Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **489,** 3236–3250 (2019)
Advance Access publication 2019 August 26

doi:10.1093/mnras/stz2373

Sparse Bayesian mass mapping with uncertainties: peak statistics and feature locations

M. A. Price,[⋆] J. D. McEwen, X. Cai and T. D. Kitching (for the LSST Dark Energy Science Collaboration)
Mullard Space Science Laboratory, University College London, London RH5 6NT, UK

Accepted 2019 August 23. Received 2019 August 23; in original form 2019 January 8

ABSTRACT

Weak lensing convergence maps – upon which higher order statistics can be calculated – can be recovered from observations of the shear field by solving the lensing inverse problem. For typical surveys this inverse problem is ill-posed (often seriously) leading to substantial uncertainty on the recovered convergence maps. In this paper we propose novel methods for quantifying the Bayesian uncertainty in the location of recovered features and the uncertainty in the cumulative peak statistic – the peak count as a function of signal-to-noise ratio (SNR). We adopt the sparse hierarchical Bayesian mass-mapping framework developed in previous work, which provides robust reconstructions and principled statistical interpretation of reconstructed convergence maps without the need to assume or impose Gaussianity. We demonstrate our uncertainty quantification techniques on both Bolshoi *N*-body (cluster scale) and Buzzard V-1.6 (large-scale structure) *N*-body simulations. For the first time, this methodology allows one to recover approximate Bayesian upper and lower limits on the cumulative peak statistic at well-defined confidence levels.

Key words: gravitational lensing: weak – methods: statistical – techniques: image processing – cosmological parameters – dark matter.

1 INTRODUCTION

In an empty universe the *null geodesics* along which photons travel correspond directly to straight lines. However, in the presence of a non-uniform distribution of matter the null geodesics are perturbed *via* gravitational interaction with the local matter over or underdensity i.e. the photons are *gravitationally lensed* (Schneider 2005; Munshi et al. 2008; Heavens 2009; Grimm & Yoo 2018). As this gravitational interaction is sensitive only to the total matter distribution, and the overwhelming majority of matter is typically dark, gravitational lensing provides a natural probe of dark matter itself (Clowe et al. 2006).

Collections of associated photons emitted from a distant object travel along separate geodesics which are perturbed in different ways by the intervening matter distribution, e.g. photons travelling closer to matter over densities will interact more strongly and therefore be perturbed more than those farther away. As such the geometry of a distant object is warped (Bartelmann & Schneider 2001) – i.e. colloquially the object is *lensed*.

Provided the propagating photons at no time come closer than one Einstein radius to the intervening matter over and underdensities, the object is *weakly lensed*. Weak gravitational lensing of distant galaxies manifests itself at first order into two quantities; the spin-0 convergence κ which is a magnification, and the spin-2 shear γ which is a perturbation to the galaxy ellipticity (third flattening).

Both the shear γ and the convergence κ have dominant intrinsic (i.e. in the absence of lensing effects) underlying values which makes measurements of the lensing effect difficult. In fact, there is no a priori way to know the intrinsic brightness of a galaxy (resulting in an inherent degeneracy – the mass-sheet degeneracy) and so the convergence is not an observable quantity. In fact, the standard convergence is not gauge invariant and is therefore fundamentally unobservable (Grimm & Yoo 2018). However, as the intrinsic ellipticity distribution of galaxies has zero mean one can average to recover the shearing contribution, hence the shear is an observable quantity. As such, measurements of the shear field are taken and inverted to form estimators of the convergence. Typically this inverse problem is seriously ill-posed and so substantial uncertainty on the reconstructed convergence map is introduced.

A wealth of information may be calculated directly from the shear field (often in the form of second-order statistics; Kilbinger 2015 – such as the power spectrum as in Alsing et al. 2016; Taylor et al. 2018) though recently there is increasing interest in extracting non-Gaussian information from the convergence field, e.g. peak

---

[⋆] E-mail: m.price.17@ucl.ac.uk

© 2019 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

statistics, Minkowski functionals, extreme value statistics (Coles & Chiang 2000; Munshi & Coles 2017; Peel et al. 2017a, 2018; Fluri et al. 2018).

Primarily, the interest has arisen as higher order statistics of the convergence field have been shown to provide complementary constraints on dark matter cosmological parameters which are typically poorly constrained by second-order statistics (Pires et al. 2010).

However, to make principled statistical inferences from the convergence field, the inversion from γ to κ must be treated in a principled statistical manner – something which until recently was missing from convergence reconstruction algorithms which were either not framed in a statistical framework (e.g. Kaiser & Squires 1993; VanderPlas et al. 2011; Lanusse et al. 2016; Wallis et al. 2017; Jeffrey et al. 2018) or made assumptions of Gaussianity (e.g. Corless, King & Clowe 2009; Schneider et al. 2015; Alsing et al. 2016). As the information of interest in higher order convergence statistics is non-Gaussian, assumptions of Gaussianity in the reconstruction process severely degrade the quality of the cosmological information.

Recently, a mass-mapping framework was developed (see Price et al. 2018a) which addressed precisely this issue. This new sparse hierarchical Bayesian mass-mapping formalism can be rapidly computed, can be extended to big data, and provides a principled statistical framework for quantifying uncertainties on reconstructed convergence maps. Notably, it has been shown to accurately reconstruct very high dimensional Bayesian estimators many orders of magnitude faster than *state-of-the-art* proximal MCMC algorithms – it was specifically benchmarked against Px-MALA (Pereyra 2015; Durmus, Moulines & Pereyra 2018) in Price et al. (2018b).

In this paper, we propose two novel uncertainty quantification techniques, aimed to answer two frequently asked questions of the recovered convergence map. The first of these questions asks where a feature of interest in the reconstructed convergence map could have been observed – typically this has been addressed by bootstrapping; however we can now infer it directly in a Bayesian manner. The second question asks given a magnitude threshold what is the maximum and minimum number of peaks which could have been observed, within some well-defined confidence.

The structure of this article is as follows. To begin, in Section 2 we provide cursory introduction to weak lensing from a mathematical perspective, with emphasis on the weak lensing planar forward model in subSection 2.2. Following this we provide a brief overview of Bayesian inference and the previously developed (Price et al. 2018a) sparse hierarchical Bayesian mass-mapping algorithm in Section 3. An introduction to Bayesian credible regions, specifically the highest posterior density credible region is provided in Section 3.1. In Section 4 we develop a novel Bayesian inference approach to quantifying the uncertainty in reconstructed feature location, which we then showcase on illustrative *N*-body cluster simulation data in Section 5. We then introduce a novel Bayesian inference approach for recovery of principled uncertainties on the aggregate peak count statistic in Section 6. Following this we showcase this Bayesian inference approach to quantify uncertainty in the aggregate peak statistic in Section 7 on *N*-body large-scale structure (LSS) illustrative simulation data. Finally we draw conclusions in Section 8.

2 WEAK GRAVITATIONAL LENSING

In this section we provide a brief introduction to weak gravitational lensing, with emphasis on how this effect manifests itself into observable quantities. For a detailed background review of the field see Bartelmann & Schneider (2001) and Schneider (2005). For a more mathematical background of the field, with emphasis on statistical methods see Munshi et al. (2008), Heavens (2009), and Grimm & Yoo (2018). For a background of specifically the peak statistics see Lin (2016).

2.1 Mathematical background

In a non-uniform distribution of matter the null geodesics along which photons travel are no longer straight lines, instead they are now sensitive to the local matter distribution. When many photons are propagating from a distant object to us here and now, the local matter distribution adjusts the geometry of the object we observe – the object is *gravitationally lensed.*

Provided the trajectory of the propagating photons at no time comes closer than one Einstein radius θ_{E} to the intervening matter overdensities then the lens equation,

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

remains effectively singular and we are in the *weak lensing* regime. Equivalently one can define the weak lensing regime to be convergence fields for which $\kappa \ll 1$ – ensuring that the shear signal remains linear. Here the Einstein radius is given by

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

where G is the gravitational constant, M_{lens} is lensing mass, c is the speed of light *in vacuo*, and $f_K(r)$ is the angular diameter distance defined as

$$f_K(r) = \begin{cases} \sin(r) & \text{if} \quad K = 1, \\ r & \text{if} \quad K = 0, \\ \sinh(r) & \text{if} \quad K = -1, \end{cases} \tag{3}$$

where r is the comoving distance and K is the curvature of the universe, which has been observed to be consistent with 0 by Planck Collaboration VI (2018).

As galaxies are naturally sparsely distributed across the sky, most observations fall within the weak lensing regime. The weak gravitational lensing effect can be described by a lensing potential $\phi(r, \omega)$ which is the integrated Newtonian potential $\Phi(r, \omega)$ along the line of sight

$$\phi(r, \omega) = \frac{2}{c^2}\int_0^r \mathrm{d}r' \frac{f_K(r-r')}{f_K(r)f_K(r')}\Phi(r', \omega), \tag{4}$$

where $\omega = (\theta, \psi)$ are angular spherical coordinates. A further constraint exists, such that the local Newtonian potential $\Phi(r, \omega)$ must satisfy the Poisson equation

$$\nabla^2\Phi(r, \omega) = \frac{3\Omega_{\mathrm{M}}H_0^2}{2a(r)}\delta(r, \omega), \tag{5}$$

for matter density parameter Ω_{M}, Hubble constant H_0, and scale parameter $a(r)$. Combined, equations (4) and (5) allow one to make inferences of cosmological parameters from observations of the lensing potential $\phi(r, \omega)$.

At linear order, gravitational lensing manifests itself as two quantities: the spin-0 convergence field κ (magnification) and the spin-2 shear field γ (perturbation to ellipticity). It can be shown that (Bartelmann & Schneider 2001; Schneider 2005) these quantities can be related to the lensing potential $\phi(r, \omega)$ by

$$\kappa(r, \omega) = \frac{1}{4}(\eth\bar{\eth} + \bar{\eth}\eth)\, \phi(r, \omega), \tag{6}$$

$$\gamma(r, \omega) = \frac{1}{2}\eth\eth\, \phi(r, \omega), \tag{7}$$

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

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

$$\bar{\eth} \equiv -\sin^{-s}\theta\left(\frac{\partial}{\partial\theta} - \frac{i\partial}{\sin\theta\partial\psi}\right)\sin^{s}\theta. \tag{9}$$

2.2 Lensing planar forward model

Often second-order statistics (Kilbinger 2015) related to the shear γ are computed (e.g. the shear power spectrum as in Alsing et al. 2016 and Taylor et al. 2018), though increasingly there is interest in extracting weak lensing information from the convergence directly – typically higher order non-Gaussian information.

Unfortunately, due to an inherent degeneracy the convergence is not an observable quantity (Bartelmann & Schneider 2001; Schneider 2005; Grimm & Yoo 2018) – this effect is colloquially referred to as the *mass-sheet degeneracy*. However, as the intrinsic ellipticity distribution of galaxies has zero mean, averaging many ellipticity observations within a given pixel provides a good estimate of the shear field.

In fact, there exists a further degeneracy between κ and γ such that the true observable is the *reduced shear* g but for the context of the paper we will assume $\gamma \approx g \ll 1$ – see Mediavilla et al. (2016, p.153), Price et al. (2018a), or Wallis et al. (2017) for details on how to account for the non-linear reduced shear.

As both κ and γ are related to ϕ a relation between κ and γ can be formed. Therefore, typically measurements of the shear field are taken and inverted to form estimates of the underlying convergence field. For small fields of view the *flat sky approximation* can be made, which reduces the spin-raising $\eth$ and lowering $\bar{\eth}$ operators to (Bunn et al. 2003)

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

From equations (6) and (7) it is clear that the forward model in Fourier space between κ and γ is given by

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

with the mapping operator being

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

where we have dropped the spin subscripts for clarity. To recover an estimator of the convergence one must invert this forward model.

The most naive inversion technique is that of Kaiser–Squires (KS) inversion (Kaiser & Squires 1993), which is direct inversion in Fourier space, i.e.

