Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **506,** 3678–3690 (2021) https://doi.org/10.1093/mnras/stab1983
Advance Access publication 2021 July 12

Sparse Bayesian mass mapping with uncertainties: hypothesis testing of structure

M. A. Price,★ J. D. McEwen, X. Cai, T. D. Kitching, C. G. R. Wallis and (for the LSST Dark Energy Science Collaboration)

Mullard Space Science Laboratory, University College London, London RH5 6NT, UK

Accepted 2021 July 2. in original form 2021 May 17

ABSTRACT

A crucial aspect of mass mapping, via weak lensing, is quantification of the uncertainty introduced during the reconstruction process. Properly accounting for these errors has been largely ignored to date. We present a new method to reconstruct maximum a posteriori (MAP) convergence maps by formulating an unconstrained Bayesian inference problem with Laplace-type l_1-norm sparsity-promoting priors, which we solve via convex optimization. Approaching mass mapping in this manner allows us to exploit recent developments in probability concentration theory to infer theoretically conservative uncertainties for our MAP reconstructions, without relying on assumptions of Gaussianity. For the first time, these methods allow us to perform hypothesis testing of structure, from which it is possible to distinguish between physical objects and artefacts of the reconstruction. Here, we present this new formalism, and demonstrate the method on simulations, before applying the developed formalism to two observational data sets of the Abell 520 cluster. Initial reconstructions of the Abell 520 catalogues reported the detection of an anomalous 'dark core' – an overdense region with no optical counterpart – which was taken to be evidence for self-interacting dark matter. In our Bayesian framework, it is found that neither Abell 520 data set can conclusively determine the physicality of such dark cores at 99 per cent confidence. However, in both cases the recovered MAP estimators are consistent with both sets of data.

Key words: gravitational lensing: weak – methods: data analysis – methods: statistical – techniques: image processing – dark matter.

1 INTRODUCTION

Gravitational lensing is an astrophysical phenomenon, which can be observed on galactic to cosmic spatial scales, through which distant images are distorted by the intervening mass density field. Due to its nature, lensing is sensitive to the total mass distribution (both visible and invisible) along a line of sight (Bartelmann & Schneider 2001; Schneider 2005; Munshi et al. 2008; Heavens 2009). Therefore, as the majority of massive structures in the universe are predominantly dark matter, lensing provides a novel way to probe the nature of dark matter itself. Weak gravitational lensing (WL) is a regime in which one makes the approximation that lensed sources have (at no time) come radially closer than an Einstein radius to the intervening mass concentrations – which ensures that sources are not multiply imaged. The effect of weak lensing on distant source galaxies is two-fold: the galaxy size is magnified by a convergence field κ and the galaxy ellipticity (third flattening) is perturbed from an underlying intrinsic value by a shearing field γ.

Due to the mass-sheet degeneracy, the weak lensing convergence field is not directly observable. In the weak lensing regime, the shearing field does not suffer such degeneracies and can accurately be modelled from observed ellipticities. Therefore, observations of γ are typically inverted to recover estimators of κ. Such estimators are colloquially named *dark matter mass maps*, and constitute one of the principal observables for cosmology (Clowe et al. 2006). Standard cosmological protocol is to extract weak lensing information in the form of second-order statistics (Kilbinger 2015; Alsing et al. 2016; Taylor et al. 2018), which are then compared to theory. In this approach, mass maps are not required. However, as two-point global statistics are by definition sensitive only to Gaussian contributions, and weak lensing is inherently non-Gaussian, it is informative to consider higher order statistics (Coles & Chiang 2000; Munshi & Coles 2017). Many higher order statistical techniques can be performed directly on mass maps (κ-fields), which motivates investigation into alternate mass-map reconstruction methodologies.

Reconstructing mass maps from shear observations requires solving an ill-posed (often seriously) inverse problem. Many approaches to solving this lensing inverse problem have been developed (e.g. Kaiser & Squires 1993; VanderPlas et al. 2011; Lanusse et al. 2016; Wallis et al. 2017; Chang et al. 2018; Jeffrey et al. 2018), with the industry standard being Kaiser–Squires (KS; Kaiser & Squires 1993). Although these approaches often produce reliable convergence estimators, they lack principled statistical approaches to uncertainty quantification and often assume Gaussianty during the reconstruction process, or post-process by Gaussian smoothing, which is suboptimal when one wishes to analyse small-scale non-Gaussian structure.

★ E-mail: m.price.17@ucl.ac.uk

Most methods refrain from quantifying uncertainties in reconstructions, but those that do often do so by assuming Gaussian priors and adopting Markov chain Monte Carlo (MCMC) techniques (Corless, King & Clowe 2009; Schneider et al. 2015; Alsing et al. 2016).The computational cost of MCMC approaches is large. Recent developments in probability concentration theory have led to advancements in fast approximate uncertainty quantification techniques (Cai, Pereyra & McEwen 2017a,b; Pereyra 2017).

In this article, we present a new mass-mapping formalism. We formulate the lensing inverse problem as a sparse hierarchical Bayesian inference problem from which we derive an unconstrained convex optimization problem. We solve this optimization problem in the analysis setting, with a wavelet-based, sparsity-promoting, l_1-norm prior – similar priors have been shown to be effective in the weak lensing setting (Leonard, Lanusse & Starck 2014; Lanusse et al. 2016; Peel, Lanusse & Starck 2017; Jeffrey et al. 2018). Formulating the problem in this way allows us, for the first time, to recover maximum a posteriori (MAP) estimators, from which we can exploit analytical methods (Cai et al. 2017b; Pereyra 2017) to recover approximate highest posterior density (HPD) credible regions, and perform hypothesis testing of structure in a variety of ways. We apply our algorithm to a range of catalogues drawn from *N*-body simulations – Bolshoi cluster catalogues (Klypin, Trujillo-Gomez & Primack 2011) – and the debated Abell 520 (A520) cluster catalogues (Clowe et al. 2012; Jee et al. 2014). We then demonstrate the aforementioned uncertainty quantification techniques on our MAP reconstructions from these catalogues.

The structure of this article is as follows. In Section 2, we provide a brief overview of the weak lensing paradigm and motivate a sparsity-based approach. In Section 3, we provide the details of our algorithm, as well as some updates to super-resolution image recovery. In Section 4, we present the uncertainty quantification techniques, both mathematically and mechanistically. In Sections 5 and 6, we apply both our reconstruction algorithm and the uncertainty quantification techniques to the aforementioned data sets and analyse the results. Finally, in Section 7, we draw conclusions from this work and propose future avenues of research.

Section 3 relies on a moderate level of understanding in the fields of proximal calculus and compressed sensing, and Section 4 relies on a general understanding of Bayesian inference. As such, for the reader solely interested in practical application of these techniques, we recommend Section 5 onwards.

2 WEAK GRAVITATIONAL LENSING

The following section presents a brief review of the mathematical background relevant to the weak lensing formalism, though a deeper description can be found in popular review articles (Bartelmann & Schneider 2001; Schneider 2005).

2.1 Weak lensing regime

Gravitational lensing refers to the deflection of distant photons as they propagate from their origin to us, the observer. This deflection is caused by local Newtonian potentials that are, in turn, sourced by the total local matter over- or underdensity. As such, weak lensing is sensitive to both the visible and invisible matter distribution – making it an ideal probe of dark matter in the Universe.

The WL regime is satisfied when propagating photons (from a distant source) have an angular position on the source plane β (relative to the line of sight from observer through the lensing mass) greater than the Einstein radius $\theta_{\rm E}$ of the intervening mass. This assertion ensures that the solution of the first-order lens equation is singular:

$$\beta = \theta - \theta_{\rm E}^2 \frac{\theta}{|\theta|^2}, \tag{1}$$

where the Einstein radius is defined to be

$$\theta_{\rm E} = \sqrt{\frac{4GM}{c^2}\frac{f_K(r-r')}{f_K(r)f_K(r')}}, \tag{2}$$

where f_K is the angular diameter distance in a cosmology with curvature K, c is the speed of light in a vacuum, G is the gravitational constant, and M is the lensing mass. Perhaps more generally, the weak lensing regime can be defined as convergence fields for which $\kappa \ll 1$ – ensuring that the shear signal remains linear.

Due to the sparse nature of the distribution of galaxies across the sky, most sources are (to a good approximation) within the WL regime. The WL effect is best expressed in terms of a lensing potential ϕ, defined to be the integral of the Newtonian potential Φ along a given line of sight:

$$\phi(r,\omega) = \frac{2}{c^2}\int_0^r {\rm d}r' \frac{f_K(r-r')}{f_K(r)f_K(r')}\Phi(r',\omega), \tag{3}$$

where r and $r^{'}$ are comoving distances, and $\omega = (\theta, \psi)$ are angular spherical coordinates. The local Newtonian potential must satisfy the Poisson equation and as such is related to the matter overdensity field:

$$\nabla^2\Phi(r,\omega) = \frac{3\Omega_{\rm M}H_0^2}{2a(r)}\delta(r,\omega), \tag{4}$$

where $\Omega_{\rm M}$ is the matter density parameter, H_0 is the current Hubble constant, $a(r)$ is the scale factor, and δ is the fractional overdensity.

To first order, there are two primary ways in which light from distant sources is distorted by this lensing potential. Images are magnified by a spin-0 convergence field κ and sheared by a spin-2 shear field γ. These quantities can be shown (Bartelmann & Schneider 2001) to be related to the lensing potential by

$$\kappa(r,\omega) = \frac{1}{4}(\eth\bar{\eth} + \bar{\eth}\eth)\,\phi(r,\omega), \tag{5}$$

$$\gamma(r,\omega) = \frac{1}{2}\eth\eth\,\phi(r,\omega), \tag{6}$$

where $\eth$ and $\bar{\eth}$ are the spin s raising and lowering operators respectively and are in general defined to be,

$$\eth \equiv -\sin^s\theta\left(\frac{\partial}{\partial\theta} + \frac{i\partial}{\sin\theta\partial\psi}\right)\sin^{-s}\theta, \tag{7}$$

$$\bar{\eth} \equiv -\sin^{-s}\theta\left(\frac{\partial}{\partial\theta} - \frac{i\partial}{\sin\theta\partial\psi}\right)\sin^{s}\theta. \tag{8}$$

where we have omitted spin subscripts for clarity.

