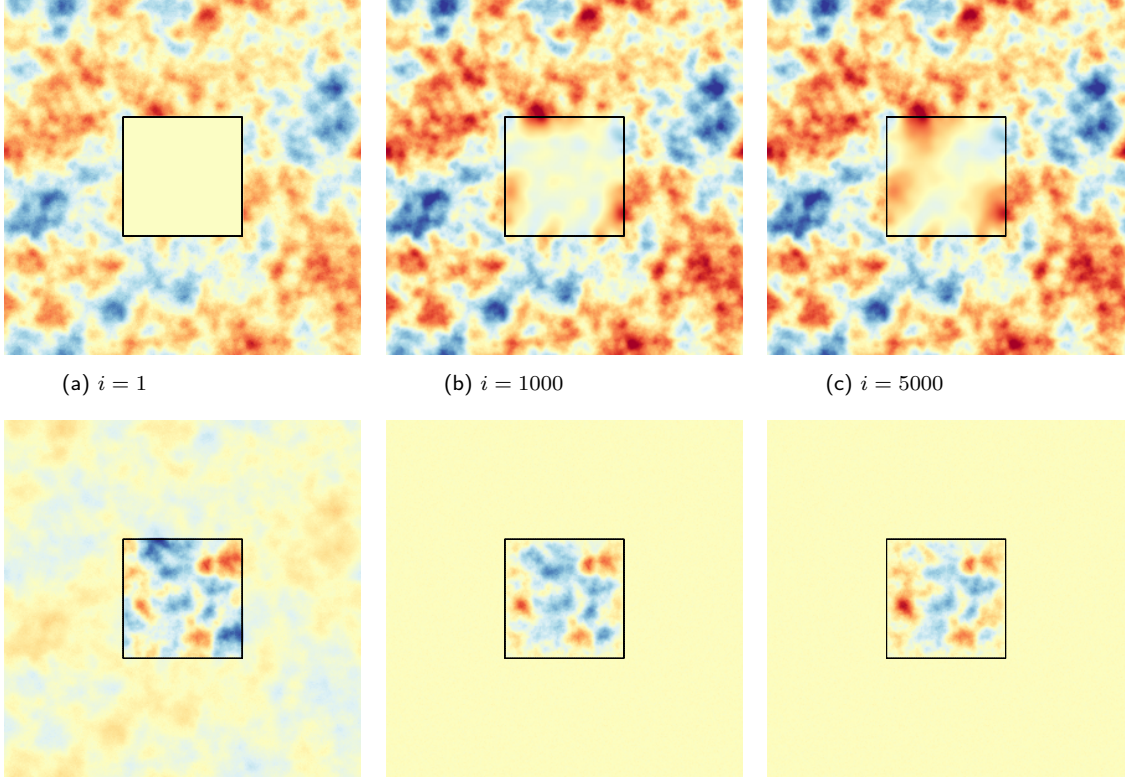


Figure 12: *Top:* Reconstructed density field at given iteration. *Bottom:* Difference of true density field with reconstruction at each iteration. Note that we have used the same mask/noise properties, as well as color scale, as in Sec 3.3.

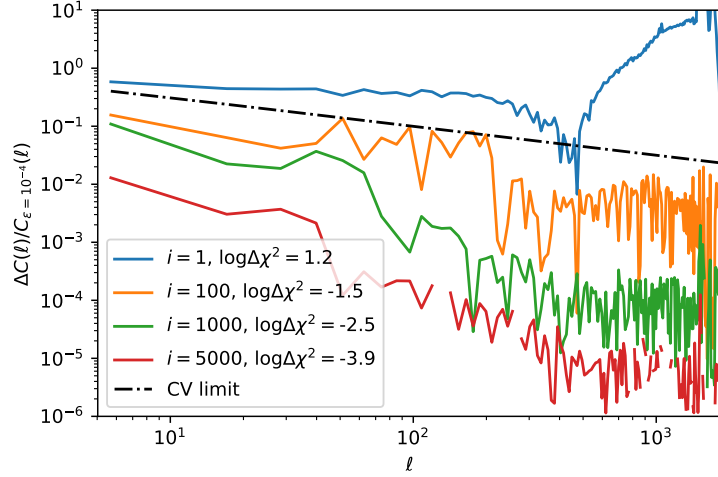


Figure 13: Change in convergence properties as a function of scale. We compare against the high convergence solution ($\epsilon = 10^{-4}$) rather than the true solution as the presence of noise will bias the end power spectra and calculating the noise bias and the Fisher information matrix (Hessian) for each step of the iteration would be computationally expensive.