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

where we have again dropped function arguments and subscripts for clarity. This approach attempts to mitigate the effect of noise by convolving the recovered convergence estimate with a broad Gaussian smoothing kernel, which severely degrades the quality of non-Gaussian information. This poses a somewhat serious issue as non-Gaussian information is precisely the information that is to be extracted from the convergence field. Therefore for higher order convergence statistics the KS estimator is patently suboptimal.

Moreover, decomposition of spin fields on bounded manifolds is well known to be degenerate (Bunn et al. 2003) and so inversion of shear to convergence for masked fields is inherently ill-defined – in particular the KS estimator is known to break down for non-trivial masking. In fact the lensing inverse problem is often seriously ill-posed, therefore motivating methods regularized by prior information.

3 SPARSE BAYESIAN MASS MAPPING

Many mass-mapping algorithms have been considered (e.g. Kaiser & Squires 1993; VanderPlas et al. 2011; Jee et al. 2016; Lanusse et al. 2016; Wallis et al. 2017; Chang et al. 2018; Jeffrey et al. 2018), however in the context of this paper we wish to conduct principled statistical analysis of the reconstructed convergence map, and so we opt for the sparse hierarchical Bayesian algorithm presented in Price et al. (2018a) and benchmarked against MCMC algorithms in Price et al. (2018b).

Recently a sparse hierarchical Bayesian framework for convergence reconstruction was presented (Price et al. 2018a) which is not limited to Gaussian priors – in fact the prior need not even be differentiable. In this work we adopt this mass-mapping algorithm, which we briefly describe below.

First, by Bayes' theorem the *posterior distribution* of the convergence κ reads

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

which shows how one should infer the *posterior* $p(\kappa|\gamma)$ from the *likelihood function* (data fidelity term) $p(\gamma|\kappa)$ and the *prior* (regularization term) $p(\kappa)$ (see e.g. Trotta 2017, for a clear introduction to Bayesian inference in a cosmological setting). In the scope of this paper we do not consider the *Bayesian evidence* $\int_{\mathbb{C}^N} p(\gamma|\kappa)p(\kappa)\mathrm{d}\kappa$ as it acts as a normalization term and so does not affect the recovered solution. Typically the Bayesian evidence is used for model comparison which is an avenue of study in of itself.

In a Bayesian inference problem one often wishes to find the solution κ which maximizes the posterior given data and some model. From the monotonicity of the logarithm function, maximization of the posterior is equivalent to minimization of the *log-posterior* such that

$$\underset{\kappa}{\operatorname{argmax}}\left\{p(\kappa|\gamma)\right\} \equiv \underset{\kappa}{\operatorname{argmin}}\left\{-\log(p(\kappa|\gamma))\right\}, \tag{15}$$

where the convergence κ which maximizes the posterior is given by κ^{map}, where MAP stands for *maximum* a posteriori. Provided the posterior is log-concave the minimization of the log-posterior takes the form of a *convex optimization* problem, which can be rapidly computed even in high-dimensional settings.

Let $\kappa \in \mathbb{C}^N$ be the discretized complex convergence field and $\gamma \in \mathbb{C}^M$ be the discretized complex shear field, where M is the number of binned shear measurements and N is the dimensionality of the recovered convergence field. Suppose we can define a *measurement operator* $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N} : \kappa \mapsto \gamma$ which maps κ on to γ. On the plane, the measurement operator is given by (e.g. Price et al. 2018a)

$$\boldsymbol{\Phi} = \mathbf{F}^{-1}\mathbf{D}\mathbf{F}, \tag{16}$$

where $\mathbf{D}$ is the planar forward model in Fourier space as defined in equation (12), and $\mathbf{F}$ ($\mathbf{F}^{-1}$) is the forward (inverse) fast Fourier transform.

Now suppose that some noise n is contaminating our measurements, then measurements are obtained through the *measurement equation*

$$\gamma = \mathbf{\Phi}\kappa + n. \tag{17}$$

Within this article we will consider the typical case in which the noise $n \sim \mathcal{N}(0, \sigma_n^2) \in \mathbb{C}^M$, i.e. i.i.d. (independent and identically distributed) zero mean Gaussian noise of variance σ_n^2. In this setting the likelihood function is given by a multivariate Gaussian, which for diagonal covariance $\Sigma = \sigma_n^2 \mathbb{I}$ is given compactly as

$$p(\gamma|\kappa) \propto \exp\left(\frac{-\|\mathbf{\Phi}\kappa - \gamma\|_2^2}{2\sigma_n^2}\right), \tag{18}$$

which (as in Price et al. 2018a) is regularized by a sparsity promoting Laplace-type l_1-norm wavelet prior

$$p(\kappa) \propto \exp\Big(- \mu\|\mathbf{\Psi}^\dagger\kappa\|_1\Big), \tag{19}$$

where $\mu \in \mathbb{R}_+$ is the jointly inferred MAP regularization parameter (Pereyra, Bioucas-Dias & Figueiredo 2015) – the derivation and implementation of which may be found in Price et al. (2018a). Sparsity promoting priors in wavelet dictionaries have explicitly been shown to be effective in the weak lensing setting (Lanusse et al. 2016; Peel, Lanusse & Starck 2017b; Jeffrey et al. 2018; Price et al. 2018a).

Using these terms, the minimization of the log-posterior becomes

$$\kappa^{\rm map} = \underset{\kappa}{\operatorname{argmin}} \left\{ \mu\|\mathbf{\Psi}^\dagger\kappa\|_1 + \frac{\|\mathbf{\Phi}\kappa - \gamma\|_2^2}{2\sigma_n^2} \right\}, \tag{20}$$

which is then solved iteratively by implementing a proximal forward–backward splitting algorithm (e.g. Combettes & Pesquet 2011).

Note that one can choose any convex log-priors 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).

3.1 Bayesian credible regions

In Bayesian analysis a posterior credible region $C_\alpha \in \mathbb{C}^N$ at confidence $100(1-\alpha)$ per cent is a set which satisfies

$$p(\kappa \in C_\alpha|\gamma) = \int_{\kappa \in \mathbb{C}^N} p(\kappa|\gamma)\mathbb{I}_{C_\alpha} \mathrm{d}\kappa = 1 - \alpha, \tag{21}$$

where $\mathbb{I}_{C_\alpha}$ is an indicator function defined such that

$$\mathbb{I}_{C_\alpha} = \begin{cases} 1 & \text{if} \quad \kappa \in C_\alpha, \\ 0 & \text{if} \quad \kappa \notin C_\alpha. \end{cases} \tag{22}$$

There are in general a large number of posterior regions (hypervolumes) which satisfy this integral. The decision-theoretically optimal region – in the sense of minimum enclosed volume – is called the *highest posterior density* (HPD) credible region (Robert 2001) and is defined to be

$$C_\alpha := \{\kappa : f(\kappa) + g(\kappa) \leq \epsilon_\alpha\}, \tag{23}$$

where $f(\kappa) = \mu\|\Psi^\dagger\kappa\|_1$ is the log-prior term and $g(\kappa) = \|\Phi\kappa - \gamma\|_2^2/2\sigma_n^2$ is our data fidelity term (log-likelihood function). Here ϵ_α defines a level set (i.e. isocontour) of the log-posterior set such that equation (21) is satisfied. In practice the true HPD credible region is difficult to compute due to the high-dimensional integral in equation (21), motivating computationally efficient approximate techniques.

Recently a conservative approximation of the HPD credible set was proposed by Pereyra (2017) which exploits developments in probability concentration theory. The approximate HPD credible set C'_α is given by

$$C'_\alpha := \{\kappa : f(\kappa) + g(\kappa) \leq \epsilon'_\alpha\}, \tag{24}$$

with approximate level-set threshold

$$\epsilon'_\alpha = f(\kappa^{\rm map}) + g(\kappa^{\rm map}) + \tau_\alpha\sqrt{N} + N, \tag{25}$$

where the bounding term $\tau_\alpha = \sqrt{16\log(3/\alpha)}$ which in turn is constrained to confidence $\alpha \in (4\exp{(-N/3)}, 1)$. The error of this approximation is bounded above by

$$0 \leq \epsilon'_\alpha - \epsilon_\alpha \leq \eta_\alpha\sqrt{N} + N, \tag{26}$$

where $\eta_\alpha = \sqrt{16\log(3/\alpha)} + \sqrt{1/\alpha}$. This upper bound is typically conservative, meaning the isocontour is at all times larger than the true isocontour (i.e. this estimator will never produce an underestimate). In Price et al. (2018b) the bound on recovered local error bars was found to be ±10 to 15 per cent larger than the true MCMC – yet could be computed $\mathcal{O}(10^6)$ times faster. A similar comparison was done by Cai, Pereyra & McEwen (2018a) in a radio interferometric setting.

The concept of approximate HPD credible regions is particularly useful as it allows us to explore high-dimensional posteriors – many orders of magnitude larger than *state-of-the-art* MCMC techniques are currently able to accommodate – in a computationally efficient manner.

4 BAYESIAN PEAK LOCATIONS

Often one wishes to know the location of a feature of interest within the reconstructed convergence $\kappa^{\rm map}$. Typically, this uncertainty is assessed *via* bootstrapping of the recovered image for a large number of simulated noise fields (as in e.g. Peel et al. 2017b).

With the concept of approximate HPD credible regions in mind, we propose a novel Bayesian approach to quantifying uncertainty in the peak location which we will refer to as the *'Bayesian location'*.

In essence the Bayesian location is computed as follows: A feature of interest is removed from the recovered convergence map, this feature is then inserted back into the convergence map at a new position to create a surrogate convergence map, if this surrogate map is within the approximate credible set then the position at which the feature was inserted cannot be rejected, if the surrogate is not in the approximate credible set then the position can be rejected. This process is computed for a sample of the total possible insertion positions, eventually providing an isocontour of 'acceptable' positions. This isocontour, at a well-defined confidence level, is the Bayesian location.

4.1 Bayesian location

Suppose we recover a (MAP) convergence field $\kappa^{\rm map}$ *via* optimization of the objective function defined in equation (20) which contains a feature of interest (e.g. a large peak). Let us define the subset of pixels which contain this feature to be $\Omega_{\mathcal{Z}} \subset \Omega$, where Ω is the entire image domain.