2.2 Standard mass-mapping techniques

Typically, we wish to make inferences about the projected matter overdensity $\delta(r, \omega)$ that is most directly accessible by inverting the integral equation (Schneider 2005)

$$\kappa(r,\omega) = \frac{3\Omega_{\rm M}H_0^2}{2c^2}\int_0^r {\rm d}r' \frac{f_K(r')f_K(r-r')}{f_K(r)}\frac{\delta(f_K(r')r',r')}{a(r)}. \tag{9}$$

This poses a difficulty as the convergence κ is only determined to the degeneracy $\kappa \rightarrow \kappa^{'} = \eta\kappa + (1-\eta)$ and is therefore not

directly observable – this degeneracy is often referred to as the *mass-sheet degeneracy*. However, as the intrinsic ellipticity distribution of galaxies has zero mean, if one averages many galaxy ellipticities within a given pixel the true shear γ can be recovered – which makes γ an observable field. As such one typically collects observations of γ that are and subsequently used to construct estimators of κ.

For small sky fractions, we can approximate the field of view as a plane (though this approximation degrades quickly with sky fraction; Wallis et al. 2017). In this planar approximation, ð and $\bar{\text{ð}}$ reduce to

$$\text{ð} \approx -(\partial_x + i\partial_y) \quad \text{and} \quad \bar{\text{ð}} \approx -(\partial_x - i\partial_y). \tag{10}$$

Combining equations (5) and (6), we find the planar forward model in Fourier space:

$$\hat{\gamma}(k_x, k_y) = \mathbf{D}_{k_x,k_y}\hat{\kappa}(k_x, k_y), \tag{11}$$

with the mapping operator being,

$$\mathbf{D}_{k_x,k_y} = \frac{k_x^2 - k_y^2 + 2ik_xk_y}{k_x^2 + k_y^2}. \tag{12}$$

Hereafter, we drop the k_x, k_y subscripts for clarity. It is informative to note that this forward model is undefined at the origin ($k=\sqrt{k_x^2+k_y^2}=0$) – which corresponds to the mass-sheet degeneracy (Bartelmann & Schneider 2001)

The most naive inversion of this forward model is KS inversion,

$$\hat{\kappa}^{\rm KS} = \mathbf{D}^{-1}\hat{\gamma}, \tag{13}$$

which is direct inversion in Fourier space (Kaiser & Squires 1993). KS inversion of the forward model, given by equation (11), performs adequately, provided the space over which it is defined is complete, and the sky fraction is small. However, masking and survey boundaries are inherent in typical WL surveys, leading to significant contamination of the KS estimator. Often maps recovered with the KS estimator are convolved with a Gaussian kernel to reduce the impact of these contaminations but this is suboptimal. This smooths away a large fraction of the small-scale non-Gaussian information, which cosmologists are increasingly interested in extracting from WL surveys.

3 SPARSE MAP ESTIMATORS

Several alternate approaches for solving the inverse problem between convergence κ and shear γ that do not assume or impose Gaussianity have been proposed, some of which are based on the concept of wavelets and sparsity (Pires, Starck & Amara 2009; Jullo et al. 2014; Lanusse et al. 2016; Peel et al. 2017).

We propose a mass-mapping algorithm that relies on sparsity in a given wavelet dictionary. Moreover, we formulate the problem such that we can exploit recent developments in the theory of probability concentration, which have been developed further to produce novel uncertainty quantification techniques (Pereyra 2017). Crucially, this allows us to recover principled statistical uncertainties on our MAP reconstructions (as in Cai et al. 2017a,b) as will be discussed in detail in the following section.

As mentioned previously, galaxies have an intrinsic ellipticity. To mitigate the effect of intrinsic ellipticity we choose to project the ellipticity measurements on to a grid and average. If we assume that galaxies have no preferential orientation in the absence of lensing effects, then the average intrinsic ellipticity tends to zero. This is a good approximation for the purposes of this paper, but weak correlation between the intrinsic alignments of galaxies has been observed (Troxel & Ishak 2015; Piras et al. 2018).

3.1 Hierarchical Bayesian framework

Hierarchical Bayesian inference provides a rigorous mathematical framework through which theoretically optimal solutions can be recovered. Moreover, it allows one to construct measures of the uncertainty on recovered point estimates.

As is common for hierarchical Bayesian models, we begin from Bayes' theorem for the posterior distribution,

$$p(\kappa|\gamma) = \frac{p(\gamma|\kappa)p(\kappa)}{\int_{\mathbb{C}^N} p(\gamma|\kappa)p(\kappa)\mathrm{d}\kappa}, \tag{14}$$

where $p(\gamma|\kappa)$ is the likelihood function representing data fidelity, N is the dimensionality of κ, and $p(\kappa)$ is a prior on the statistical nature of κ. The denominator is called the *Bayesian evidence* that is constant and so can be dropped for our purposes. Typically, the Bayesian evidence is used for model comparison, which we will not be considering within the context of this paper. Given Bayes' theorem, and the monotonicity of the logarithm function, we can easily show that the maximum posterior solution is defined by,

$$\underset{\kappa}{\text{argmax}}\{p(\kappa|\gamma)\} = \underset{\kappa}{\text{argmin}}\{-\log(\ p(\kappa|\gamma)\ )\}. \tag{15}$$

This step is crucial, as it allows us to solve the more straightforward problem of minimizing the log-posterior rather than maximizing the full posterior. Conveniently, in most physical situations the operators associated with the log-posterior are convex. Drawing from the field of convex optimization, the optimal solution for the posterior can be recovered extremely quickly – even in high dimensional settings.

3.2 Sparsity and inverse problems

Let $\gamma \in \mathbb{C}^M$ be the discretized complex shear field extracted from an underlying discretized convergence field $\kappa \in \mathbb{C}^N$ by a *measurement operator* $\mathbf{\Phi} \in \mathbb{C}^{M\times N} : \kappa \mapsto \gamma$. In the planar setting, $\mathbf{\Phi}$ can be modelled by,

$$\mathbf{\Phi} = \mathbf{M}\mathbf{F}^{-1}\mathbf{D}\mathbf{F}. \tag{16}$$

Here, $\mathbf{F}$ is the discrete fast Fourier transform (FFT), $\mathbf{F}^{-1}$ is the inverse discrete fast Fourier transform (IFFT), $\mathbf{M}$ is a standard masking operator, and $\mathbf{D}$ is a diagonal matrix applying the scaling of the forward model in Fourier space as defined in equation (12). In the case of independent and identically distributed *i.i.d.* Gaussian noise, measurement of γ will be contaminated such that

$$\gamma = \mathbf{\Phi}\kappa + \mathcal{N}(0, \sigma_i^2), \tag{17}$$

where $\mathcal{N}(0, \sigma_i^2) \in \mathbb{C}^M$ is additive i.i.d. Gaussian noise of variance σ_i^2 for pixel i. Often in WL experiments, the total number of binned measurements is less than the number of pixels to be recovered, $M < N$, and the inverse problem becomes ill-posed.

In such a setting, the Bayesian likelihood function (data fidelity term) is given by the product of Gaussian likelihoods defined on each pixel with pixel noise variance σ_i^2, which is to say an overall multivariate Gaussian likelihood of known covariance $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_M) \in \mathbb{R}^{M\times M}$. Let $\Phi_i\kappa$ be the value of $\mathbf{\Phi}\kappa$ at

pixel i, then the overall likelihood is then defined as,

$$p(\gamma|\kappa) \propto \prod_{i=0}^{M} \exp\left(\frac{-(\Phi_i\kappa - \gamma_i)^2}{2\sigma_i^2}\right) = \prod_{i=0}^{M} \exp\left(\frac{-1}{2}(\bar{\Phi}_i\kappa - \bar{\gamma}_i)^2\right),$$
$$= p(\gamma|\kappa) \propto \exp\left(\frac{-\|\bar{\boldsymbol{\Phi}}\kappa - \bar{\gamma}\|_2^2}{2}\right), \quad (18)$$

where $\|\cdot\|_2$ is the ℓ_2-norm and $\bar{\boldsymbol{\Phi}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Phi}$ is a composition of the measurement operator and an inverse covariance weighting. Effectively, this covariance weighting leads to measurements $\bar{\gamma} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\gamma$ that whiten the typically non-uniform noise variance in the observational data γ.

This likelihood function allows one to map from the number count of observations per pixel to a corresponding noise variance (assuming an intrinsic ellipticity dispersion of $\sim$0.37), from which the noise (under and central limit theory argument of Gaussianity) may be correctly incorporated into the reconstruction. In practice, this requires only the number density of observations per pixel, which is trivially inferred from raw observational data catalogues.

To regularize this inverse problem, we then define a sparsity promoting Laplace-type prior:

$$p(\kappa) \propto \exp(-\mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1), \quad (19)$$

where $\boldsymbol{\Psi}$ is an appropriately selected wavelet dictionary, and $\mu \in \mathbb{R}_+$ is a regularization parameter – effectively a weighting between likelihood and prior. Note that one may choose any convex log-prior within this formalism e.g. an ℓ_2-norm prior from which one essentially recovers Weiner filtering (see Padmanabhan, Seljak & Pen 2003; Horowitz, Seljak & Aslanyan 2018, for alternate iterative Weiner filtering approaches). From equations (14) and (15), the unconstrained optimization problem that minimizes the log-posterior is,

$$\kappa^{\rm map} = \underset{\kappa}{\rm argmin}\left\{\mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1 + \frac{\|\bar{\boldsymbol{\Phi}}\kappa - \gamma\|_2^2}{2}\right\}, \quad (20)$$

where the bracketed term is called the *objective function*. To solve this convex optimization problem, we adopt a forward–backward splitting algorithm (e.g. Combettes & Pesquet 2009). A full description of this algorithm applied in the current context is outlined in Cai et al. (2017b).