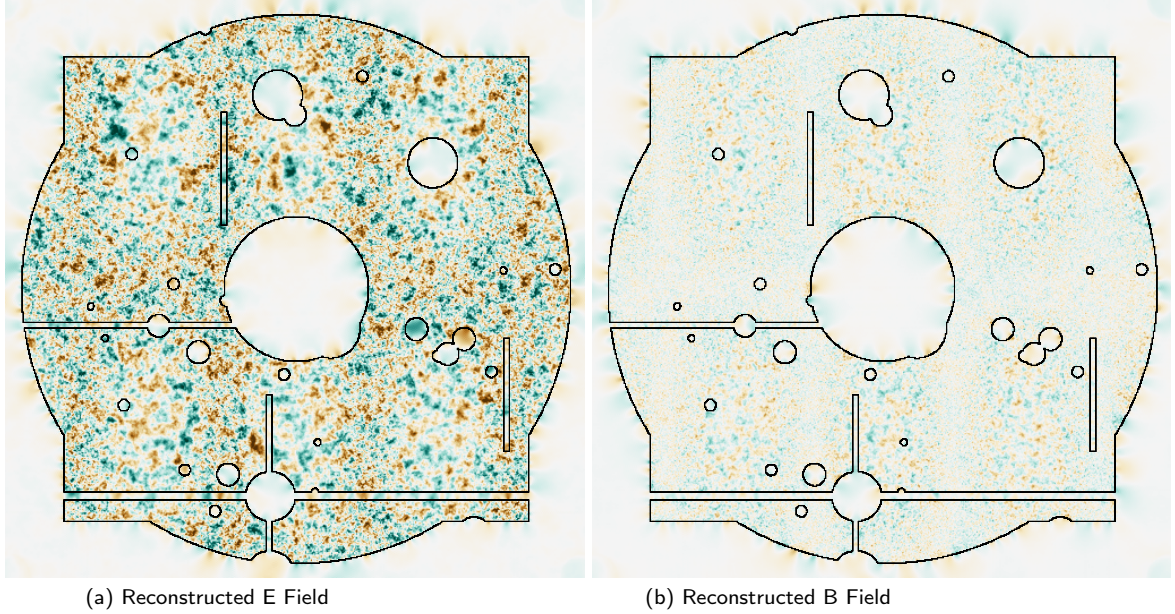


Figure 14: Maximum likelihood E and B potential fields for cosmic shear case. Noise properties, mask, and color scale are the same as in Section 3.4.

We compare our convergence accuracy as a function of scale to the cosmic variance limit in Figure 13. While in practice one wants the error on the reconstruction to be well below this limit, it provides a useful guideline for the necessary accuracy for reconstruction. Note that the properties of this reconstruction are a function primarily of the survey geometry; a hypothetical full sky survey with no masked region and similar noise properties would converge much quicker to the optimal solution. Similarly, a case with smaller masked regions (for example only stellar masking) would find much faster convergence of the large scale modes.

C Joint E & B Cosmic Shear Reconstruction

In the main text we only explored reconstructing the primary (i.e. curl-free) E -mode lensing potential of cosmic shear. However, there are various potential sources of B -mode effects within realistic observed lensing maps, such as instrumental effects, clustering of source galaxies [41], and intrinsic alignments of galaxy shapes [42, 43]. To control for these effects it is useful to perform a joint optimization of both E and B modes from the shear maps. The same tools could also be applied directly to the CMB polarization field from the Q and U maps [32].

The observed shear fields γ_1 and γ_2 can be expressed in terms of the E and B potentials as

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} (\partial_x^2 - \partial_y^2) & -2\partial_x\partial_y \\ 2\partial_x\partial_y & (\partial_x^2 - \partial_y^2) \end{bmatrix} \begin{bmatrix} \phi_E \\ \phi_B \end{bmatrix}, \quad (\text{C.1})$$

where we assume flat sky. Our response matrix now takes two signal fields (ϕ_E and ϕ_B) to two data fields (γ_1 and γ_2) and we perform the optimization over the signal fields.