To begin with, extract the feature $\mathcal{Z} = \kappa^{\rm map}\mathbb{I}_{\Omega_\mathcal{Z}}$, i.e. a convergence field which contains only the feature of interest. Now we adopt the process of *segmentation inpainting* (Cai et al. 2018a; Cai, Pereyra & McEwen 2018b; Price et al. 2018a) to create a

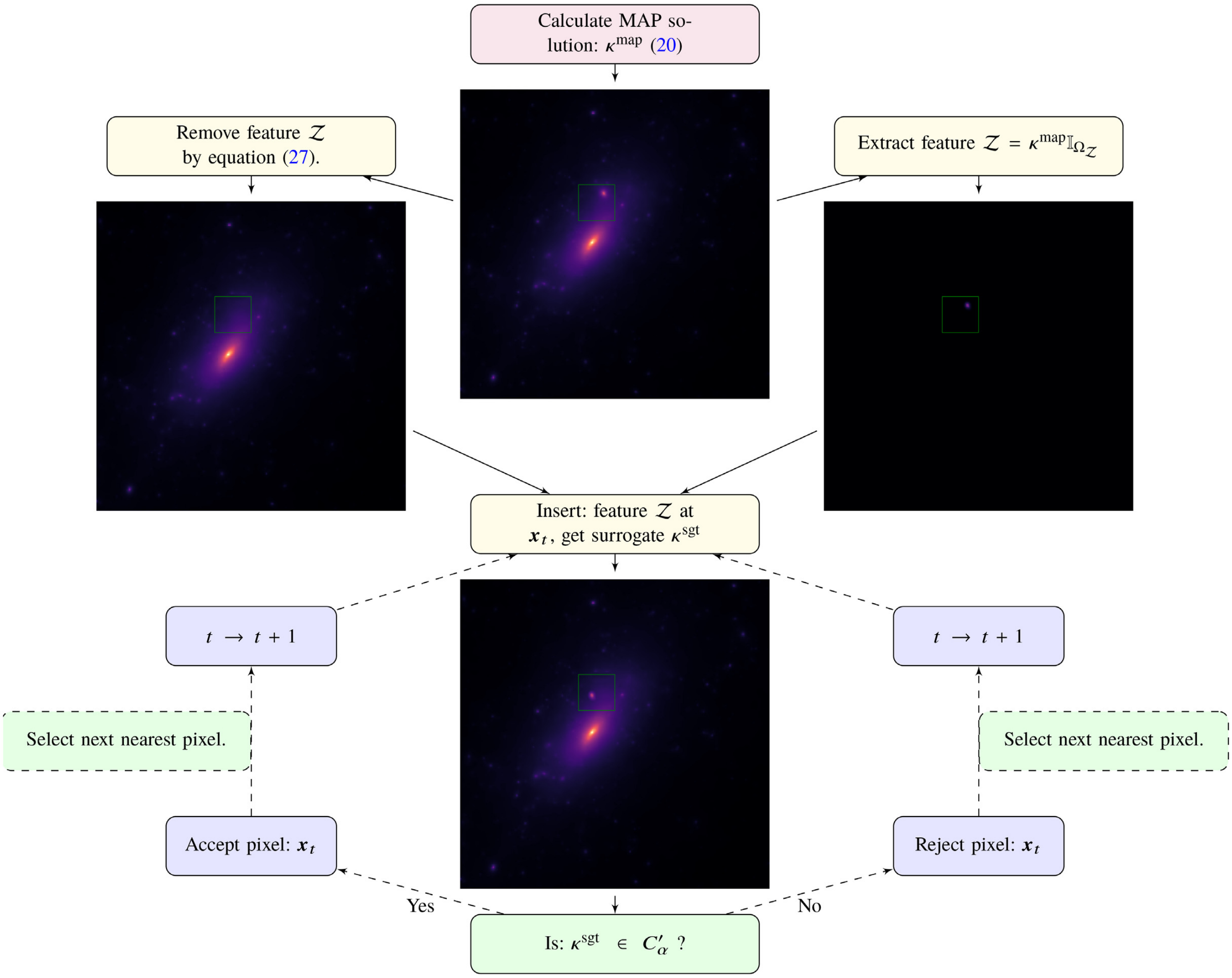

Calculate MAP solution: κ^{map} (20)
Remove feature $\mathcal{Z}$ by equation (27).
Extract feature $\mathcal{Z} = \kappa^{\text{map}}\mathbb{I}_{\Omega_\mathcal{Z}}$
Insert: feature $\mathcal{Z}$ at $\boldsymbol{x}_t$, get surrogate κ^{sgt}
$t \rightarrow t+1$
$t \rightarrow t+1$
Select next nearest pixel.
Select next nearest pixel.
Accept pixel: $\boldsymbol{x}_t$
Reject pixel: $\boldsymbol{x}_t$
Yes
No
Is: $\kappa^{\text{sgt}} \in C'_\alpha$?

Figure 1. Schematic representation of the inverse nested iterations to determine the *Bayesian location* (see Section 4). The Bayesian location is a positional uncertainty on a feature of interest $\mathcal{Z}$ within a recovered convergence field. Once a complete ring of pixels have been rejected the algorithm returns a binary map of accepted pixels which we call the Bayesian location. Any pixel outside of this location is rejected at $100(1-\alpha)$ per cent confidence. Alternately the probability isocontour bounding the set of acceptable pixels can be located by N-splitting circular bisection as described in Section 4.2 and Appendix A.

convergence field realization without the feature of interest $\mathcal{Z}$ but with background signal replaced.

Mathematically segmentation inpainting is represented by the iterations

$$\kappa^{(t+1),\text{sgt}} = \kappa^{\text{map}}\mathbb{I}_{\Omega\setminus\Omega_\mathcal{Z}} + \Lambda\text{soft}_\lambda(\Lambda^\dagger\kappa^{(t),\text{sgt}})\mathbb{I}_{\Omega_\mathcal{Z}}, \tag{27}$$

where Λ is an appropriately selected dictionary – for this purpose we simply use the Daubechies 8 (DB8) wavelet dictionary with 8-levels and λ is the soft-thresholding parameter.

Following the wavelet inpainting, in order to separate the true feature from the background residual convergence the signal which was inpainted into the region $\Omega_\mathcal{Z}$ is subtracted from the extracted feature $\mathcal{Z}$ – effectively accounting for the residual background signal which would likely have been present even in the absence of the feature $\mathcal{Z}$. At this junction the surrogate convergence κ^{sgt} is hypothesis tested for physicality (Cai et al. 2018a; Price et al. 2018a).

If a feature is not found to be physical, the algorithm terminates at this point as, fundamentally, it is illogical to evaluate the uncertainty in position of an object of which you cannot statistically determine the existence.

Now that we have successfully isolated $\mathcal{Z}$ we can insert it back into the surrogate field κ^{sgt} at a perturbed position. It is then sufficient to check if

$$f(\kappa^{\text{sgt}\prime}) + g(\kappa^{\text{sgt}\prime}) \leq \epsilon'_\alpha, \tag{28}$$

where $\kappa^{\text{sgt}'}$ represents the surrogate with the feature $\mathcal{Z}$ inserted at a perturbed location.

If the inequality does hold, then the conclusion is that at $100(1-\alpha)$ per cent confidence we cannot say that the feature could not be found at this location. If the equality does not hold then $\mathcal{Z}$ in its observed form could not have been found at the new location at $100(1-\alpha)$ per cent confidence. The question then becomes, which perturbed positions are acceptable and which are not.

With the above in mind, we propose a typical inverse nested iterative scheme to determine the pixel-space isocontour for a given feature in the reconstructed convergence field. Schematically this iterative process is outlined in Fig. 1. Essentially, inverse nesting is: start in a ring 1-pixel from the MAP peak location in the first iteration, moving the ring one pixel outwards after each iteration.

4.2 N-splitting circular bisection

Inverse nested iterations are sufficiently fast for low-dimensional reconstructions (256 × 256), however as the dimensionality of the reconstructed domain grows it becomes increasingly beneficial to adopt more advanced algorithms to compute the Bayesian location in an efficient manner.

We propose a novel iterative algorithm for computing the pixel-space position isocontour which we call *N-splitting circular bisection* (N-splitting), the full details of which can be found in Appendix A. A brief outline of N-splitting is given below.

Suppose we wish to compute positions on the Bayesian location isocontour at *equiangular intervals* $\Delta\Theta$ (defined clockwise about the peak location) relative to the y-axis. Then we require $n = 2\pi/\Delta\Theta$ sampling angles which are trivially given by

$$\Theta_i = i\,\Delta\Theta, \tag{29}$$

where i is an iterative factor which sets the angle for a given direction Θ_i.

Once Θ_i is defined for a single direction, the distance d'_α along direction Θ_i such that the objective function saturates the level-set threshold ϵ'_α is found by bisection. Mathematically, this is formally defined to be

$$d^i_\alpha = \min_d \Big\{ \ d \in \Gamma_i \mid f(\kappa_d^{\mathrm{sgt}}) + g(\kappa_d^{\mathrm{sgt}}) > \epsilon'_\alpha \ \Big\}, \tag{30}$$

$$\Gamma_i = \Big\{ q_1 \sin(\Theta_i), q_2 \cos(\Theta_i) \mid q_1, q_2 \in \mathbb{R}_+ \Big\}, \tag{31}$$

$$d = \sqrt{q_1^2 + q_2^2}, \tag{32}$$

where Γ_i is the subset of the real domain which lie on the line of infinite extent along a given direction Θ_i sourced at the peak location, and κ_d^{sgt} is the surrogate convergence map constructed by inserting the feature of interest $\mathcal{Z}$ into a perturbed location $[q_1 \sin(\Theta_i), q_2 \cos(\Theta_i)]$.

Once a representative set of positions on the location isocontour are computed, the *convex hull* is taken – the convex hull is simply the smallest convex set which contains all samples of the location isocontour. The boundary of this closed convex set of acceptable pixels is taken as the Bayesian location.

5 ILLUSTRATIVE EXAMPLE OF THE BAYESIAN LOCATION

In this section we perform sparse Bayesian reconstructions of a large cluster extracted from the Bolshoi N-body simulation (Klypin, Trujillo-Gomez & Primack 2011; Lanusse et al. 2016), upon which we construct and assess the performance of Bayesian locations for each of the four largest subhaloes. In line with previous work of Price et al. (2018a) and in the related article of Lanusse et al. (2016) we refer to this extracted cluster as Bolshoi-3.

We grid the Bolshoi convergence field on to a discretized complex map of dimension (1024 × 1024) so as to demonstrate that the sparse Bayesian approach can construct Bayesian estimators efficiently even when the dimension of the problem is of $\mathcal{O}(10^6)$ or larger – dimensions at which MCMC techniques can become highly computationally challenging.

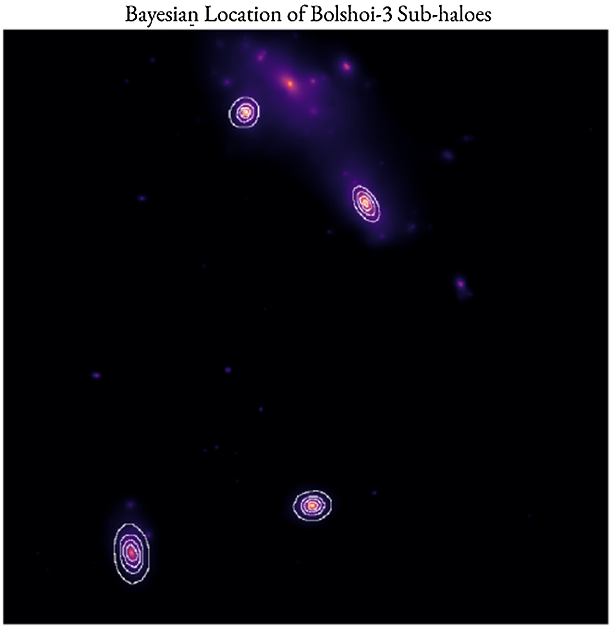


Figure 2. Combined plot of the 99 per cent confidence Bayesian locations at SNR = 12, 15, 17, 20 dB. The outer rings represent the noisier position isocontours whereas as the data become cleaner the isocontour ring becomes smaller (therefore the rings represent isocontours at SNR = 12, 15, 17, 20 dB, from the outer rings inwards, respectively). N-splitting circular bisection (see Section 4.2) was used to efficiently compute each isocontour. For input SNR's below ≈10 the smaller local features cannot be determined physically *via* the initial hypothesis test, and so we truncate our analysis at SNR = 12.

5.1 Methodology

First, we construct a complex discretized set of artificial shear measurements $\tilde{\gamma} \in \mathbb{C}^M$ by

$$\tilde{\gamma} = \Phi\kappa, \tag{33}$$

where κ is the input Bolshoi-3 convergence map. We then contaminate these mock measurements with noise n, which for simplicity we select to be i.i.d. Gaussian noise $n \sim \mathcal{N}(0, \sigma_n^2)$ of zero mean and variance σ_n^2. The variance is selected such that the SNR of the noisy artificial shear maps can be varied, and is therefore set to be

$$\sigma_n = \sqrt{\frac{\| \boldsymbol{\Phi}\kappa \|_2^2}{N}} \times 10^{-\frac{\mathrm{SNR}}{20}}. \tag{34}$$

The MAP convergence field κ^{map} is recovered *via* the sparse Bayesian mass-mapping algorithm using DB10 wavelets (10-levels), and the Bayesian location for the set of 4 peaks is constructed. For a detailed discussion of how noise levels in decibels (dB) translate to physical quantities such as galaxy number density see Price et al. (2018a).