Let $f(\kappa) = \mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1$ denote our prior term, and $g(\kappa) = \|\bar{\boldsymbol{\Phi}}\kappa - \gamma\|_2^2/2$ denote our data fidelity term. Then our optimization problem can be re-written compactly as,

$$\underset{\kappa}{\rm argmin}\left\{f(\kappa) + g(\kappa)\right\}. \quad (21)$$

The forward–backward iteration step is then defined to be,

$$\kappa^{(i+1)} = {\rm prox}_{\lambda^{(i)}f}(\kappa^{(i)} - \lambda^{(i)}\nabla g(\kappa^{(i)})), \quad (22)$$

for iteration i, with gradient,

$$\nabla g(\kappa) = \bar{\boldsymbol{\Phi}}^\dagger(\boldsymbol{\Phi}\kappa - \gamma). \quad (23)$$

If the wavelet dictionary $\boldsymbol{\Psi}$ is a tight frame (i.e. $\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} = \mathbb{I}$), the proximity operator is given by,

$${\rm prox}_{\lambda f}(z) = z + \boldsymbol{\Psi}({\rm soft}_{\lambda\mu}(\boldsymbol{\Psi}^\dagger z) - \boldsymbol{\Psi}^\dagger z), \quad (24)$$

where ${\rm soft}_\lambda(z)$ is the point-wise soft-thresholding operator (Combettes & Pesquet 2009) and λ is a parameter related to the step size (which is in turn related to the Lipschitz differentiability of the log-prior) that should be set according to Cai et al. (2017b). The iterative algorithm is given explicitly in the primary iterations of algorithm 1. Adaptations for frames that are not tight can be found in Cai et al. (2017b) and are readily available within our framework.

Our algorithm has distinct similarities to the GLIMPSE algorithm presented by Lanusse et al. (2016), but crucially differs in several aspects. Most importantly we formulate the problem in a hierarchical Bayesian framework that allows us to recover principled statistical uncertainties. In addition to this we include Bayesian inference of the regularization parameter, a robust estimate of the noise level (which can be folded into the hierarchical model), and we use super-resolution operators instead of non-discrete fast Fourier transforms.

Algorithm 1 Forward-backward analysis algorithm

Input: $\gamma \in \mathbb{C}^M$, $\kappa^{(0)} \in \mathbb{C}^N$, λ, $\mu^{(0)} = i = t = 0$, $T_1, T_2 \in \mathbb{R}_+$
Output: $\kappa^{\rm map} \in \mathbb{C}^N$, $\mu \in \mathbb{R}_+$
Precomputation:
Do:
1: Calculate $\kappa^{(t)} = \underset{\kappa}{\rm argmin}\left\{f(\kappa) + g(\kappa)\right\}$,
2: Update $\mu^{(t+1)} = \frac{(N/k)+\alpha-1}{f(\kappa^{(t)})+\beta}$,
3: $t = t + 1$,
4: On convergence, μ becomes fixed.
Until: Iteration limit reached.
Primary Iterations:
Do:
1: update $\nu^{(i+1)} = \kappa^{(i)} - \lambda\bar{\boldsymbol{\Phi}}^\dagger(\bar{\boldsymbol{\Phi}}\kappa^{(i)} - \gamma)$,
2: compute $\eta = \boldsymbol{\Psi}^\dagger\nu^{(i+1)}$,
3: update $\kappa^{(i+1)} = \nu^{(i+1)} + \boldsymbol{\Psi}({\rm soft}_{\lambda\mu}(\eta) - \eta)$,
4: $i = i + 1$.
Until: Stopping criterion satisfied.
i.e. $\frac{\|\kappa^{(i)} - \kappa^{(i+1)}\|_2}{\|\kappa^{(i)}\|_2} < T_1$ and $\frac{{\rm obj}(\kappa^{(i)}) - {\rm obj}(\kappa^{(i+1)})}{{\rm obj}(\kappa^{(i)})} < T_2$.

3.3 Reduced shear

Due to a degeneracy between γ and κ, the true observable quantity is in fact the *reduced shear* g (Bartelmann & Schneider 2001),

$$g = \frac{\gamma}{1 - \kappa}. \quad (25)$$

Deep in the weak lensing regime one can safely approximate $\gamma \approx g \ll 1$, which ensures that the optimization problem remains linear. However, when reconstructing regions close to massive structures (galaxy clusters) this approximation is no longer strictly valid and we must unravel this additional factor. We adopt the procedure outlined in Wallis et al. (2017), which we also outline schematically in Fig. 1 – this method can be found in detail in Mediavilla et al. (2016, p. 153). We find that these corrections typically converge after $\sim$5–10 iterations.

3.4 Regularization parameter selection

One key issue of sparsity-based reconstruction methods is the selection of the regularization parameter μ. Several methodologies have arisen (Paykari et al. 2014; Lanusse et al. 2016; Peel et al. 2017; Jeffrey et al. 2018) for selecting μ, though often the regularization parameter is chosen somewhat arbitrarily – as the integrity of the MAP solution is assumed to be weakly dependent on the choice of μ. However, to extract principled statistical uncertainties on the recovered images, one must select this parameter in a principled statistical manner.

We apply the hierarchical Bayesian formalism developed by Pereyra, Bioucas-Dias & Figueiredo (2015) – the details of which are

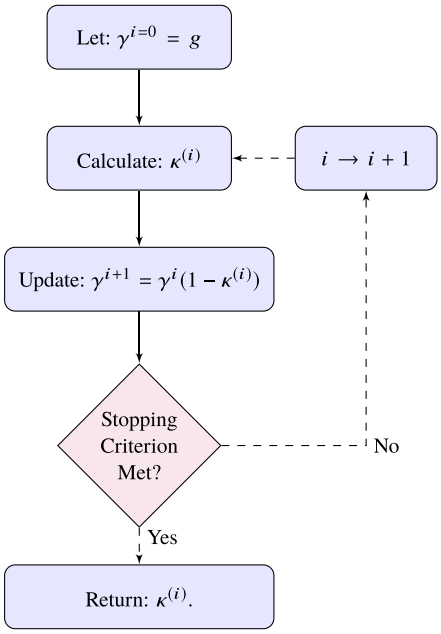

Figure 1. Schematic of reduced shear iterations. An initial guess of the MAP solution $\kappa_i^{\rm map}$ is constructed; the current best shear estimates γ_i are then used in tandem to construct a new estimate of the true shear field γ_{i+1}.

elegantly presented by the authors. Though we will outline roughly the underlying argument here.

First define a sufficient statistic f to be k-homogeneous if $\exists\ k \in \mathbb{R}_+$ such that,

$$f(\eta x) = \eta^k f(x),\ \forall x \in \mathbb{R}^N,\ \forall \eta > 0. \tag{26}$$

All norms, composite norms and composition of norms with linear operators are 1-homogeneous – and so our ℓ_1-norm has k of 1. If a sufficient statistic f is k-homogeneous, then the normalization factor $C(\mu)$ of $p(\kappa|\mu)$ is given by (Pereyra et al. 2015),

$$C(\mu) = A\mu^{-N/k}, \tag{27}$$

where A is a constant independent from μ. The proposed Bayesian inference model then implements a gamma-type hyper-prior – which is a typical hyper-prior for scale parameters,

$$p(\mu) = \frac{\beta^\alpha}{\Gamma(\alpha)}\mu^{\alpha-1}e^{-\beta\mu}\mathbb{I}_{\mathbb{R}_+}(\mu), \tag{28}$$

where without loss of generality $\alpha = \beta = 1$. The result is effectively insensitive to their value (in numerical experiments values of $\alpha, \beta \in [10^{-2}, 10^5]$ produced essentially no difference in μ).

Now, let us extend the inference problem of the log-posterior to the case where μ is an additionally unknown parameter. In this context, we compute the joint MAP estimator $(\kappa^{\rm map}, \mu^{\rm map}) \in \mathbb{C}^N \times \mathbb{R}_+$ which maximizes $p(\kappa, \mu|\gamma)$ such that,

$$\mathbf{0}_{N+1} \in \partial_{\kappa,\mu} \log p(\kappa^{\rm map}, \mu^{\rm map}|\gamma), \tag{29}$$

where $\mathbf{0}_i$ is the i-dimensional null vector and $\partial_s h(s')$ is the set of subgradients of function $h(s)$ at s'. This in turn implies both that,

$$\mathbf{0}_N \in \partial_\kappa \log p(\kappa^{\rm map}, \mu^{\rm map}|\gamma), \tag{30}$$

and

$$\mathbf{0} \in \partial_\mu \log p(\kappa^{\rm map}, \mu^{\rm map}|\gamma). \tag{31}$$

From equation (30), we recover the optimization problem with known regularization parameter μ given in equation (20). However, from equations (27), (28), and (31), it follows that the MAP regularization parameter μ is given by (Pereyra et al. 2015),

$$\mu^{\rm map} = \frac{\frac{N}{k} + \alpha - 1}{f(\kappa^{\rm map}) + \beta}, \tag{32}$$

where we recall that N is the total dimension of our convergence space.