To study the joint reconstruction, we use the same starting E field as in Subsection 3.4, but also induce a B field which has a power spectrum with the same shape as the E field but an amplitude 10^{-5} times smaller. We then do a joint reconstruction of both fields, yielding results shown in Fig 14.

The power spectrum estimation is now slightly more complicated as the Hessian matrix now has EE, BB, as well as EB, BE components to sum over, as shown in Fig 15. The EB/BE components represent leakage between the channels, which in this case is dominated by E power leaking into B. We show this power spectrum reconstruction in Figure 16. This is visually apparent in the reconstruction as we find an over-abundance of B power in the reconstructed map, which then gets down-weighted when this leakage is accounted for. In addition, as the B mode power is dominated by noise, it is difficult to accurately reconstruct its power from this one mock observation.

Alternatively, rather than perform a full Fisher-like analysis for the B-mode power one could instead perform multiple realizations of the B-mode leakage and average them together to form a “leakage bias” in analogous way as for the noise terms. This would have the possible advantage of requiring an additional optimization for each realization of the B-mode power, rather than an optimization for each mode of injected power as for the full Fisher analysis. If one is interested in studying many modes at once, treating the B-modes like noise bias would be computationally expedient.

References

- [1] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*, .
- [2] G. B. Rybicki and W. H. Press, *Interpolation, realization, and reconstruction of noisy, irregularly sampled data*, *ApJ* **398** (Oct., 1992) 169–176.
- [3] K. B. Fisher, O. Lahav, Y. Hoffman, D. Lynden-Bell, and S. Zaroubi, *Wiener reconstruction of density, velocity and potential fields from all-sky galaxy redshift surveys*, *MNRAS* **272** (Feb., 1995) 885–908, [[astro-ph/9406009](#)].
- [4] F. S. Kitaura and T. A. Enßlin, *Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization*, *MNRAS* **389** (Sept., 2008) 497–544, [[arXiv:0705.0429](#)].
- [5] E. F. Bunn, Y. Hoffman, and J. Silk, *The Wiener-filtered COBE DMR Data and Predictions for the Tenerife Experiment*, *ApJ* **464** (June, 1996) 1, [[astro-ph/9509045](#)].
- [6] M. Tegmark, A. de Oliveira-Costa, and A. J. Hamilton, *High resolution foreground cleaned CMB map from WMAP*, *PRD* **68** (Dec., 2003) 123523, [[astro-ph/0302496](#)].
- [7] E. Komatsu, D. N. Spergel, and B. D. Wandelt, *Measuring Primordial Non-Gaussianity in the Cosmic Microwave Background*, *ApJ* **634** (Nov., 2005) 14–19, [[astro-ph/0305189](#)].
- [8] U. Seljak, *Cosmography and Power Spectrum Estimation: A Unified Approach*, *ApJ* **503** (Aug., 1998) 492–501, [[astro-ph/9710269](#)].
- [9] F. Elsner and B. D. Wandelt, *Efficient Wiener filtering without preconditioning*, *A&A* **549** (Jan., 2013) A111, [[arXiv:1210.4931](#)].
- [10] J. Alsing, A. Heavens, A. H. Jaffe, A. Kiessling, B. Wandelt, and T. Hoffmann, *Hierarchical cosmic shear power spectrum inference*, *MNRAS* **455** (Feb., 2016) 4452–4466, [[arXiv:1505.07840](#)].