5.2 Analysis and computational efficiency

To demonstrate this uncertainty quantification technique we construct 99 per cent confidence Bayesian locations for the four largest subhaloes in the Bolshoi-3 cluster, for input SNR in dB of $\in \{12, 15, 17, 20\}$.

In Figs 2 and 3 it is apparent that, as expected, the positional uncertainty isocontour at 99 per cent confidence is smaller for less noisy data, growing in proportion to the noise. In our analysis 32 N-splitting directions (pointings) were used, though as can be seen

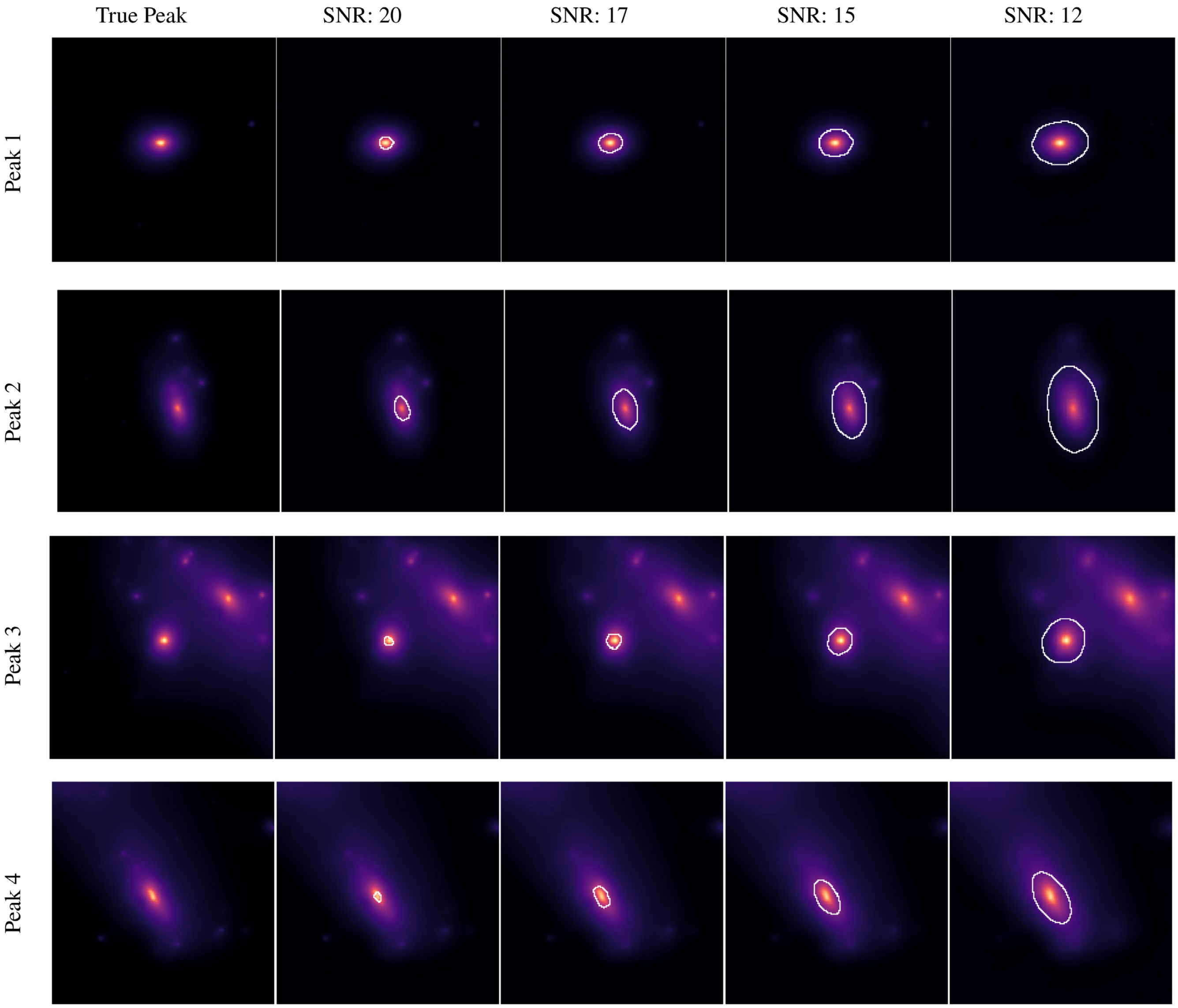


Figure 3. *Left to right:* Sparse Bayesian reconstructions of Bolshoi-3 peaks 1 to 4 (*top to bottom*, respectively) followed by Bayesian locations (see Section 4) at 99 per cent confidence for input SNR of 20.0 to 12.0 dB – which are overlaid on the sparse Bayesian MAP recovered convergence maps κ^{map} at the corresponding SNR level. As the input artificial shear becomes more contaminated with noise, the relative information content decreases, and so inferred uncertainty of the reconstructed peak positions increases, as one would logically expect. Note the asymmetry in the 99 per cent isocontour, which motivates the N-splitting searching algorithm (see Section 4.2 and Appendix A) rather than a naive circular inference (e.g. finding the maximal x and y displacements and assuming a circular isocontour). Finally, observe that the 99 per cent isocontour for peaks 3 and 4 are proportionally more tightly constrained than peaks 1 and 2. This is due to the local information density typically being higher in more signal dense regions – perturbations to pixels in more information dense regions are more tightly constrained and can therefore move less distance before saturating the approximate level-set threshold ϵ'_α. This effect has been observed in the context of *local credible intervals* as presented in Cai et al. (2018a) and introduced to the weak lensing setting in Price et al. (2018b).

in Figs 2 and 3 as few as 16 directions would easily have been sufficient given the smoothness of the isocontour.

Computationally, reconstruction of the MAP convergence field and computation of the Bayesian location for the complete set of peaks took ∼5 min on a standard 2016 MacBook Air. Notably, this is a conservative (Pereyra 2017) and tight (Price et al. 2018b) approximate Bayesian inference in an over 10^6-dimensional space on a personal laptop in minutes. Further to this, the sparse Bayesian algorithmic structure can be easily parallelizable and so this computational efficiency can be optimized further.

6 AGGREGATE UNCERTAINTY IN PEAK COUNTS

Building on the notion of an approximate HPD credible region presented in Section 3.1 we now ask the question: given a reconstructed convergence field κ^{map}, and at a given SNR threshold K, what is the maximum and minimum peak count at $100(1-\alpha)$ per cent confidence.

In this article we choose to define a *peak* in κ^{map} by a pixel $\kappa^{\text{map}}(\boldsymbol{x})$ which is larger than the 8 pixels which surround it (Lin 2016). A point of the peak statistic is computed as follows: A threshold K

is taken on κ^{map}, and the *peak count* (number of peaks which have intensity larger than K) is taken on the subset of pixels larger than the threshold.

Formally we define the *excursion set* $\Omega^+ \subset \Omega$ as

$$\Omega^+ = \Big\{ \boldsymbol{x} \mid \kappa^{\mathrm{map}}(\boldsymbol{x}) \ > \ K \Big\}, \tag{35}$$

where Ω is the complete set of recovered pixels. Define a further subset $\Pi \subset \Omega^+$ as the set of peaks in Ω^+:

$$\Pi(\kappa^{\mathrm{map}}) = \Big\{ \boldsymbol{x} \mid \kappa^{\mathrm{map}}(\boldsymbol{x}) \ > \ \kappa^{\mathrm{map}}(\boldsymbol{x}'), \ \forall \, \boldsymbol{x}' \in \mathcal{N}(\boldsymbol{x}) \Big\}, \tag{36}$$

where $\mathcal{N}(\boldsymbol{x})$ represents the set of immediately surrounding pixels.

Note that this definition is not valid for pixels on the boundary of the field, and so these pixels are necessarily not considered. This not only excludes the outer edge of κ^{map} but also any pixels surrounding masked regions (of which there are typically many).

Conceptually, we would then like to know at a given threshold K what is the maximum and minimum number of peaks which could exist such that the *surrogate solution* κ^{sgt} still belongs to the approximate HPD credible set C'_α.

Let $\eta^{\mathrm{max}}_\alpha$ be the *upper bound* on the number of peaks, and $\eta^{\mathrm{min}}_\alpha$ be the *lower bound* on the number of peaks, for a given threshold K, at $100(1-\alpha)$ per cent confidence. Further let η be the number of peaks calculated from the MAP solution κ^{map} at threshold K. Formally these quantities are given by

$$\eta \ \equiv \ |\Pi(\kappa^{\mathrm{map}})|, \tag{37}$$

$$\eta^{\mathrm{max}}_\alpha \ \equiv \max_{\kappa^{\mathrm{sgt}}} \Big\{ \, |\Pi(\kappa^{\mathrm{sgt}})| \in \mathbb{R}_+ \mid f(\kappa^{\mathrm{sgt}}) + g(\kappa^{\mathrm{sgt}}) \leq \epsilon'_\alpha \Big\}, \tag{38}$$

$$\eta^{\mathrm{min}}_\alpha \ \equiv \min_{\kappa^{\mathrm{sgt}}} \Big\{ \, |\Pi(\kappa^{\mathrm{sgt}})| \in \mathbb{R}_+ \mid f(\kappa^{\mathrm{sgt}}) + g(\kappa^{\mathrm{sgt}}) \leq \epsilon'_\alpha \Big\}, \tag{39}$$

where $|\Pi(\kappa)|$ is the *cardinality of the peak set* of a convergence field κ.

It is not all obvious how to locate the extremum of optimization problems given in equations (38) and (39) as they are inherently non-linear, non-convex problems. We can, however, propose a logical iterative approach to calculate well motivated approximations to the upper and lower peak count limits $\eta^{\mathrm{max}}_\alpha$ and $\eta^{\mathrm{min}}_\alpha$.

6.1 Approximate Bayesian lower bound on peak counts

It is perhaps conceptually more straightforward to minimize the cardinality of the peak set and so we will first describe this process.

To calculate an approximate bound on $\eta^{\mathrm{min}}_\alpha$ we begin by creating the initial peak set Π from the recovered convergence κ^{map}. The peak in $\Pi(\kappa^{\mathrm{map}})$ with lowest magnitude is located. The shortest distance r_{min} from the pixel location $\boldsymbol{x}$ to a pixel $\boldsymbol{x'}$ such that $\kappa^{\mathrm{map}}(\boldsymbol{x}') = y$ (where y is some magnitude at which it is assumed the peaks influence is sufficiently small) is computed in Euclidean space as $r_{\mathrm{min}} = |\boldsymbol{x} - \boldsymbol{x}'|$ – within this paper we simply set $y = 0$.

Let us define the region of interest $\Omega_{\mathcal{A}} \subset \Omega$ to be a circular aperture centred on the peak pixel location $\boldsymbol{x}$ with radius r_{min}. Conceptually, this acts as a proxy for the area effected by a large overdensity sourced at the location of the peak.