It is precisely this optimal μ value that we wish to use in our hierarchical Bayesian model. Hereafter, we drop the map superscript from μ for clarity. To calculate μ we perform preliminary iterations defined by

$$\kappa^{(t)} = \underset{\kappa}{\rm argmin}\{f(\kappa; \mu^{(t)}) + g(\kappa)\}, \tag{33}$$

where $g(\kappa)$ is our likelihood term and,

$$\mu^{(t+1)} = \frac{\frac{N}{k} + \alpha - 1}{f(\kappa^{(t)}) + \beta}. \tag{34}$$

Typically, we find that these preliminary iterations take ∼5–10 iterations to converge, and recover close to optimal parameter selection for a range of test cases – note that here the optimal selection of μ is that which maximizes the SNR of a recovered image.

Another factor that can influence the quality of reconstructions is the selection of wavelet dictionary. In this paper, we consider Daubechies (eight levels) and SARA dictionaries (Carrillo, McEwen & Wiaux 2012; Carrillo et al. 2013), though a wide variety of wavelet dictionaries exist, see e.g. starlets (Starck, Murtagh & Fadili 2015). The eight-level SARA dictionary is a combination of the Dirac and Daubechies 1–8 wavelet dictionaries. It is important to note that we use the SARA dictionary, not the complete SARA scheme (Carrillo et al. 2012, 2013), which involves an iterative re-weighting scheme that is not considered here.

3.5 Super-resolution image recovery

Gridding of weak lensing data is advantageous in that it can provide a good understanding of the noise properties – a necessary feature for principled uncertainty quantification. However, an inherent drawback of projecting data into a grid is the possibility of creating an incomplete space due to low sampling density – often referred to as masking. Decomposition of spin signals on bounded manifolds is inherently degenerate (Bunn et al. 2003b); specifically the orthogonality of eigenfunctions is locally lost at the manifold boundaries, leading to signal leakage between Fourier (or on the sphere, harmonic) modes.

One approach to mitigate this problem is to avoid the necessity of gridding by substituting a *non-uniform discrete Fourier transform* (NFFT) into the RHS of equation (16) as presented by Lanusse et al. (2016). A downside of this NFFT approach is that the noise is more difficult to handle, leading to complications when considering uncertainty quantification. Another approach is to perform super-resolution image recovery, which we present in the context of our algorithm.

Suppose the dimension of our gridded measurement space is M, as before, and the desired dimension of our solution space is N', where $N' \geq N$. In this setting, we have shear measurements $\gamma \in \mathbb{C}^M$ and recovered convergence $\kappa \in \mathbb{C}^{N'}$. Let us now define a *super-resolution* (subscript SR) measurement operator to be,

$$\mathbf{\Phi}_{\rm SR} = \mathbf{M}\,\mathbf{F}_{\rm lr}^{-1}\,\mathbf{D}\,\mathbf{Z}\,\mathbf{F}_{\rm hr} \tag{35}$$

where $\mathbf{F}_{\rm hr}$ is a high-resolution (dimension N') fast Fourier transform, $\mathbf{Z} \in \mathbb{C}^{N\times N'}$ is a Fourier space down-sampling which maps $\tilde{\kappa}' \in \mathbb{C}^{N'}$ on to $\tilde{\kappa} \in \mathbb{C}^{N}$, where tilde represents Fourier coefficients, $\mathbf{D}$ is the planar forward model given by equation (11), and $\mathbf{M}$ is a standard masking operator. Finally, $\mathbf{F}_{\rm lr}^{-1}$ is a low resolution (dimension M) inverse fast Fourier transform. For completeness, the super-resolution adjoint measurement operator is given by,

$$\mathbf{\Phi}^{\dagger}_{\rm SR} = \mathbf{F}_{\rm hr}^{-1}\, \mathbf{Z}^{\dagger}\, \mathbf{D}^{\dagger}\, \mathbf{F}_{\rm lr}\, \mathbf{M}^{\dagger}, \tag{36}$$

where $\mathbf{M}^{\dagger}$ is adjoint masking (gridding), $\mathbf{D}^{\dagger}$ is the adjoint of $\mathbf{D}$ (which is self-adjoint hence $\mathbf{D}^{\dagger} = \mathbf{D}$), and $\mathbf{Z}^{\dagger} \in \mathbb{C}^{M'\times M}$ is zero padding in Fourier space that acts by mapping $\tilde{\gamma} \in \mathbb{C}^{M}$ to $\tilde{\gamma}' \in \mathbb{C}^{M'}$. Note that when considering the KS estimate in the super-resolution setting a rescaling function to account for the different Fourier normalization factors must be introduced (which we absorb into the Fourier operators). As before, this super-resolution measurement operator is concatenated with the inverse covariance weighting to form an analogous composite operator $\bar{\mathbf{\Phi}}_{\rm SR}$ which is used throughout the following analysis.

Conceptually super-resolution allows partial inpainting of higher resolution Fourier modes. In this way, one is able to recover high-resolution structure for images from comparatively low resolution data sets. Such high-resolution structure is of course dependent on the prior information injected when solving the inverse problem. Interestingly this raises another consideration: in scenarios where the pixel-level observation count is very low the noise level dilutes high-frequency components and can limit the efficacy of reconstruction algorithms. In such a setting, gridding observational data on to a lower resolution map, with inherently lower pixel-level noise, and performing a super-resolution reconstruction can recover far better estimates of the high-frequency modes, and thus often recovers greater reconstruction fidelity.

4 BAYESIAN UNCERTAINTY QUANTIFICATION

Estimators recovered from algorithms of the form presented in the previous section are MAP solutions to, in general, ill-conditioned inverse problems, and as such have significant intrinsic uncertainty. Theoretically, MCMC techniques could be applied to recover the complete posterior in the context of Gaussian (Schneider et al. 2015; Alsing et al. 2016) and sparsity-promoting (Pereyra 2013; Cai et al. 2017a) priors but these approaches are computationally demanding for high-dimensional problems where N is large. As N can easily be larger than 10^6 (e.g. when considering 1024×1024 resolution images), MCMC approaches are often not feasible.

In Pereyra (2017), a methodology based on probability concentration is presented, which uses MAP estimators to estimate theoretically conservative approximate Bayesian credible regions (specifically HPD credible regions) of the posterior, $p(\kappa|\gamma)$. As this approach requires only knowledge of the MAP solution and the objective function, the Bayesian credible regions can be approximated efficiently in high-dimensional settings.

4.1 Highest posterior density regions

A posterior credible region at confidence level $100(1-\alpha)$ per cent is a subset $C_\alpha \in \mathbb{C}^N$ that satisfies the integral,

$$p(\kappa \in C_\alpha|\gamma) = \int_{\kappa\in\mathbb{C}^N} p(\kappa|\gamma)\mathbb{I}_{C_\alpha}\, {\rm d}\kappa = 1-\alpha, \tag{37}$$

where $\mathbb{I}_{C_\alpha}$ is the set indicator function for C_α defined by $\mathbb{I}_{C_\alpha}(\kappa) = 1\ \forall \kappa \in C_\alpha$ and 0 elsewhere. One possible region which satisfies this property is the HPD region defined by,

$$C_\alpha := \{\kappa : f(\kappa) + g(\kappa) \leq \epsilon_\alpha\}, \tag{38}$$

where ϵ_α defines an iso-contour (i.e. level-set) of the log-posterior set such that the integral in (37) is satisfied. This region can be shown (Robert 2001) to have minimum volume and is thus decision-theoretically optimal. However, due to the dimensionality of the integral in (37) calculation of the HPD credible region is difficult. A conservative approximation of C_α was recently proposed (Pereyra 2017) and shown to be effective in the inverse imaging setting of radio interferometric imaging (Cai et al. 2017b). This approximate HPD is defined by

$$C'_\alpha := \{\kappa : f(\kappa) + g(\kappa) \leq \epsilon'_\alpha\}, \tag{39}$$

where the approximate threshold ϵ'_α is given by

$$\epsilon'_\alpha = f(\kappa^{\rm map}) + g(\kappa^{\rm map}) + \tau_\alpha\sqrt{N} + N, \tag{40}$$

with constant $\tau_\alpha = \sqrt{16\log(3/\alpha)}$. For a detailed derivation of this approximation, see Pereyra (2017). Provided $\alpha \in [4\exp(-N/3), 1]$ the deviation of this adapted threshold is bounded and grows at most linearly with respect to N. The error of this approximate threshold is bounded by

$$0 \leq \epsilon'_\alpha - \epsilon_\alpha \leq \eta_\alpha\sqrt{N} + N, \tag{41}$$

where $\eta_\alpha = \sqrt{16\log(3/\alpha)} + \sqrt{1/\alpha}$. In high-dimensional settings (N large), this error may naively appear large; however, in practice the error is relatively small.

4.2 Hypothesis testing

Extending the concept of HPD credible regions, one can perform *knock-out* hypothesis testing of the posterior to determine the physicality of recovered structure (Cai et al. 2017b).

To perform such tests, one first creates a surrogate image $\kappa^{\rm sgt}$ by masking a feature of interest $\Omega_D \subset \Omega$ in the MAP estimator $\kappa^{\rm map}$. It is then sufficient to check if,

$$f(\kappa^{\rm sgt}) + g(\kappa^{\rm sgt}) \leq \epsilon'_\alpha. \tag{42}$$

If this inequality holds, we interpret that the physicality of Ω_D is undetermined and so no strong statistical statement can be made. Should the objective function evaluated at $\kappa^{\rm sgt}$ be larger than ϵ'_α then it no longer belongs to the approximate credible set C'_α and therefore (as ϵ'_α is conservative) it cannot belong to the HPD credible set C_α. Therefore, for $\kappa^{\rm sgt}$ that do not satisfy the above inequality we determine the structure Ω_D to be strictly physical at $100(1-\alpha)$ per cent confidence level. A schematic of hypothesis testing is provided in Fig. 2.