- [11] D. Kodi Ramanah, G. Lavaux, and B. D. Wandelt, *Wiener filter reloaded: fast signal reconstruction without preconditioning*, *MNRAS* **468** (June, 2017) 1782–1793, [[arXiv:1702.08852](#)].
- [12] J. Papez, L. Grigori, and R. Stompor, *Solving linear equations with messenger-field and conjugate gradients techniques - an application to CMB data analysis*, *ArXiv e-prints* (Mar., 2018) [[arXiv:1803.03462](#)].
- [13] J. Jewell, S. Levin, and C. H. Anderson, *Application of Monte Carlo Algorithms to the Bayesian Analysis of the Cosmic Microwave Background*, *ApJ* **609** (July, 2004) 1–14, [[astro-ph/0209560](#)].
- [14] D. S. Seljebotn, T. B erland, H. K. Eriksen, K.-A. Mardal, and I. K. Wehus, *Multi-resolution Bayesian CMB component separation through Wiener-filtering with a pseudo-inverse preconditioner*, *ArXiv e-prints* (Oct., 2017) [[arXiv:1710.00621](#)].
- [15] J. Alsing, A. Heavens, and A. H. Jaffe, *Cosmological parameters, shear maps and power spectra from CFHTLenS using Bayesian hierarchical inference*, *MNRAS* **466** (Apr., 2017) 3272–3292, [[arXiv:1607.00008](#)].
- [16] P. Simon, A. N. Taylor, and J. Hartlap, *Unfolding the matter distribution using three-dimensional weak gravitational lensing*, *MNRAS* **399** (Oct., 2009) 48–68, [[arXiv:0907.0016](#)].
- [17] T. A. En lin and M. Frommert, *Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty*, *PRD* **83** (May, 2011) 105014, [[arXiv:1002.2928](#)].
- [18] U. Seljak, G. Aslanyan, Y. Feng, and C. Modi, *Towards optimal extraction of cosmological information from nonlinear data*, *ArXiv e-prints* (June, 2017) [[arXiv:1706.06645](#)].
- [19] J. R. Bond, A. H. Jaffe, and L. Knox, *Estimating the power spectrum of the cosmic microwave background*, *PRD* **57** (Feb., 1998) 2117–2137, [[astro-ph/9708203](#)].
- [20] M. Tegmark, *How to measure CMB power spectra without losing information*, *PRD* **55** (May, 1997) 5895–5907, [[astro-ph/9611174](#)].
- [21] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.
- [22] Planck Collaboration, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, and et al., *Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters*, *A&A* **594** (Sept., 2016) A11, [[arXiv:1507.02704](#)].
- [23] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al., *Planck 2015 results. XIII. Cosmological parameters*, *A&A* **594** (Sept., 2016) A13, [[arXiv:1502.01589](#)].
- [24] K. M. G rski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, *ApJ* **622** (Apr., 2005) 759–771, [[astro-ph/0409513](#)].
- [25] L. Van Waerbeke, Y. Mellier, M. Radovich, E. Bertin, M. Dantel-Fort, H. J. McCracken, O. Le F vre, S. Foucaud, J.-C. Cuillandre, T. Erben, B. Jain, P. Schneider, F. Bernardeau, and B. Fort, *Cosmic shear statistics and cosmology*, *A&A* **374** (Aug., 2001) 757–769, [[astro-ph/0101511](#)].
- [26] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, *PhysRep* **340** (Jan., 2001) 291–472, [[astro-ph/9912508](#)].
- [27] B. Jain and U. Seljak, *Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes*, *ApJ* **484** (July, 1997) 560–573, [[astro-ph/9611077](#)].