Now, define a *down-scaling kernel* $\mathcal{S}_{K,\Omega_{\mathcal{A}}} \in \mathbb{C}^{N \times N}$ which has the action of scaling the magnitude of the subset $\kappa^{\mathrm{map}}\mathbb{I}_{\Omega_{\mathcal{A}}}$ of pixels belonging to the region of interest $\Omega_{\mathcal{A}}$ on to $[0, K]$. Application of the down-scaling operator returns a surrogate solution κ^{sgt}. Mathematically this is

$$\kappa^{\mathrm{sgt}} = \mathcal{S}_{K,\Omega_{\mathcal{A}}}\big(\kappa^{\mathrm{map}}\big) = \kappa^{\mathrm{map}}\mathbb{I}_{\Omega \setminus \Omega_{\mathcal{A}}} + \frac{K}{\max\big(\kappa^{\mathrm{map}}\mathbb{I}_{\Omega_{\mathcal{A}}}\big)} (\kappa^{\mathrm{map}}\mathbb{I}_{\Omega_{\mathcal{A}}}). \tag{40}$$

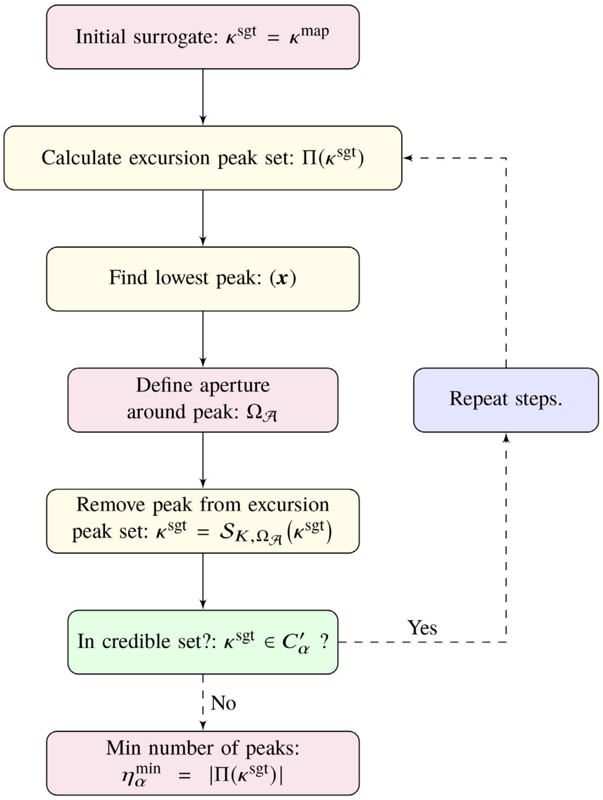

Initial surrogate: $\kappa^{\mathrm{sgt}} = \kappa^{\mathrm{map}}$
Calculate excursion peak set: $\Pi(\kappa^{\mathrm{sgt}})$
Find lowest peak: $(\boldsymbol{x})$
Define aperture around peak: $\Omega_{\mathcal{A}}$
Repeat steps.
Remove peak from excursion peak set: $\kappa^{\mathrm{sgt}} = \mathcal{S}_{K,\Omega_{\mathcal{A}}}(\kappa^{\mathrm{sgt}})$
In credible set?: $\kappa^{\mathrm{sgt}} \in C'_\alpha$?
Yes
No
Min number of peaks: $\eta^{\mathrm{min}}_\alpha = |\Pi(\kappa^{\mathrm{sgt}})|$

Figure 4. Schematic representation of the iteration steps in finding the Bayesian lower bound $\eta^{\mathrm{min}}_\alpha$ at confidence $100(1-\alpha)$ per cent of the peak count $|\Pi|$ for a given MAP reconstruction κ^{map}.

Application of $\mathcal{S}_{K,\Omega_{\mathcal{A}}}$ to an isolated region $\Omega_{\mathcal{A}}$ conserves the local topology of the field – which is precisely what we want as it means we are making no assumptions about the halo profile around a peak. Removing a peak by application of $\mathcal{S}_{K,\Omega_{\mathcal{A}}}$ creates a surrogate solution κ^{sgt} which is likely to minimize the increase in the objective function.

As such $\mathcal{S}_{K,\Omega_{\mathcal{A}}}$ is a good strategy for excluding peaks from $\Pi(\kappa^{\mathrm{map}})$ as it will likely maximize the number of peaks which can be removed from $\Pi(\kappa^{\mathrm{map}})$ before the level-set threshold ϵ'_α is saturated. Thus, it will likely be near decision-theoretically optimal at minimizing equation (39), which is precisely what we want.

A schematic of the iterative process proposed to find the Bayesian lower bound on the peak statistic can be seen in Fig. 4. In words, the process is as follows. Within each iteration, the lowest intensity peak within the peak set is removed forming a new surrogate convergence field κ^{sgt}, the objective function is recalculated and if the objective function is below the approximate level-set threshold ϵ'_α then the lowest peak within κ^{sgt} is now removed, so on and so forth until the objective function rises above ϵ'_α, at which the iterations are terminated and the minimum number of peaks is recovered.

6.2 Approximate Bayesian upper bound on peak counts

Now we invert our perspective in order to approximate the maximum number of peaks which could be observed at a given threshold K at $100(1-\alpha)$ per cent confidence. Here we will be considering the non-linear maximization problem constructed in equation (38).

First, we introduce the notion of the *inclusion set* Ω^-, conjugate to Ω^+ such that $\Omega^- \cup \Omega^+ \equiv \Omega$ and $\Omega^- \cap \Omega^+ = \varnothing$,

$$\Omega^- = \left\{ \boldsymbol{x} \mid \kappa^{\text{map}}(\boldsymbol{x}) \le K \right\}, \tag{41}$$

with this in mind, we can now cast the maximization problem into a minimization problem analogous to that used before.

We now wish to minimize the number of peaks that belong to the *inclusion set* Ω^- which is by definition equivalent to maximizing the number of peaks which belong to the *excursion set* Ω^+ – which is precisely what we want.

Analogously to Section 6.1 to construct our approximate bound we calculate the further subset $\Pi^- \subset \Omega^-$ which is defined similarly to the relation in equation (36) such that

$$\Pi^-(\kappa^{\text{map}}) = \left\{ \boldsymbol{x} \mid \kappa^{\text{map}}(\boldsymbol{x}) \ > \ \kappa^{\text{map}}(\boldsymbol{x}'), \ \forall \, \boldsymbol{x}' \in \mathcal{N}(\boldsymbol{x}) \right\}, \tag{42}$$

i.e. the subset of peaks below a threshold K.

In contrast to Section 6.1 we now locate the largest peak in Π^-. Suppose that this peak is found at $\Pi^-(\boldsymbol{x})$, we now construct a circular aperture about $\boldsymbol{x}$ with radius r_{min} as defined before. Let this circular aperture set of pixels be $\Omega_\mathcal{A} \subset \Omega$.

Now we define an *up-scaling kernel* $\mathcal{S}^\dagger_{K,\Omega_\mathcal{A}} \in \mathbb{C}^{N \times N}$ which has action

$$\mathcal{S}^\dagger_{K,\Omega_\mathcal{A}}\left(\kappa^{\text{map}}\right) = \kappa^{\text{map}} \mathbb{I}_{\Omega \setminus \Omega_\mathcal{A}} + \frac{K + \Delta}{\max\left(\kappa^{\text{map}} \mathbb{I}_{\Omega_\mathcal{A}}\right)} (\kappa^{\text{map}} \mathbb{I}_{\Omega_\mathcal{A}}), \tag{43}$$

which is very slightly different to the down-scaling operator in the numerator of the second term. Here Δ is an infinitesimal quantity added such that the rescaled peak within $\Omega_\mathcal{A}$ falls infinitesimally above the threshold K and is therefore counted as a peak. In practice we set Δ to be $\sim 10^{-5}$ and find that adjusting this quantity by $\mathcal{O}(10^2)$ has negligible effect on the recovered uncertainties.

With these conceptual adjustments we then follow the same iterations discussed in Section 6.1 to find the approximate Bayesian upper bound on the peak count η^{max}_α. Schematically this is given in Fig. 5.

Finally we return the tuple $\left(\eta^{\text{min}}_\alpha, \eta, \eta^{\text{max}}_\alpha\right)$ which is in the form (minimum, most likely, maximum) at $100(1 - \alpha)$ per cent confidence.

6.3 Limitations of rescaling

Suppose the SNR threshold K is large enough such that during iterations in schematic of Fig. 4 the cardinality of the excursion peak set $|\Pi(\kappa^{\text{sgt}})| \rightarrow 0$. In this situation even though the approximate level-set threshold ϵ'_α is not saturated, the algorithm is forced to stop as there are simply no more peaks to exclude (push down). At this point the strategy for removing peaks becomes locally ill-defined. Effectively this is a clipping artefact. To avoid this effect entirely, if $|\Pi(\kappa^{\text{sgt}})| = 0$ at any point within the iterations at a given threshold, the lower bound η^{min}_α at threshold K is set to 0, i.e. we are infinitely uncertain by construction.

Analogously, consider the case when K is small enough that during the iterations in schematic of Fig. 5 the cardinality of the inclusion peak set $|\Pi^-(\kappa^{\text{sgt}})| \rightarrow 0$. In this situation there are simply no more peaks to include (pull up). Again we remove this clipping effect by setting η^{max}_α at threshold K is set to $|\Pi(\kappa^{\text{sgt}})|$.

Typically these clipping effects only occur for very small $K \le 2$ or very large $K \ge 8$ thresholds, and so a wealth of information can be extracted from the intervening scales. Low thresholds clip the upper limit η^{max}_α as the cardinality of the peak set drops to 0 quickly, but the objective function rises comparatively slowly, as this SNR range is statistically dominated by noise. High threshold clip the lower limit η^{min}_α simply due to the inherently low count of peaks at high-SNR thresholds.

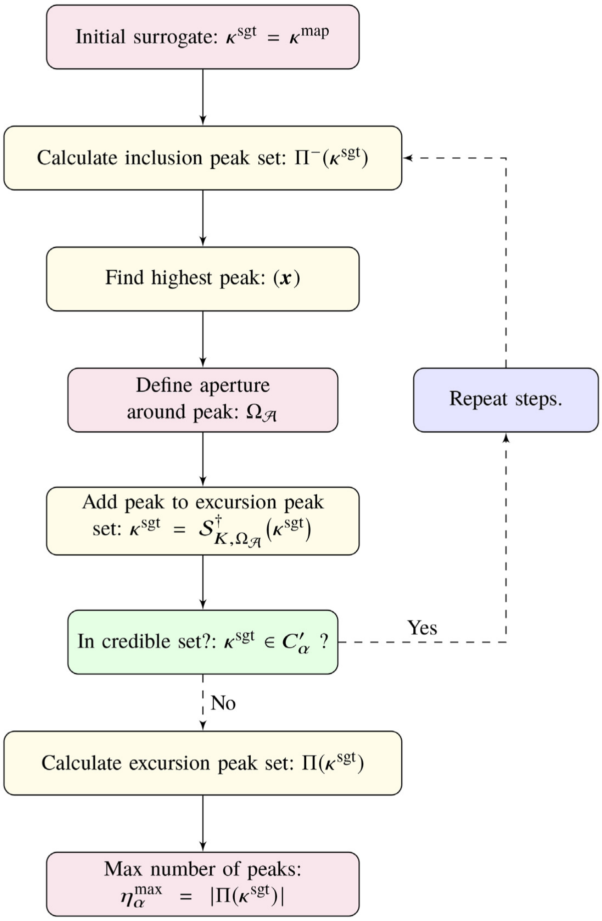

Initial surrogate: $\kappa^{\text{sgt}} = \kappa^{\text{map}}$
Calculate inclusion peak set: $\Pi^-(\kappa^{\text{sgt}})$
Find highest peak: $(\boldsymbol{x})$
Define aperture around peak: $\Omega_\mathcal{A}$
Repeat steps.
Add peak to excursion peak set: $\kappa^{\text{sgt}} = \mathcal{S}^\dagger_{K,\Omega_\mathcal{A}}(\kappa^{\text{sgt}})$
In credible set?: $\kappa^{\text{sgt}} \in C'_\alpha$?
Yes
No
Calculate excursion peak set: $\Pi(\kappa^{\text{sgt}})$
Max number of peaks: $\eta^{\text{max}}_\alpha = |\Pi(\kappa^{\text{sgt}})|$

Figure 5. Schematic representation of the iteration steps in finding the Bayesian upper bound η^{max}_α at confidence $100(1 - \alpha)$ per cent of the peak count $|\Pi|$ for a given MAP reconstruction κ^{map}.

Further to this, the decision-theory approach adopted here for locating the maximal and minimal values of the cumulative peak statistic is based on several assumption: removing lower peaks increases the objective function by less than larger peaks; the extent of a peak (dark matter overdensity) is approximated by a circular aperture; and removal of a peak has little to no effect on locations in the image domain which are outside of this aperture. All three of these assumptions are very reasonable.