In pixel-space, we begin by masking out a feature of interest, creating a rough surrogate image – setting the pixels associated with a selected structure to 0 – this rough surrogate is then passed through an appropriate wavelet filter Λ as part of *segmentation-inpainting* to replace generic background structure into the masked region. Mathematically, this amounts to the iterations,

$$\kappa^{(i+1),{\rm sgt}} = \kappa^{\rm map}\mathbb{I}_{\Omega-\Omega_D} + \Lambda^{\dagger}{\rm soft}_{\lambda_t}(\Lambda\kappa^{(i),{\rm sgt}})\mathbb{I}_{\Omega_D}, \tag{43}$$

where $\Omega_D \subset \Omega$ is the subset of masked pixels, $\mathbb{I}_{\Omega-\Omega_D}$ is the set indicator function, and λ_t is a thresholding parameter that should be chosen appropriately for the image.

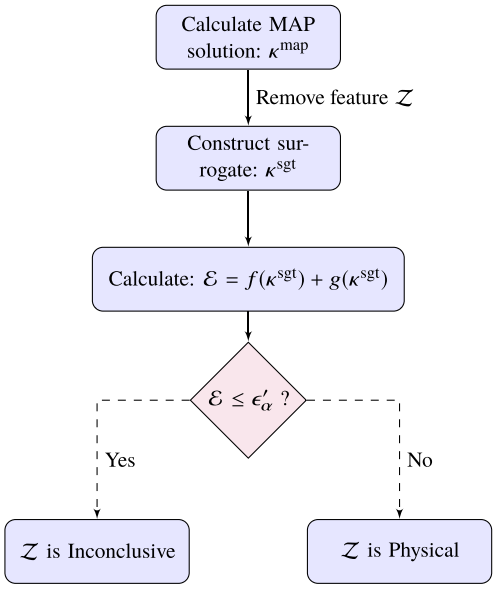

Figure 2. Schematic of hypothesis testing. The feature $\mathcal{Z}$ is entirely general and can be constructed by any well-defined operator on the MAP solution $\kappa^{\rm map}$.

A second straightforward method for generating surrogate images is to blur local pixel substructure into one collective structure – in a process called *segmentation-smoothing*. This approach provides a simple way to determine if the substructure in a given region is physical or likely to be an artefact of the reconstruction process.

For example, if several massive peaks are located near one another, one can blur these structures into a single cohesive peak. This would be useful when considering peak statistics on convergence maps – which is often used to constrain the cosmological parameters associated with dark matter.

One can conduct such blurring of structure by: specifying a subset of the reconstructed pixels $\Omega_D \subset \Omega$; convolving $\kappa^{\rm map}$ with a Gaussian smoothing kernel; and replacing pixels that belong to Ω_D with their smoothed counterparts. This can be displayed algorithmically as,

$$\kappa^{\rm sgt} = \kappa^{\rm map}\mathbb{I}_{\Omega-\Omega_D} + (\kappa^{\rm map} * \mathcal{G}(0,\chi))\mathbb{I}_{\Omega_D}, \tag{44}$$

where $\mathcal{G}(0,\chi)$ is a chosen Gaussian smoothing kernel and $*$ is a trivially extended 2D version of the usual 1D Fourier convolution operator,

In the scope of this paper, we focus primarily on pixel-space features, but it is important to stress that *knock-out* approach is entirely general and can be applied to any well-defined feature of an MAP estimator – i.e. masking certain Fourier space features, removal of global small-scale structure, etc.

5 ILLUSTRATION ON SIMULATIONS

We now consider a selection of realistic simulations to illustrate our sparse reconstruction method on cluster scales that are particularly challenging for myriad factors. Further to this, we showcase the aforementioned uncertainty quantification methods in a variety of idealized cluster scale MAP reconstructions. We place emphasis on uncertainty quantification rather that the reconstruction fidelity.

5.1 Data sets

In this paper, we focus primarily on four large clusters (those with significant friends of friends, i.e. significant substructure) extracted from the Bolshoi *N*-body simulation (Klypin et al. 2011). On the cluster scale, we showcase our formalism on a variety of Bolshoi *N*-body simulation data sets. The Bolshoi *N*-body cluster simulation catalogues we work with in this paper are those used in Lanusse et al. (2016), which were extracted using the CosmoSim web-tool.[1] Construction of these weak lensing realizations assumed a redshift of 0.3, with a 10×10 arcmin^2 field of view, and have convergence normalized with respect to lensing sources at infinity. Explicitly this results in pixel-dimensions of $\sim$2.5 arcsec. Due to the relatively low particle density, these images were subsequently denoized by a multiscale Poisson denoizing algorithm.

5.2 Methodology

Typically, we begin by creating an artificial shear field $\hat{\gamma} \in \mathbb{C}^M$ from a known *ground-truth* convergence field κ, that is extracted from a given data set. This is a common approach in the imaging community and presents a closed scenario in which the true input is known. These $\hat{\gamma}$ fields are created by,

$$\hat{\gamma} = \mathbf{\Phi}\kappa + \mathcal{N}\left(0, \sigma_i^2\right), \tag{45}$$

where σ_i (i.e. the noise covariance) is determined entirely from a predefined number density of observations $n_{\rm gal}$ per arcmin^2, an assumed intrinsic ellipticity dispersion of 0.37, and the resolution of the images (in this case 10 arcmin $\times$ 10 arcmin). In this way, the noise can be tuned to directly mimic that present in practical settings. Using the simulated noise covariance (which in practice would be provided by the observation team), we then utilize the SOPT[2] framework to perform our reconstruction algorithm on $\hat{\gamma}$ such that we recover a MAP estimator of the convergence $\kappa^{\rm map}$. From this reconstructed convergence field, a recovered SNR is computed and a selection of hypothesis tests are conducted to showcase the power of this formalism.

In the case where the underlying clean γ are unavailable (i.e. application to A520 data), we conduct the same analysis as before but instead of creating artificial noisy $\hat{\gamma}$ maps we used the real noisy observational data.

Throughout our analysis the recovered SNR (dB) is defined to be,

$$\text{SNR} = 20 \times \log_{10}\left(\frac{\|\kappa\|_2}{\|\kappa - \kappa^{\rm map}\|_2}\right), \tag{46}$$

when the ground-truth convergence is known. Furthermore, we quantify the topological similarity between the true convergence and the estimator via the Pearson correlation coefficient that is defined to be

$$r = \frac{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa^{\rm map}(i) - \bar{\kappa}^{\rm map}\}\{\kappa(i) - \bar{\kappa}\}}{\sqrt{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa^{\rm map}(i) - \bar{\kappa}^{\rm map}\}^2}\sqrt{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa(i) - \bar{\kappa}\}^2}}, \tag{47}$$

where $\bar{x} = \langle x \rangle$. The correlation coefficient $r \in [-1, 1]$ quantifies the structural similarity between two data sets: 1 indicates maximal positive correlation, 0 indicates no correlation, and -1 indicates maximal negative correlation.

[1] https://www.cosmosim.org

[2] A highly optimized sparse optimization solver (https://github.com/astro-informatics/SOPT).

Table 1. Contains both reconstruction SNR and Pearson correlation coefficient (topological correlation) metrics for the raw KS (no smoothing), an optimally smoothed KS (grid search for smoothing kernel which maximizes the recovered SNR), and our sparse reconstructions of the Bolshoi-3 cluster simulated with realistic noise derived from the presented number density of galaxy observations $n_{\rm gal}$. The difference column is calculated as the difference between the Sparse and smoothed KS recovered SNR. Note that dB is a logarithmic scale therefore increases of ∼20 dB are extreme reductions in RMS error.

Input $n_{\rm gal}$	KS	KS Smooth	Sparse	Difference
		SNR (dB)		
500	2.917	6.276	27.506	+21.230
100	− 4.497	5.774	21.955	+16.181
30	− 10.400	5.340	21.462	+16.122
10	− 15.970	5.041	14.409	+9.368
		Pearson correlation		
500	0.166	0.902	0.977	+0.075
100	0.076	0.796	0.970	+0.174
30	0.039	0.689	0.955	+0.266
10	0.029	0.716	0.949	+0.233

5.3 Bolshoi cluster catalogues

The Bolshoi cluster data used consists of four large clusters extracted from the Bolshoi *N*-body simulation (Klypin et al. 2011; Lanusse et al. 2016). These images were then multiscale Poisson denoized to create suitable ground truth simulations. We choose to analyse the same clusters considered in Lanusse et al. (2016), as they showcase a wide variety of structure on all scales. Hereafter, we restrict ourselves to the SARA dictionary (Carrillo et al. 2012) truncated at the 4th Daubechies wavelet (DB4) for simplicity – i.e. the combination of the Dirac, and DB1 to DB4 wavelet dictionaries only.

To investigate the SNR gain of our formalism over KS in the cluster scale setting, we created realizations of noisy pseudo-shear maps for assumed number density of galaxy observations $n_{\rm gal} \in [500, 100, 30, 10]$ from one Bolshoi cluster map, upon which we applied our reconstruction algorithm pipeline. The results of which are presented in Table 1. It should be noted that for comparisons sake the KS estimate without convolution with a Gaussian smoothing kernel is provided in addition to an optimally smoothed KS estimator. This has been done to highlight the difference in reconstruction fidelity between the raw KS estimator and the KS estimator after post-processing (Gaussian smoothing), a discrepancy often not addressed by the community. As this post-processing convolution is known to degrade the quality of non-Gaussian information (which cosmologists are becoming increasingly interested in) such plots demonstrate the trade-off between non-Gaussian information and reconstruction fidelity.

As can be seen in Fig. 3 and Table 1, sparse approaches significantly outperform the smoothed (and non-smoothed) KS approach in all cases, over all metrics tracked. Importantly sparse approaches are able to recover reasonable results even when the noise level entirely dilutes the true signal, as in the $n_{\rm gal} = 10$ setting, making such approaches on (at least) cluster data very attractive for future studies.