- [28] N. Padmanabhan, U. Seljak, and U. L. Pen, *Mining weak lensing surveys.*, *New Astronomy* **8** (2003) 581–603, [[astro-ph/0210478](#)].
- [29] V. Böhm, S. Hilbert, M. Greiner, and T. A. Enßlin, *Bayesian weak lensing tomography: Reconstructing the 3D large-scale distribution of matter with a lognormal prior*, *PRD* **96** (Dec., 2017) 123510, [[arXiv:1701.01886](#)].
- [30] N. Kaiser, G. Squires, and T. Broadhurst, *A Method for Weak Lensing Observations*, *ApJ* **449** (Aug., 1995) 460–+.
- [31] N. Jeffrey, F. B. Abdalla, O. Lahav, F. Lanusse, J.-L. Starck, A. Leonard, D. Kirk, C. Chang, E. Baxter, T. Kacprzak, S. Seitz, V. Vikram, L. Whiteway, T. M. C. Abbott, S. Allam, S. Avila, E. Bertin, D. Brooks, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, M. Crocce, C. E. Cunha, C. B. D’Andrea, L. N. da Costa, C. Davis, J. De Vicente, S. Desai, P. Doel, T. F. Eifler, A. E. Evrard, B. Flaugher, P. Fosalba, J. Frieman, J. Garcia-Bellido, D. W. Gerdes, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, K. Honscheid, B. Hoyle, D. J. James, M. Jarvis, K. Kuehn, M. Lima, H. Lin, M. March, P. Melchior, F. Menanteau, R. Miquel, A. A. Plazas, K. Reil, A. Roodman, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, and A. R. Walker, *Improving Weak Lensing Mass Map Reconstructions using Gaussian and Sparsity Priors: Application to DES SV*, *ArXiv e-prints* (Jan., 2018) [[arXiv:1801.08945](#)].
- [32] C. M. Hirata and U. Seljak, *Reconstruction of lensing from the cosmic microwave background polarization*, *PRD* **68** (Oct., 2003) 083002, [[astro-ph/0306354](#)].
- [33] J. Carron and A. Lewis, *Maximum a posteriori CMB lensing reconstruction*, *PRD* **96** (Sept., 2017) 063510, [[arXiv:1704.08230](#)].
- [34] W. Hu and T. Okamoto, *Mass Reconstruction with Cosmic Microwave Background Polarization*, *ApJ* **574** (Aug., 2002) 566–574, [[astro-ph/0111606](#)].
- [35] B. Horowitz, S. Ferraro, and B. D. Sherwin, *Reconstructing Small Scale Lenses from the Cosmic Microwave Background*, *ArXiv e-prints* (Oct., 2017) [[arXiv:1710.10236](#)].
- [36] R. M. Szepletowski, D. J. Bacon, J. P. Dietrich, M. Busha, R. Wechsler, and P. Melchior, *Density mapping with weak lensing and phase information*, *MNRAS* **440** (May, 2014) 2191–2200, [[arXiv:1306.5324](#)].
- [37] P. Simon, *Improving three-dimensional mass mapping with weak gravitational lensing using galaxy clustering*, *A&A* **560** (Dec., 2013) A33, [[arXiv:1203.6205](#)].
- [38] M. A. Price, J. D. McEwen, X. Cai, T. D. Kitching, and C. G. R. Wallis, *Sparse Bayesian mass-mapping with uncertainties: hypothesis testing of structure*, [[arXiv:1812.04014](#)].
- [39] H. Wang, H. J. Mo, X. Yang, Y. P. Jing, and W. P. Lin, *ELUCID-Exploring the Local Universe with the Reconstructed Initial Density Field. I. Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics*, *ApJ* **794** (Oct., 2014) 94, [[arXiv:1407.3451](#)].
- [40] J. Jasche and B. D. Wandelt, *Bayesian physical reconstruction of initial conditions from large-scale structure surveys*, *MNRAS* **432** (June, 2013) 894–913, [[arXiv:1203.3639](#)].
- [41] P. Schneider, L. van Waerbeke, and Y. Mellier, *B-modes in cosmic shear from source redshift clustering*, *A&A* **389** (July, 2002) 729–741, [[astro-ph/0112441](#)].
- [42] C. M. Hirata and U. Seljak, *Intrinsic alignment-lensing interference as a contaminant of cosmic shear*, *PRD* **70** (Sept., 2004) 063526, [[astro-ph/0406275](#)].
- [43] M. A. Troxel and M. Ishak, *The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology*, *PhysRep* **558** (Feb., 2015) 1–59, [[arXiv:1407.6990](#)].