Although further computational optimizations are not an immediate concern since our approach is already highly computationally efficient, we acknowledge that this iterative approach for removing peaks can easily be formulated as a bisection style problem which is likely to drastically reduce the computation time further – particularly for low thresholds, as it mitigates the number of trivial noise peak removal recalculations which are done in the brute force approach presented above. In future, should computational efficiency become of primary interest this speed up will be considered.

7 ILLUSTRATIVE EXAMPLE OF PEAK UNCERTAINTIES

In this section we apply the sparse Bayesian mass-mapping pipeline to high-resolution (2048 × 2048) convergence maps extracted from

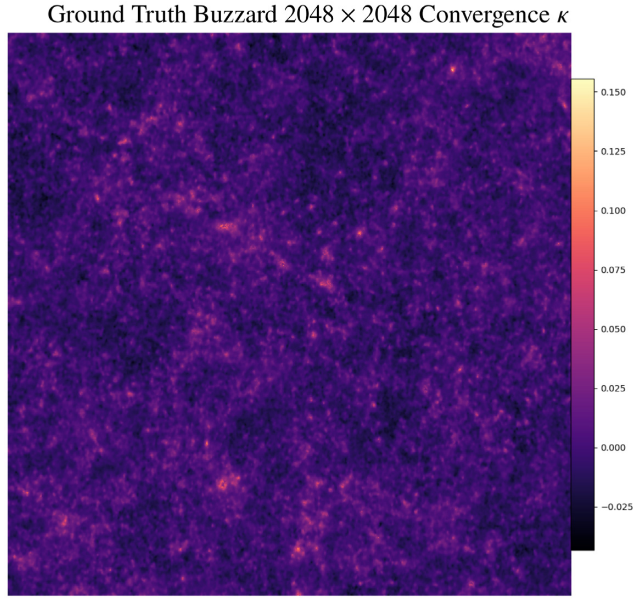


Figure 6. Input 2048 × 2048 convergence map extracted from the Buzzard *N*-body simulation.

the Buzzard V-1.6 *N*-body[1] simulation (see Fig. 6), upon which we construct the cumulative peak statistic (number of peaks above a threshold as a function of the threshold). Additionally, we recover the 99 per cent approximate Bayesian constraints on the peak count at each threshold, from which we infer the 68 per cent constraint so as to aid the reader in comparison to typical 1σ error bars quoted in related literature.

7.1 Simulated data sets

The Buzzard V-1.6 *N*-body simulation convergence catalogue (DeRose, Wechsler & Rykoff, in preparation; Wechsler, in preparation) is extracted by full ray tracing with the origin located at the box corner – leading to a quarter sky coverage. For wide fields the *flat sky approximation* breaks down (Wallis et al. 2017) and so this quarter sky coverage was reduced to smaller planar patches.

The complete quarter sky convergence catalogue was projected into a coarse HEALPIX[2] (Górski et al. 2005) pixelization ($N_{\rm side} = 4$). Inside of each pixel, we further tessellated the largest square region which we then project into a 2048 × 2048 grid. These gridded convergence maps formed our ground truth, discretized convergence fields.

As HEALPIX samples in such a way as to provide equal area pixels, and the Buzzard simulation galaxy density is fairly uniform, each extracted square region contained $\sim 2 \times 10^7$ galaxies leading to ~ 5 galaxies per pixel.

Due to a comparatively low density of samples, Poisson noise is prevalent, and thus extracted planar regions were passed through a multiscale Poisson denoising algorithm. This consisted of a forward Anscombe transform (in order to Gaussianize the Poisson noise), several TV-norm (total-variation) denoising optimizations of differing scale, followed by an inverse Anscombe transform (as in Lanusse et al. 2016; Price et al. 2018b). A more involved treatment could be applied, but this approach is sufficient to demonstrate our peak reconstructions.

7.2 Application to Buzzard V-1.6

We select at random one of many planar patches produced for the following application. Following the methodology presented in Section 5.1 we generate an artificial shear catalogue which we then contaminate with independent and identically distributed (i.i.d.) Gaussian noise such that the SNR of mock shear measurements is 30 dB – i.e. an idealized noise level simply for illustrative purposes.

The MAP convergence estimator $\kappa^{\rm map}$ is recovered from these noisy mock shear measurements *via* our sparse Bayesian mass-mapping framework. From $\kappa^{\rm map}$ we then calculate $\sigma^2 = \langle (\kappa^{\rm map})^2 \rangle$ which we then use as a measure of the noise level in the reconstructed convergence field. Implementing the uncertainty quantification technique presented in Section 6 we then construct the cumulative peak statistic for SNR thresholds $K \in [2\sigma,\ 8\sigma)$ at increments of 0.25σ with upper and lower 99 per cent approximate Bayesian confidence limits.

Fig. 7 displays the recovered cumulative peak statistic in both a linear and logarithmic scale. Typically, similar figures in the literature will quote 1σ error bars, and so for comparisons sake we convert the Bayesian 99 per cent confidence limits into the 68 per cent confidence limits which are comparable to 1σ constraints (in Fig. 7 we provide both confidence limits for clarity).

Complete reconstruction of the peak statistics for 24 threshold bins, each with approximate Bayesian upper and lower bounds, for a 2048 × 2048 resolution convergence map, with DB11 (11-level) wavelets, took $\sim$2 h on a 2016 MacBook Air. This is a non-trivial Bayesian inference in over 4×10^6 dimensions, and so 2 h is a very reasonable computation time – though further speed-ups are possible, e.g. we can trivially parallelize the calculations for each threshold leading to an increase in computational efficiency by a factor of the number of thresholds taken (in our case 24).

Additionally, the computational bottleneck is for lower thresholds as many low-intensity peaks must be removed, and thus an adaptive scheme could be implemented as discussed previously to avoid unnecessary sampling of these thresholds. With the aforementioned speed-ups, computation of the complete peak statistic is likely to take $\mathcal{O}$(minutes) on a personal laptop.

Following this initial analysis we reduce the SNR to investigate the effect of increased noise on shear measurements to the cumulative peak statistics within our Bayesian framework. We first decrease the SNR to 25 dB, seen in Fig. 8. Following which, we then reduce the input SNR further to 20 dB, the corresponding results being plotting in Fig. 9. This higher noise level of 20 dB is still a very optimistic (somewhat unrealistic) estimate of what upcoming surveys may reach; however in this paper we are primarily focused on demonstrating the methodology and leave detailed realistic simulations and forecasting for future work. A detailed description of how these noise levels in dB translate into observation constraints (e.g. galaxy number density etc.) can be found in (Price et al. 2018a).

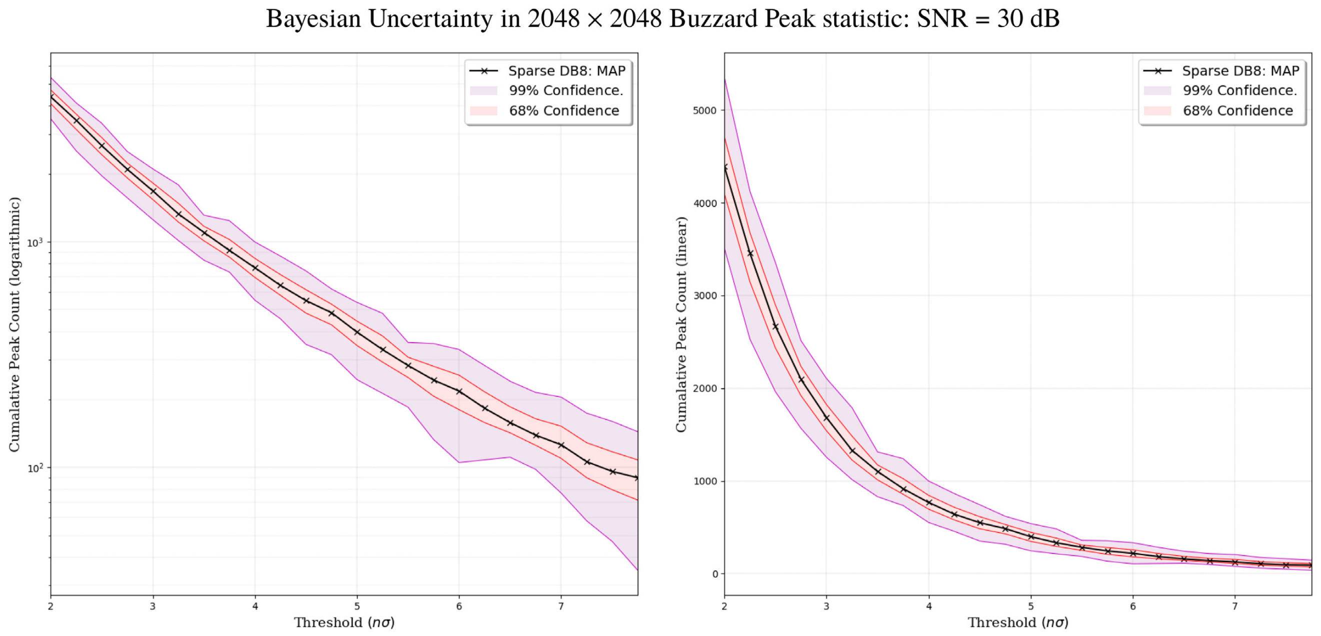
7.3 Analysis of peak statistic

Figs 7, 8, and 9 clearly show that as the noise level in the discretized complex shear field increases the isocontours of the cumulative peak statistic at 99 per cent and 68 per cent loosen noticeably. Therefore this, unsurprisingly, indicates that cleaner measurements are likely to give tighter constraints on cosmological parameters – though it should be noted that increasing the number of data points (i.e.

---

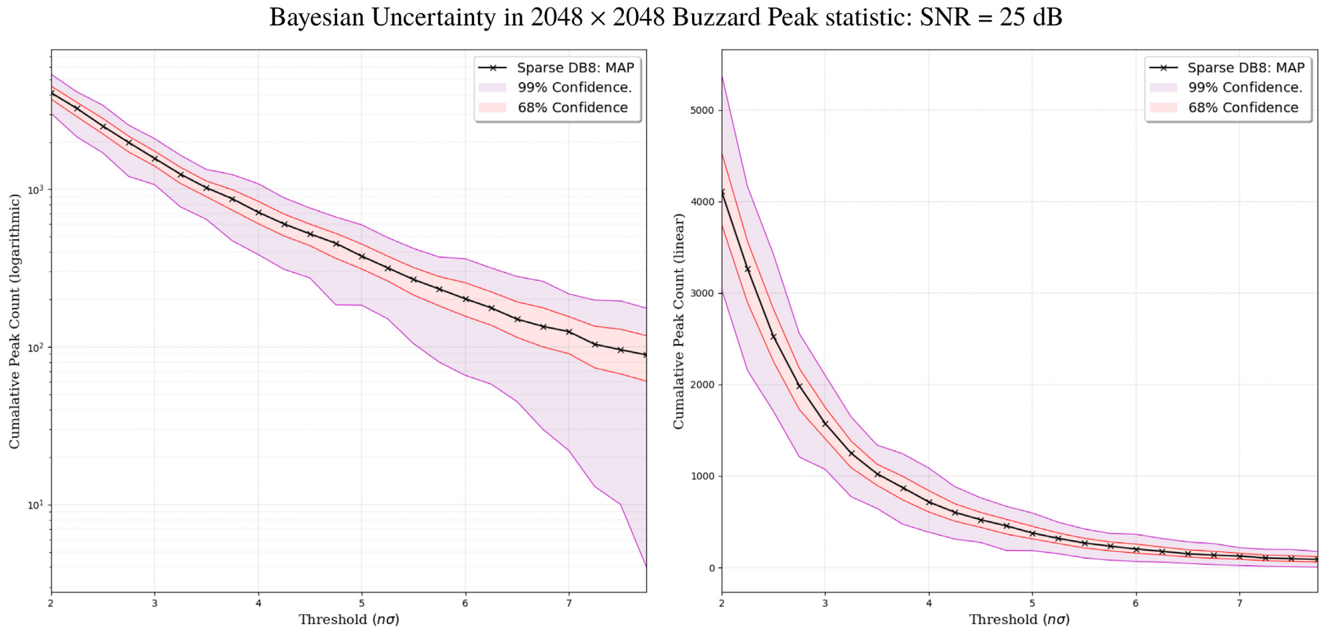
[1] Obtained due to our affiliation with the LSST-DESC collaboration.