5.3.1 Hypothesis testing: Bolshoi clusters

Perhaps more interestingly, we now perform a series of hypothesis tests as discussed in Section 4.2. For each of the remaining three Bolshoi clusters, we construct three possible example hypothesis tests which one may wish to perform. In this case, these hypotheses were either structure removal followed by segmentation-inpainting; or Gaussian smoothing of certain structures (i.e. smoothing multiple peaks into a single larger peak which may be of interest when conducting peak-count analysis). Though these are both extremely useful considerations, it is important to stress the generality of our approach such that any well-defined operation on the reconstructed image, with a clear understandable hypothesis, is applicable.

To ensure the method behind hypothesis testing is clear, we will walk through a typical application. The top row of Fig. 4 displays the hypothesis tests applied to the first Bolshoi cluster. Conceptually, the correct way to interpret Hypothesis 1 (H1, red) is: *'The central dark core is likely just an artefact of the reconstruction'*.

This structure is then removed from the image by segmentation-inpainting (lower left image), and the objective function is then recalculated. It is found that the objective function is now larger than the approximate level-set threshold $\epsilon'_{99\ \text{per cent}}$, the surrogate segmentation-inpainted image falls outside of the 99 per cent HPD credible region, and so the hypothesis is rejected. This implies that the structure is not simply an artefact, but is necessary to the integrity of the reconstruction, i.e. this structure is now determined to be physical at 99 per cent confidence. However, had removing this region not raised the objective function above $\epsilon'_{99\ \text{per cent}}$, then the conclusion is that there is insufficient evidence to reject the hypothesis (which is not equivalent to saying that the region is strictly not physical).

An identical thought process can be applied to H2 and H3 of the top row in Fig. 4, H1 in the second row of Fig. 4, and all three hypothesis tests presented in the final row. In each case, a substructure of the $\kappa^{\rm map}$ is removed via segmentation-inpainting and it is queried whether the resulting surrogate solution $\kappa^{\rm sgt} \in C'_\alpha$. Each of the large substructures on the final row, and H2 of the second row, are determined to be physical at 99 per cent confidence. Conversely, the comparatively smaller substructures considered in H2 and H3 of the top row do not saturate the level-set threshold, and are therefore undetermined. All numerical data related to hypothesis testing of the Bolshoi cluster reconstructions can be found in Table 2.

H2 and H3 of the middle row of Fig. 4 have a different interpretation. In these cases, the central region has been blurred by segmentation-smoothing (convolution with a Gaussian smoothing kernel) – the difference between these two cases being simply the degree of smoothing. Here, the hypothesis is: *'The central region is likely to be just a single peak, rather than two'*. As in the previous example, the objective function is recalculated and is now greater than $\epsilon'_{99\ \text{per cent}}$ and so the hypothesis is rejected. The natural conclusion is thus that the data **is** sufficient to determine that at least two peaks are physically present at 99 per cent confidence.

6 APPLICATION TO ABELL 520 OBSERVATIONAL CATALOGUES

We perform an application of our entire reconstruction pipeline to real observational data sets. We select two observational data sets of the A520 cluster (Clowe et al. 2012; Jee et al. 2014) – hereafter, for clarity we refer to them as C12 and J14 (as in Peel et al. 2017).[3]

[3] http://www.cosmostat.org/software/glimpse

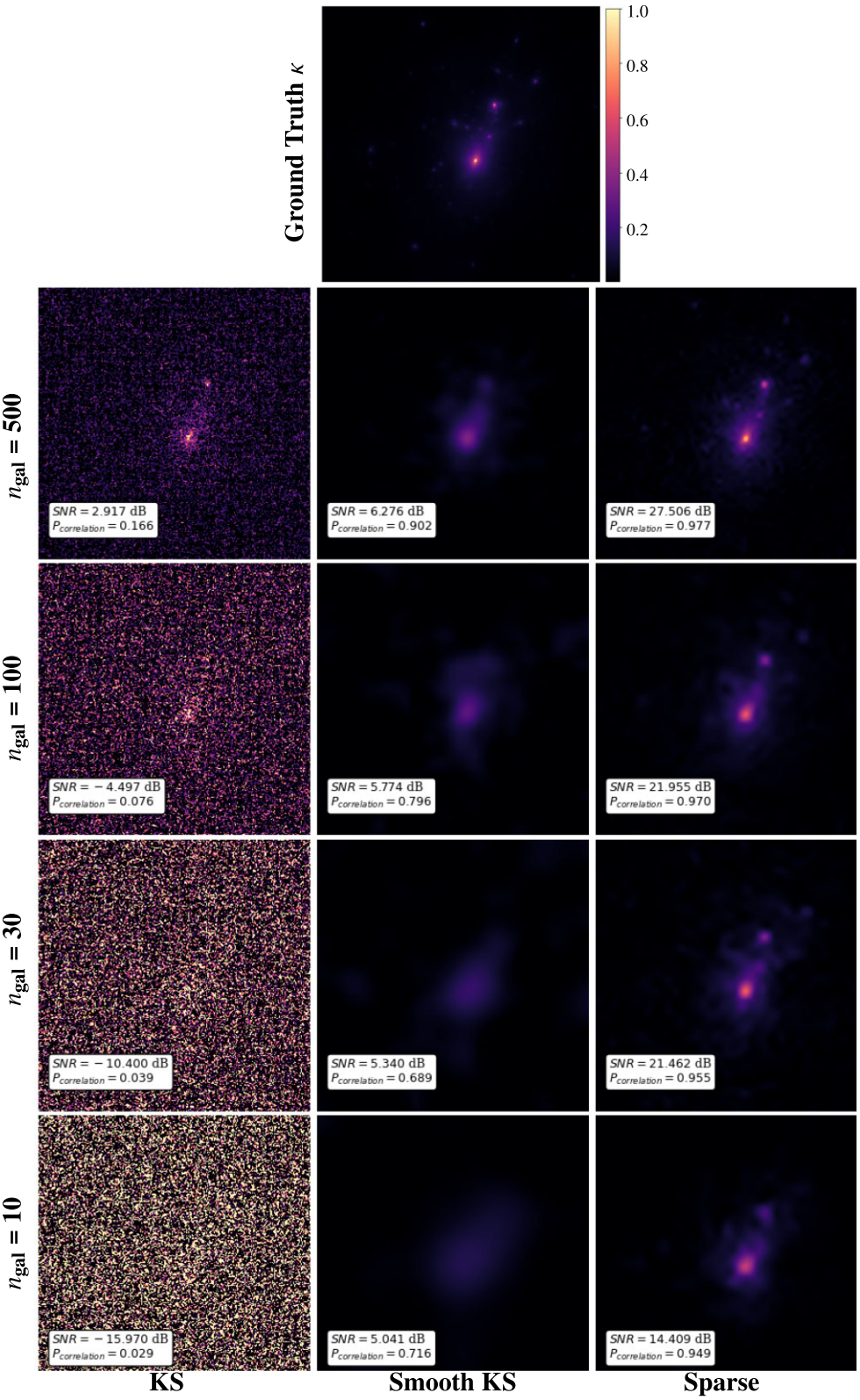

Figure 3. Top to bottom: Ground truth convergence map, simulations with noise levels corresponding to $n_{\rm gal} \in [500, 100, 30, 10]$, respectively. Notice the clear effectiveness of sparse reconstruction over the standard KS method for a range of input SNR values. The numerical details can be found in Table 1. The vertical labels indicate the input $n_{\rm gal}$ used to simulate realistic noise for a given row, whereas horizontal labels indicate the reconstruction type. An optimal (grid searched to maximize the recovered SNR) Gaussian smoothing kernel was applied to the KS recovery to yield the KS (smooth) recovery in an attempt to remove noise from the KS estimator (obviously this is not possible in practice, where the ground truth is unknown: results shown therefore present the best possible performance for the smoothed KS estimator). Clearly, in all cases, the super-resolution sparse approach produces convergence maps that are far more representative of the ground truth across the aforementioned metrics.