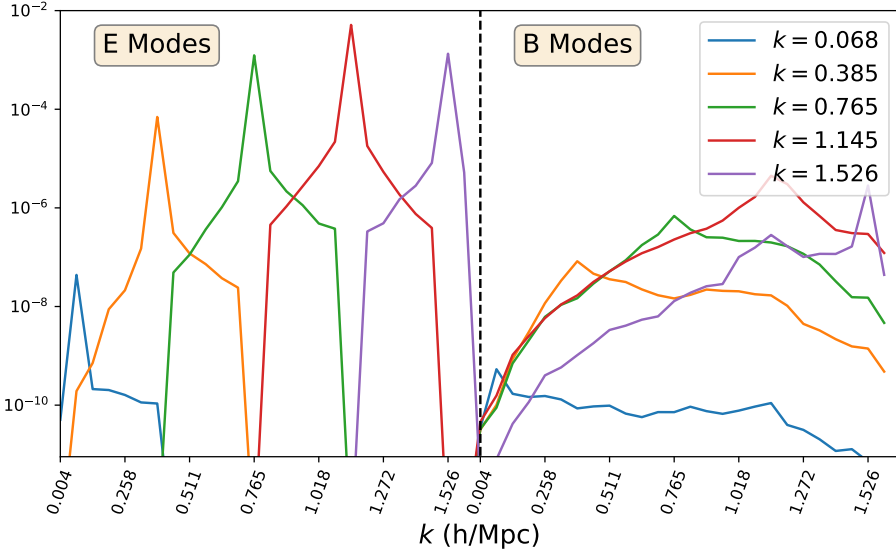
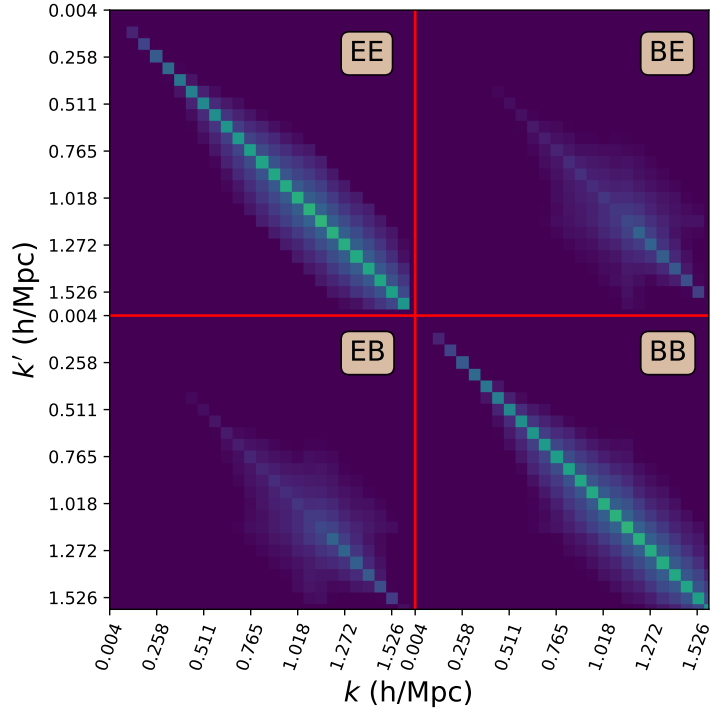


Figure 15: *Top:* Full two dimensional Fisher matrix for the cosmic shear E/B joint reconstruction case. The matrix can be viewed as 4 blocks, with EE and BB the response of each type of mode to itself and the BE and EB reflecting the leakage between the modes induced by the survey geometry. *Bottom:* Vertical cuts of the Fisher matrix. Note that to reduce numerical noise in the final reconstruction we have zeroed out terms in the EE blocks far from the diagonal. Since B modes are significantly sub-dominant in this example, a similar truncation of BB was not necessary.

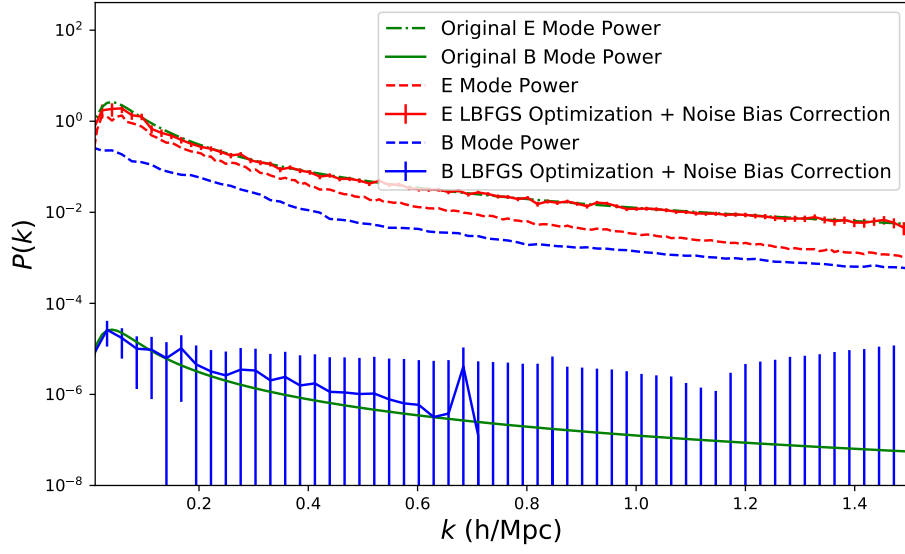


Figure 16: Reconstruction of the E and B power spectrum from mock observations using the maximum likelihood technique described in this work. Green lines indicate the original E/B power of the signal maps. The dashed red and blue lines indicate the power from the LBFSG optimized maps of E and B power respectively, while the solid lines indicate their MAP power spectra.