[2] http://healpix.sourceforge.net/documentation.php



Bayesian Uncertainty in 2048 × 2048 Buzzard Peak statistic: SNR = 30 dB

Figure 7. Cumulative peak statistic for a 2048 × 2048 planar convergence map extracted from the Buzzard V-1.6 simulation (see Section 7.1) contaminated with i.i.d. Gaussian noise such that the discretized simulated shear (see Section 5.1) are of SNR 30 dB. The *purple* outer contours are the computed upper and lower bounds at 99 per cent confidence, with the inner *red* contours representing the 68 per cent ($\sim 1\sigma$) bounds, included to aid comparison to similar literature which typically quote 1σ errors. Note that the information content drops for higher σ thresholds as fewer peaks are present, leading to larger relative uncertainty as fewer samples are recovered. Further note that this example is computed in a highly idealized low-noise setting.



Bayesian Uncertainty in 2048 × 2048 Buzzard Peak statistic: SNR = 25 dB

Figure 8. Cumulative peak statistic for a 2048 × 2048 planar convergence map extracted from the Buzzard V-1.6 simulation (see Section 7.1) contaminated with i.i.d. Gaussian noise such that the discretized simulated shear (see Section 5.1) are of SNR 25 dB. The *red* inner contours represent the upper and lower bounds at 68 per cent ($\sim 1\sigma$) confidence, with the outer *purple* contours representing the computed bounds at 99 per cent confidence.

pixels) would have a similar effect to reducing the noise level per pixel.

For an SNR of 20 dB (Fig. 9) the first feature of note is the shaded blue region which indicates that for high thresholds the lower bound on the number of peaks at 99 per cent confidence is consistent (and clipped) at 0 – this is saying that at 99 per cent confidence the true number of peaks at a threshold in the blue shaded region could be 0. Note that in the blue region the Bayesian upper bound is still entirely valid, it is only the Bayesian lower bound which within our novel approach is no longer well defined.

Clearly the upper and lower bounds on the peak count statistic is dependent on the threshold one is considering and the total area over which observations are made – for wide-field surveys, more data are collected which is likely to reduce the variance of the statistic.

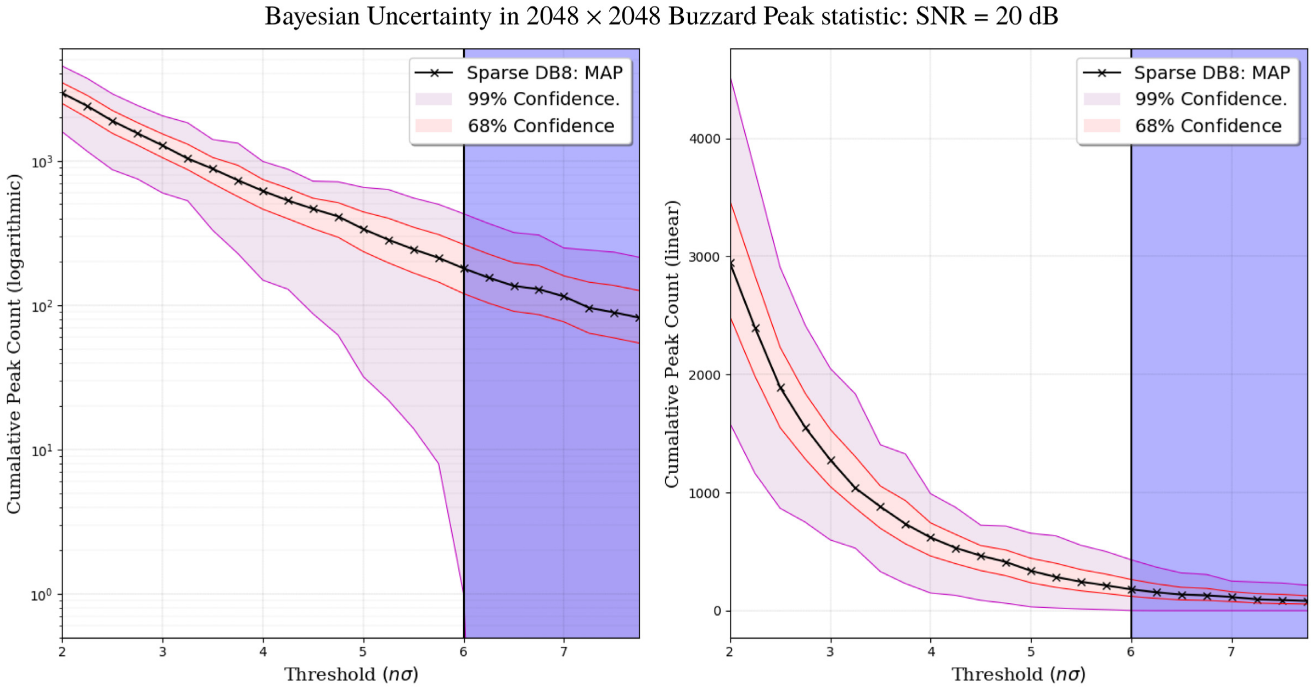


Figure 9. Cumulative peak statistic for a 2048 × 2048 planar convergence map extracted from the Buzzard V-1.6 simulation (see Section 7.1) contaminated with i.i.d. Gaussian noise such that the discretized simulated shear (see Section 5.1) are of SNR 20 dB. The *red* inner contours represent the upper and lower bounds at 68 per cent ($\sim 1\sigma$) confidence, with the outer *purple* contours representing the computed bounds at 99 per cent confidence. The shaded *blue* region indicates threshold values for which at 99 per cent confidence the data cannot rule out the possibility that no peaks exist above this threshold (note that in these regions the lower bound is technically 0 and there still exists a well-defined upper bound which is given). Comparing this plot to Fig. 7 we see that as the noise level increases the 68 per cent and 99 per cent confidence isocontours expand (as one would expect) and that in all cases the MAP peak statistics do not disagree at 99 per cent confidence.

In a general sense we summarize the mean (over all considered thresholds K) order of magnitude percentage spread on the peak statistic for the considered SNR thresholds below.

At input SNR of 20 dB, for thresholds $\in [2\sigma, 6\sigma)$ on a single 2048 × 2048 planar patch the upper and lower bounds exist and are of $\mathcal{O}(48$ per cent$)$ at 99 per cent confidence and of $\mathcal{O}(13$ per cent$)$ at 68 per cent.

At input SNR of 25 dB, for thresholds $\in [2\sigma, 8\sigma)$ on a single 2048 × 2048 planar patch the upper and lower bounds exist and are of $\mathcal{O}(25$ per cent$)$ at 99 per cent confidence and of $\mathcal{O}(7$ per cent$)$ at 68 per cent.

At input SNR of 30 dB, for thresholds $\in [2\sigma, 8\sigma)$ on a single 2048 × 2048 planar patch the upper and lower bounds exist and are of $\mathcal{O}(15$ per cent$)$ at 99 per cent confidence and of $\mathcal{O}(3$ per cent$)$ at 68 per cent.

These illustrative examples imply that for the Bayesian peak statistic to tightly constrain the cumulative peak statistic comparatively larger and or cleaner data sets may be required – or, of course, a more informative prior (though this must be well justified). However, to reduce the shot noise introduced *via* intrinsic ellipticities more galaxies must be observed within a given pixel.

One way to increase this is to simply increase the observed number density of galaxy observations, however to do so one must observe galaxies at lower magnitude (for a fixed redshift), which inherently leads to more bright distant galaxies being detected which results in galaxy blending. Hence, increasing the number density significantly above $\sim$30 gal arcmin^{-2} is typically quite difficult in practice.

Alternatively, the pixelization can be adjusted to ensure that the mean galaxy count per pixel is above a given threshold – though for weak lensing the majority of non-Gaussian information is stored at fine scales, which require small pixels, and so using larger pixels to reduce the noise level is suboptimal for information extraction.

Within the definition of the up- and down-scaling kernels (see sections 6.1 and 6.2) we define a circular aperture around a selected peak which we define to be the extent of the peak. These regions are roughly equivalent to superpixel regions as described in Cai et al. (2018a). In previous work it was shown (Price et al. 2018b) that for local credible intervals (cf. pixel level error bars) the typical error in the approximate HPD credible region is of $\mathcal{O}(12.5$ per cent$)$, and is conservative – note that the quoted 25 per cent mean RMSE error is split approximately equally between the upper and lower bounds, therefore this roughly corresponds to an mean error of 12.5 per cent on both. Therefore the bounds drawn on the peak static here are likely to be $\sim$ 12.5 per cent less tight than the true Bayesian bounds – which could be formed if one were to reconstruct the 4×10^6 dimensional posterior *via* MCMC.

In this paper (particularly the second half) we are primarily concerned with demonstrating how one may recover principled uncertainties on aggregate statistics of the convergence map – such as, but not limited to, the peak statistics. Hence we do not provide detailed analysis of how these Bayesian uncertainties may affect cosmological constraints derived from such statistics – this is saved for future work. However it is worth mentioning that one could either leverage these uncertainties to define the data covariance in a Bayesian manner (as opposed to MC which is fast but may not necessarily be fully principled, or MCMC which is $\mathcal{O}(10^6)$ times slower than our MAP approach) before and then running a standard likelihood analysis, or perform a grid search in parameter space using these uncertainties again as the data covariance. Correctly accounting the uncertainties introduced during mass mapping has been shown to be an important consideration for the future prospects of statistics such as this (Lin & Kilbinger 2018).

8 CONCLUSIONS

Using the sparse Bayesian mass-mapping framework previously developed (Price et al. 2018a,b) we have presented two novel Bayesian uncertainty quantification techniques which can be performed directly on weak lensing convergence maps.

The first of these techniques recovers the uncertainty in the location of a feature of interest within a reconstructed convergence map – e.g. a large peak – at some well-defined confidence. We call this locational uncertainty the *'Bayesian location'*.

Additionally, for computational efficiency we develop a novel sampling scheme of the position isocontour of a given feature which we call *'N-splitting circular bisection'*. We find that sampling the position isocontour in this way could be many orders of magnitude faster in high dimensions than typical inverse nesting approaches.

To evaluate this technique, we perform sparse Bayesian reconstructions of 1024 × 1024 convergence maps extracted from Bolshoi *N*-body simulation data sets upon which we compute the Bayesian location of the four largest subhaloes for a range of noise levels.

The second of these techniques quantifies the uncertainty in the cumulative peak statistic of a recovered convergence map. With this technique we can for the first time provide principled Bayesian lower and upper bounds on the number of observed peaks at a given signal-to-noise threshold, for a single observation, at well-defined confidence.

We extract 2048 × 2048 convergence maps from the Buzzard V-1.6 *N*-body simulation, upon which we calculate the cumulative peak statistic with Bayesian upper and lower bounds at 99 per cent for a range of input noise levels. We also provide the 68 per cent confidence bounds which we infer from the 99 per cent bounds to aid comparison to typical bootstrapping (MC) approaches.

For upcoming wide-field surveys convergence reconstruction will likely be done natively on the sphere (a single collective sample) to avoid projection effects, making bootstrapping approaches difficult and at worst infeasible due to the fact that they are only asymptotically exact.

Bayesian approaches require only a single set of observations to make exact inferences, and so extend trivially to the more complex spherical setting. Moreover the novel uncertainty quantification techniques presented in this paper and those presented previously in Price et al. (2018a,b) and Cai et al. (2018a) can be rapidly computed and support algorithmic structure which can be highly parallelized, making them the ideal tools for principled analysis of convergence maps.