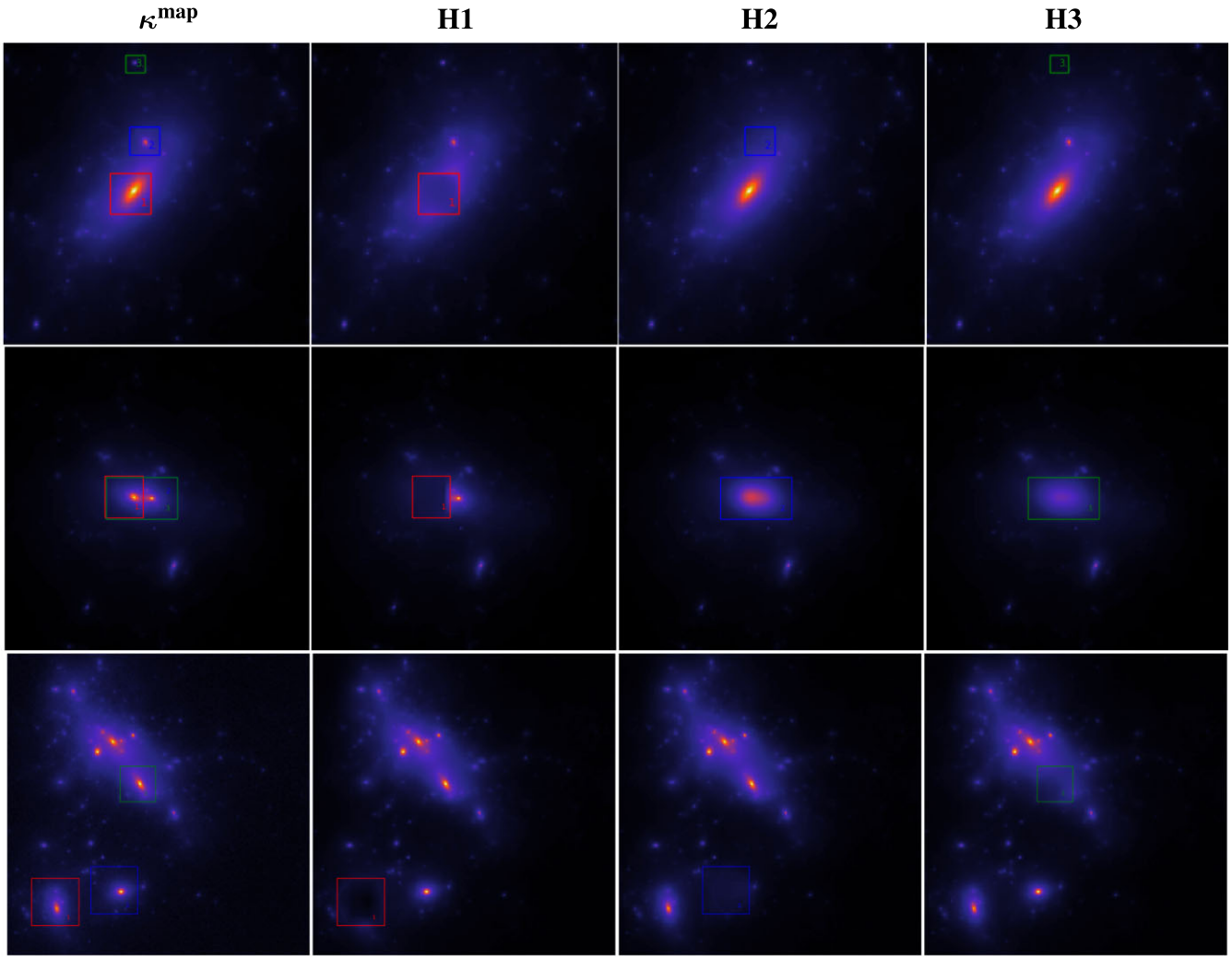

Figure 4. General: Hypothesis testing of three selected structures in the Bolshoi-1 cluster convergence field. The number density of galaxy observations n_{gal} was set idealistically to 500 arcmin^{-2} simply for demonstration purposes. Additionally super-resolution was not active and the masking was trivially set to the identity, again to simplify the example for demonstration purposes. All numerical details can be found in Table 2. Top row: We correctly determine that region 1 (*red*) is physical with 99 per cent confidence. Regions 2 (*blue*) and 3 (*green*) remain within the HPD region and are therefore inconclusive, given the data and noise level. Middle row: We correctly determine that all three null hypotheses (*red, blue* and *green*) are rejected at 99 per cent confidence. In H1, the conclusion is that the left hand peak was statistically significant. In H2 and H3, the conclusion is that an image with the two peaks merged it unacceptable, and therefore the peaks are distinct at 99 per cent confidence. Bottom row: We correctly determine that all three hypothesis regions (*red, blue*, and *green*) Ω_D are physical with 99 per cent confidence.

For a full description of the data sets, how they were constructed, and how they account for different systematics we recommend the reader look to the respective papers. These initial investigations claim to have detected several over dense regions within the merging A520 system, the most peculiar of which was a so-called dark core (location 2 in Fig. 5) for which multiwavelength observations could not determine an optical counterpart. Such a dark core would provide a contradiction to the currently understood model of collisionless dark matter – the idea being that during the collision of two massive clusters, dark matter was stripped from each cluster through self-interactions, forming an overdense residual between the two clusters, which would naturally not exhibit an optical counterpart.

The J14 catalogue contains approximately twice the number of galaxies than C12, though both are derived from the same ACS (four pointings) and Magellan images. In addition, J14 combines these images with the CFHT catalogue used in the authors' previous work (Jee et al. 2012). The C12 observing area extends over a larger angular surface than the J14, so for this analysis we limit both data sets to the region spanned by both sets. Due to the number density of measurements being very low, we are forced to project the measurements into a 32 × 32 grid – to ensure that the average number of galaxies in each grid pixel is at least above 1, though ideally we want many galaxies in each pixel to minimize the noise contribution from intrinsic ellipticity. In fact, even in this resolution the space is incomplete in several pixels, but we draw a compromise between the completeness of the space and the resolution of the data.

The data covariance was constructed directly from the number density of observations per pixel (directly inferred during catalogue gridding), with an assumed intrinsic ellipticity dispersion of 0.37. Combining this data covariance, the associated gridded data sets, and the associated mask, MAP reconstructions of the C12 and J14 convergence maps were recovered at a super-resolution magnification of 8. Reconstructions are presented in Fig. 5.

Table 2. Displays the MAP objective function, level-set threshold at 99 per cent confidence, surrogate objective function, and whether the removed region was successfully identified as being physical. This data set corresponds to Fig. 4.

Test	Initial $f(\kappa)+g(\kappa)$	Threshold $\epsilon'_{99\%}$	Surrogate $f(\kappa^{\rm sgt})+g(\kappa^{\rm sgt})$	Reject H_0 ?
		Bolshoi-1		
H1	95 426	163 408	805 513	✓
H2	95 426	163 408	134 080	×
H3	95 426	163 408	100 582	×
		Bolshoi-2		
H1	97 121	165 103	824 260	✓
H2	97 121	165 103	221 492	✓
H3	97 121	165 103	366 981	✓
		Bolshoi-3		
H1	83 419	151 401	369 939	✓
H2	83 419	151 401	234 305	✓
H3	83 419	151 401	314 089	✓

6.1 Hypothesis testing of local structure: A520 data sets

We conducted hypothesis tests on the three primary overdense regions, in addition to the contested dark core, in both the C12 and J14 data sets. In the absence of an optical counterpart, detection at high confidence of the dark core (location 2 in Fig. 5) would provide a contradiction to the collisionless model of dark matter – indicating potential self-interaction of dark matter. Due to the high estimated noise level present in the data, and the limited data resolution, only the two largest peaks in both data sets (peaks 1 and 3 of Fig. 5) sufficiently raised the objective function to reject the hypothesis at any meaningful confidence. This is to say that given the limited, noisy data and using the measurement operator and prior (ℓ_1-term) presented in this paper we can say that the data are insufficient to statistically determine the physicality of local small-scale substructure (such as the dark core) in both the C12 and J14 data sets. The initial conflict between C12 and J14 was over the existence and position of a dark core (location 2 in Fig. 5), with a notably large mass-to-light ratio, indicating the possibility of self-interacting dark matter. A subsequent inquiry was conducted (Peel et al. 2017) using the GLIMPSE reconstruction algorithm (Lanusse et al. 2016) and concluded that this peculiar peak existed in the J14 data set but not in the C12 data set.

As such, our conclusions agree well with Peel et al. (2017) (and generally with those drawn in both C12 and J14). However, within our Bayesian hierarchical formalism (which constitutes a principled statistical framework), we push this conclusion further to say that the data are insufficient to determine the physicality of these peaks.

6.2 Hypothesis testing of global structure: A520 data sets

Interestingly, we can perform a final novel hypothesis test of global structure. This hypothesis is as follows: 'The two MAP estimates are consistent with both sets of data', i.e. the MAP convergence estimate recovered from the J14 (C12) data is within the credible set (at 99 per cent confidence) of the C12 (J14) objective function. We find that the J14 (C12) MAP reconstruction is an acceptable solution to the C12 (J14) inverse problem and so the MAP solutions do not disagree – numerically this is shown in Table 3.

Given the inherent limitations of the data, we are forced to conclude: *'The data* are insufficient to determine the existence of individual substructures at high confidence – though the two largest overdense regions are found to be globally physical at 99 per cent confidence. The two MAP estimates are also found to be consistent at 99 per cent confidence'.

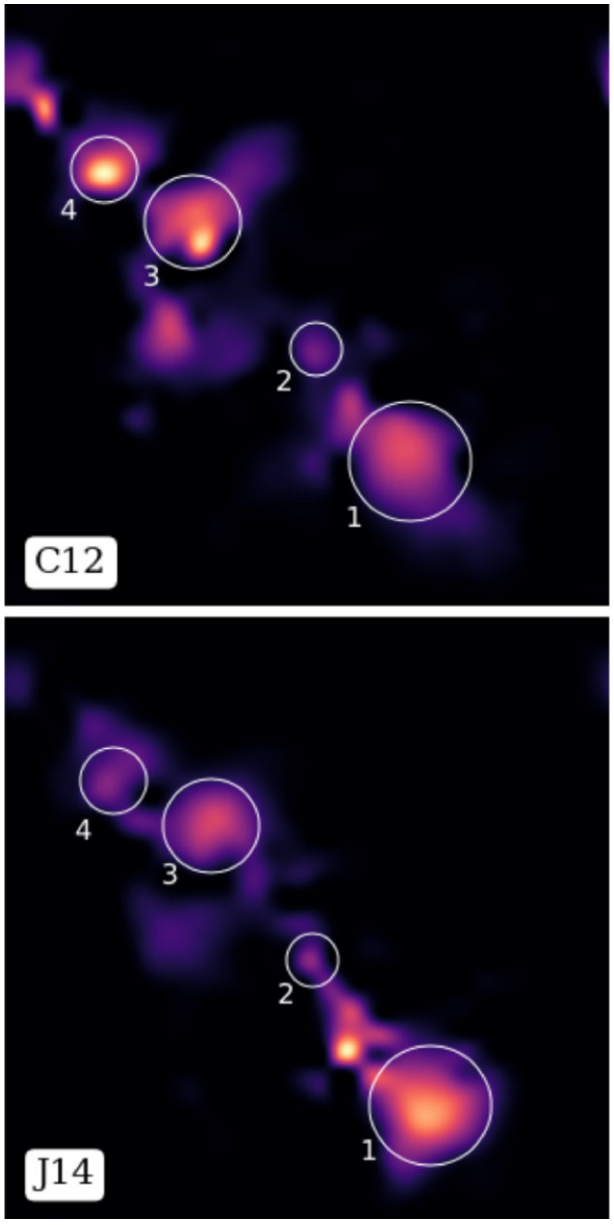

Figure 5. Top: Super-resolution sparse Bayesian reconstruction of J14 and C12, respectively. In a Bayesian manner, it is found that the two data sets do not globally disagree at 99 per cent confidence. However, given the data resolution and noise levels, only peaks 1 and 3 (in both data sets) could be determined to be statistically significant. This is not to say that they do not exist, but implies that the data quantity and quality are insufficient to make a robust, principled statistical statement that could be used as evidence of their existence. The contested peak 2 is not detected at any reasonable confidence in either data set.