ACKNOWLEDGEMENTS

Author contributions are summarized as follows. MAP: conceptualization, methodology, data curation, investigation, software, visualization, writing – original draft; JDM: conceptualization, methodology, project administration, supervision, writing – review and editing; XC: methodology, investigation, writing – review & editing; TDK: methodology, supervision, writing – review and editing.

This paper has undergone internal review in the Large Synoptic Space Telescope (LSST) Dark Energy Science Collaboration (DESC). The internal reviewers were Chihway Chang, Tim Eifler, and François Lanusse. MAP is supported by the Science and Technology Facilities Council (STFC). TDK is 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.

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., 2003, Phys. Rev. D, 67, 023501
Cai X., Pereyra M., McEwen J. D., 2018a, MNRAS, 480, 4170
Cai X., Pereyra M., McEwen J. D., 2018b, MNRAS, 480, 4154
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
Coles P., Chiang L.-Y., 2000, Nature, 406, 376
Combettes P. L., Pesquet J. C., 2011, Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer New York, p. 185
Corless V. L., King L. J., Clowe D., 2009, MNRAS, 393, 1235
Durmus A., Moulines E., Pereyra M., 2018, SIAM J. Imaging Sci., 11, 473
Fluri J., Kacprzak T., Sgier R., Refregier A., Amara A., 2018, J. Cosmol. Astropart. Phys., 10, 51
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Grimm N., Yoo J., 2018, J. Cosmol. Astropart. Phys., 7, 67
Heavens A., 2009, Nucl. Phys. B, 194, 76
Horowitz B., Seljak U., Aslanyan G., 2018, preprint (arXiv:1810.00503)
Jee M. J., Dawson W. A., Stroe A., Wittman D., van Weeren R. J., Braggen M., Brada M., Rattgering H., 2016, ApJ, 817, 179
Jeffrey N. et al., 2018, MNRAS, 479, 2871
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kilbinger M., 2015, Rep. Prog. 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
Lin C.-A., 2016, preprint (arXiv:1612.04041)
Lin C.-A., Kilbinger M., 2018, A&A, 614, A36
Mediavilla E., Muñoz J. A., Garzón F., Mahoney T. J., 2016, Astrophysical Applications of Gravitational Lensing. Cambridge Univ. Press, Cambridge
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
Peel A., Lin C.-A., Lanusse F., Leonard A., Starck J.-L., Kilbinger M., 2017a, A&A, 599, A79
Peel A., Lanusse F., Starck J.-L., 2017b, ApJ, 847, 23

Peel A., Pettorino V., Giocoli C., Starck J.-L., Baldi M., 2018, A&A, 619, A38

Pereyra M., 2015, Proximal Markov Chain Monte Carlo algorithms, Stat Comput, Springer US,

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. IEEE, Nice, France, p. 230

Pires S., Starck J.-L., Amara A., Réfrégier A., Teyssier R., 2010, in Alimi J.-M., Fuözfa A., eds, AIP Conf. Proc. Vol. 1241, Cosmological model discrimination from weak lensing data. Am. Inst. Phys., New York, p. 1118

Planck Collaboration VI, 2018, preprint (arXiv:1807.06209)

Price M. A., Cai X., McEwen J. D., Kitching T. D., Wallis C. G. R., 2018a, MNRAS, preprint (arXiv:1812.04014)

Price M. A., Cai X., McEwen J. D., Pereyra M., Kitching T. D., 2018b, MNRAS, preprint (arXiv:1812.04017)

Robert C.-P., 2001, The Bayesian Choice, Springer Science and Business Media.

Schneider P., 2005, 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

Taylor P. L., Kitching T. D., McEwen J. D., Tram T., 2018, Phys. Rev. D, 98, 023522

Trotta R., 2017, preprint (arXiv:1701.01467)

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, preprint (arXiv:1703.09233)

APPENDIX A: N-SPLITTING CIRCULAR BISECTION DETAILS

In this appendix we consider the N-splitting circular bisection (N-splitting) algorithm for iteratively sampling the Bayesian $100(1-\alpha)$ per cent confidence isocontour of the position of a feature in a reconstructed convergence map – or the *Bayesian Location* at $100(1-\alpha)$ per cent confidence.

As in the text, we begin by defining the number of directions to sample n_T from which we then form the angular increment $\Delta\Theta = 2\pi/n_T$. Starting from a vector $\mathbf{n}_0$ oriented along the positive y-axis define the $(i+1)^{\text{th}}$ pointing to be the vector,

$$\mathbf{n}_{i+1} = \mathcal{R}_{\Delta\Theta}\mathbf{n}_i, \quad \text{where} \quad i \in (1, n_T), \tag{A1}$$

and where $\mathcal{R}_{\Delta\Theta}$ is rotation by angle $\Delta\Theta$ clockwise on 2D Euclidean space – a irreducible representation of which is the standard clockwise rotation matrix,

$$\mathcal{R}_\theta = \begin{bmatrix} \cos(\Delta\Theta) & \sin(\Delta\Theta) \\ -\sin(\Delta\Theta) & \cos(\Delta\Theta) \end{bmatrix}. \tag{A2}$$

Now we know the direction along which we wish to sample we construct the $(i+1)^{\text{th}}$ bisection problem which is

$$d_\alpha^{i+1} = \min_d \left\{ d \in \Gamma_{i+1} \mid f(\kappa_d^{\text{sgt}}) + g(\kappa_d^{\text{sgt}}) > \epsilon'_\alpha \right\}, \tag{A3}$$

where κ_d^{sgt} is a surrogate convergence map with the feature of interest inserted into perturbed location $d\mathbf{n}_{i+1}$ and Γ_{i+1} is subset of the real domain which lie on the directional line centred at the original peak location with unit vector $\mathbf{n}_{i+1}$ i.e.

$$\Gamma_{i+1} = \left\{ a\mathbf{n}_{i+1} \mid a \in \mathbb{R}_+ \right\}. \tag{A4}$$

A pictorial representation of how the problem is set-up is provided in Fig. A1.

For bisection we must first make an initial guess d_0 which we define to be square root of the number of pixels contained within the mask, as this is a typical measure of the length of a masked region. This choice is particularly logical as, if a feature of interest can be removed entirely from its masked location without saturating the level-set threshold ϵ'_α then it by definition must be inconclusive, i.e. the data are insufficient evidence to say that the peak is physical.

To optimize the convergence of this algorithm further (for high sampling rates, low angular increments $\Delta\Theta \leq \pi/4$) we also propagate information between pointing's. For bisection problems associated with pointing $i > 1$ the initial guess is now set to be twice the previous optimal value $d_\alpha^{\prime,i}$. This increases the computational efficiency by ~ 20 per cent, in most cases.

Propagating information in this way relies on the assumption that the isocontour we are searching for is somewhat smooth and continuous, which is the case for most convergence reconstructions. If there is uncertainty as to the smoothness of the isocontour it is recommended that information is not propagated and the number of pointings is increased to correctly map the isocontour structure.

A1 Convergence properties

Standard inverse nesting algorithms iteratively sample the entire sub-space of the reconstructed domain bounded by the isocontour at $100(1-\alpha)$ per cent confidence, making them inefficient when one is only interested in the boundary.

Consider the case where the isocontour of a reconstructed 512×512 convergence map is a circular region of radius R. Here inverse nesting will have to sample a square region out to R, which is to say the total number of samples T_{nest} will at least be $R^2 - 1$, where 1 is removed for the central location.

For our N-splitting algorithm we define n_T pointings, and assume that the isocontour is relatively smooth. As the first bisection problem $\mathbf{n}_0$ makes a large first guess it typically takes 4–5 iterations to converge with a single pixel accuracy. The subsequent $n_T - 1$ bisection problems converge within 3–4 iterations. Therefore the total number of calculations $T_{\text{N-split}}$ is conservatively

$$T_{\text{N-split}} = 5 + 4(n_T - 1), \tag{A5}$$

which is essentially independent from R. There is in fact a small inverse dependence which is incorporated in the number of iterations needed for convergence, though this dependence is found to be small.

Comparing the computational efficiency of the two algorithms $\mathbb{E}_{512}$ where

$$\mathbb{E}_{512} \equiv \frac{T_{\text{N-split}}}{T_{\text{nest}}} = \frac{5 + 4(n_T - 1)}{R^2}. \tag{A6}$$

Typically, we find an angular separation between pointings of $\pi/4$ (i.e. 16 pointings) is sufficient to accurately recover the isocontour. Additionally, the circular radius is typically 15–30 pixels which indicates that

$$\frac{5 + 4 \times 15}{30^2} = 0.072 \leq \mathbb{E}_{512} \leq \frac{5 + 4 \times 15}{15^2} = 0.289, \tag{A7}$$

i.e. N-splitting circular bisection on 512×512 dimensional reconstructions is ∼4–14 times faster than inverse nesting.

However, in the future we will be interested in recovering high-dimensional 2048×2048 convergence maps. In this setting the number of iterations for N-splitting to converge is assumed to change by 1–2, and the number of pointings to faithfully recover the isocontour will be increase by a factor of ∼2. Additionally, the

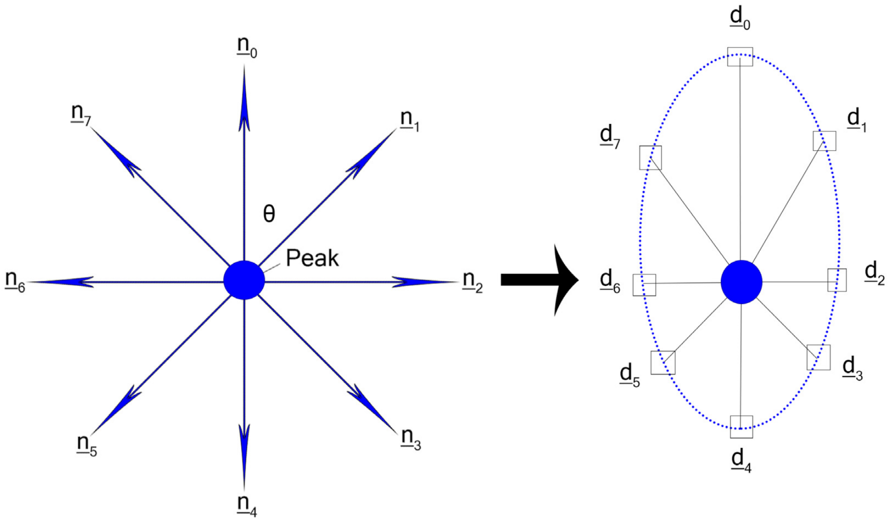

$\underline{n}_0$
$\underline{n}_1$
$\underline{n}_2$
$\underline{n}_3$
$\underline{n}_4$
$\underline{n}_5$
$\underline{n}_6$
$\underline{n}_7$
θ
Peak
$\underline{d}_0$
$\underline{d}_1$
$\underline{d}_2$
$\underline{d}_3$
$\underline{d}_4$
$\underline{d}_5$
$\underline{d}_6$
$\underline{d}_7$

Figure A1. Representation of how the problem is broken up in N-splitting circular bisection. First the n_i directions are specified (*left*) at equiangular separations θ about the peak location (*blue ball*). Bisection iterations are conducted as in equation (A3) along each of the directions, recovering a set of samples d_i of the Bayesian location isocontour at $100(1 - \alpha)$ per cent confidence (*right*). Provided a sufficient number of samples are taken, this boundary will fully represent the isocontour. We find typically ≈16 samples are needed for 512 × 512 convergence reconstructions though more or less may be needed depending on the resolution and application.

radius of the circle R increases by a factor of 4. Thus,

$$\frac{5 + 4 \times 31}{120^2} = 0.009 \leq \mathbb{E}_{2048} \leq \frac{5 + 4 \times 31}{60^2} = 0.0360, \tag{A8}$$

i.e. the conservative increase in computational efficiency of N-splitting over inverse nesting for 2048 × 2048 becomes a factor of ≈30–112.

Further optimizations are possible, such as trivially parallelizing the bisection problems of each pointing. Doing so removes the scaling with the number of pointings, but now information about starting positions cannot be propagated.

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