7 CONCLUSIONS

We have presented a sparse hierarchical Bayesian mass-mapping algorithm that provides a principled statistical framework through which, for the first time, we can conduct uncertainty quantification on recovered convergence maps without relying on any assumptions

Table 3. Displays the MAP objective function, level-set threshold at 99 per cent confidence, surrogate objective function, and whether the null hypothesis H_0 is rejected. As can be seen, both MAP solutions fail to reject the null hypothesis in the other's objective function. This leads us to conclude that the two data sets do not disagree at 99 per cent confidence. Further discussion akin to the Kullback–Leibler divergence of the two posteriors is beyond the scope of this paper, but perhaps of interest in future work.

Hypothesis test	Initial $f(\kappa)+g(\kappa)$	Threshold $\epsilon'_{99\,\text{per cent}}$	Surrogate $f(\kappa^{\text{sgt}})+g(\kappa^{\text{sgt}})$	Reject H_0?
C12 ⇔ J14	99 231	168 044	125 601	×
J14 ⇔ C12	98 943	167 243	134 391	×

of Gaussianity. Moreover, the presented formalism draws on ideas from convex optimization (rather than MCMC techniques), which makes it notably fast and allows it to scale well to big data, i.e. high-resolution and wide-field convergence reconstructions (which will be essential for future stage IV surveys, such as LSST and Euclid).

Additionally, we demonstrate a hierarchical Bayesian inference approach to automatically approximate the regularization parameter, and show that it produces near-optimal results in a variety of cases. We, however, note that this approach does not work generally, and can be unstable in extreme settings.

We showcase our Bayesian inference approach (with emphasis on the application of the uncertainty quantification techniques) on both simulation data sets and observational data (the A520 merging cluster data set). Our mass-mapping formalism is shown to produce significantly more accurate convergence reconstruction than the Kaiser–Squires estimator on all simulations considered. Hypothesis tests of substructure are demonstrated.

It is found that neither of the two A520 data sets considered could provide sufficient evidence to determine the physicality of any contested substructure (i.e. the existence of so-called dark cores) at significant confidence. It is informative to note that our methods were, in fact, sufficiently sensitive to detect the largest peaks in both data sets at 99 per cent confidence. None the less, global hypothesis tests indicate a good agreement between the two sets of data. These conclusions are roughly in agreement with those drawn previously but go further to demonstrate just how uncertain these types of cluster-scale weak lensing reconstructions inherently are (typically as a limitation of the relative information content of low-resolution, noisy data sets).

It is now natural to extend this formalism to the entire celestial sphere – a necessity of large-scale reconstruction techniques that aim to fully utilize the forthcoming Euclid and LSST[4] survey data.

ACKNOWLEDGEMENTS

This paper has undergone internal review in the LSST Dark Energy Science Collaboration. The internal reviewers were Chihway Chang, Tim Eifler, and François Lanusse. The authors would like to thank Luke Pratley for an introduction to the C++ SOPT framework and the internal reviewers for valuable discussions. MAP was supported by the Science and Technology Facilities Council (STFC). TDK was supported by a Royal Society University Research Fellowship (URF). This work was also supported by the Engineering and Physical Sciences Research Council (EPSRC) through grant EP/M0110891 and by the Leverhulme Trust. The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3-Lyon/Villeurbanne – France) funded by the Centre National de la Recherche Scientifique; 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; STFC DiRAC HPC Facilities, funded by UK BIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515.

Author contribution: MAP: methodology, data curation, investigation, software, visualization, and writing – original draft; JDM: conceptualization, methodology, project administration, supervision, and writing – review and editing; XC: methodology, investigation, and writing – review and editing; TDK: methodology, supervision, and writing – review and editing; and CGRW: methodology.

DATA AVAILABILITY

All observational data and simulated data sets utilized throughout this paper are publicly available and can be found in the corresponding references. The DarkMapper reconstruction software will be publicly released in the very near future,[5] and can be made available upon reasonable request.

REFERENCES

Alsing J., Heavens A., Jaffe A. H., Kiessling A., Wandelt B., Hoffmann T., 2016, MNRAS, 455, 4452
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bunn E. F., Zaldarriaga M., Tegmark M., de Oliveira-Costa A., 2003b, Phys. Rev. D, 67, 023501
Cai X., Pereyra M., McEwen J. D., 2018a, Monthly Notices of the Royal Astronomical Society. Uncertainty quantification for radio interferometric imaging – I. Proximal MCMC methods, Vol. 480, Oxford University Press, p. 4154
Cai X., Pereyra M., McEwen J. D., 2018b, Uncertainty quantification for radio interferometric imaging: II. MAP estimation. Monthly Notices of the Royal Astronomical Society, Vol. 480, Oxford University Press, p. 4170
Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223
Carrillo R. E., McEwen J. D., Van De Ville D., Thiran J.-P., Wiaux Y., 2013, IEEE Signal Process. Lett., 20, 591
Chang C. et al., 2018, MNRAS, 475, 3165
Clowe D., Brada M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJ, 648, L109
Clowe D., Markevitch M., Bradač M., Gonzalez A. H., Chung S. M., Massey R., Zaritsky D., 2012, ApJ, 758, 128 (C12)
Coles P., Chiang L.-Y., 2000, Nature, 406, 376
Combettes P. L., Pesquet J.-C., 2011, Proximal splitting methods in signal processing. Fixed-point algorithms for inverse problems in science and engineering, Springer, p. 185
Corless V. L., King L. J., Clowe D., 2009, MNRAS, 393, 1235
Heavens A., 2009, Nucl. Phys. B, 194, 76
Horowitz B., Seljak U., Aslanyan G., 2019, Efficient optimal reconstruction of linear fields and band-powers from cosmological data. Journal of Cosmology and Astroparticle Physics, IOP Publishing, p. 035
Jee M. J., Mahdavi A., Hoekstra H., Babul A., Dalcanton J. J., Carroll P., Capak P., 2012, ApJ, 747, 96

[4] https://www.lsst.org

[5] git@github.com:astro-informatics/darkmapper.git

Jee M. J., Hoekstra H., Mahdavi A., Babul A., 2014, ApJ, 783, 78 (J14)
Jeffrey N. et al., 2018, MNRAS, 479, 2871
Jullo E., Pires S., Jauzac M., Kneib J.-P., 2014, MNRAS, 437, 3969
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kilbinger M., 2015, Rep. Progr. Phys., 78, 086901
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Lanusse F., Starck J.-L., Leonard A., Pires S., 2016, A&A, 591, A2
Leonard A., Lanusse F., Starck J.-L., 2014, MNRAS, 440, 1281
Mediavilla E., Muñoz J. A., Garzón F., Mahoney T. J., 2016, Astrophysical Applications of Gravitational Lensing. Cambridge University Press. p. 153
Munshi D., Coles P., 2017, J. Cosmol. Astropart. Phys., 2, 010
Munshi D., Valageas P., van Waerbeke L., Heavens A., 2008, Phys. Rep., 462, 67
Padmanabhan N., Seljak U., Pen U. L., 2003, New Astron., 8, 581
Paykari P., Lanusse F., Starck J.-L., Sureau F., Bobin J., 2014, A&A, 566
Peel A., Lanusse F., Starck J.-L., 2017, ApJ, 847, 23
Pereyra M., 2016, Proximal markov chain monte carlo algorithms. Statistics and Computing, Vol. 26, Springer, p. 745
Pereyra M., 2017, SIAM J. Imaging Sci., 10, 285
Pereyra M., Bioucas-Dias J., Figueiredo M., 2015, 2015 23rd European Signal Processing Conference (EUSIPCO). Maximum-a-posteriori estimation with unknown regularisation parameters, p. 230
Piras D., Joachimi B., Schäfer B. M., Bonamigo M., Hilbert S., van Uitert E., 2018, MNRAS, 474, 1165
Pires S., Starck J.-L., Amara A., 2009, MNRAS, 395, 1265
Robert C.-P., 2001, The Bayesian Choice. Springer
Schneider P., 2005, Weak Gravitational Lensing. preprint (arXiv:astro-ph/0509252)
Schneider M. D., Hogg D. W., Marshall P. J., Dawson W. A., Meyers J., Bard D. J., Lang D., 2015, ApJ, 807, 87
Starck J.-L., Murtagh F., Fadili J., 2015, Sparse image and signal processing: Wavelets and related geometric multiscale analysis. Cambridge university press
Taylor P. L., Kitching T. D., McEwen J. D., Tram T., 2018, Testing the cosmic shear spatially-flat universe approximation with generalized lensing and shear spectra. Phys. Rev. D, Vol. 98, American Physical Society
Troxel M. A., Ishak M., 2015, Phys. Rep., 558, 1
VanderPlas J. T., Connolly A. J., Jain B., Jarvis M., 2011, ApJ, 727, 118
Wallis C. G. R., McEwen J. D., Kitching T. D., Leistedt B., Plouviez A., 2017, Mapping dark matter on the celestial sphere with weak gravitational lensing. preprint (arXiv:1703.09233)

This paper has been typeset from a TEX/LATEX file prepared by the author.