Sparse Bayesian mass mapping with uncertainties: local credible intervals

M. A. Price,[1]★ X. Cai,[1] J. D. McEwen,[1] M. Pereyra[2] and T. D. Kitching[1]
for the LSST Dark Energy Science Collaboration

[1]*Mullard Space Science Laboratory, University College London, Holmbury Hill Rd, Dorking RH5 6NT, UK*
[2]*Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK*

ABSTRACT

Until recently, mass-mapping techniques for weak gravitational lensing convergence reconstruction have lacked a principled statistical framework upon which to quantify reconstruction uncertainties, without making strong assumptions of Gaussianity. In previous work, we presented a sparse hierarchical Bayesian formalism for convergence reconstruction that addresses this shortcoming. Here, we draw on the concept of *local credible intervals* (cf. Bayesian error bars) as an extension of the uncertainty quantification techniques previously detailed. These uncertainty quantification techniques are benchmarked against those recovered via Px-MALA – a state-of-the-art proximal Markov chain Monte Carlo (MCMC) algorithm. We find that, typically, our recovered uncertainties are everywhere conservative (never underestimate the uncertainty, yet the approximation error is bounded above), of similar magnitude and highly correlated with those recovered via Px-MALA. Moreover, we demonstrate an increase in computational efficiency of $\mathcal{O}(10^6)$ when using our sparse Bayesian approach over MCMC techniques. This computational saving is critical for the application of Bayesian uncertainty quantification to large-scale stage IV surveys such as LSST and *Euclid*.

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

1 INTRODUCTION

As photons from distant sources (galaxies) travel through space–time to us, here and now their trajectories are perturbed by local mass over- and underdensities, causing the observed shapes of structures to be warped, or *gravitationally lensed*. This cosmological effect is sensitive to all matter (both visible and invisible), and so provides a natural cosmological probe of dark matter.

The gravitational lensing effect has (at first order) two distinct effects: distant galaxies are magnified by a convergence field κ and the third flattening (ellipticity) is perturbed from an underlying intrinsic value by a shear field γ. A wide range of cosmologies can be extracted from just the shear field (Alsing et al. 2016; Taylor et al. 2018), though increasingly higher order statistics (Coles & Chiang 2000; Munshi et al. 2008; Heavens 2009; Munshi & Coles 2017) are being computed on convergence maps directly.

As a result of the mass-sheet degeneracy (an a priori degeneracy of the intrinsic brightness of galaxies; see Bartelmann & Schneider 2001), the convergence field cannot be observed directly. Instead, measurements of the shear field γ must be taken and inverted through some mapping to create an estimator for κ. Typically, these inverse problems are ill-posed (often seriously) and so creating unbiased estimators for the convergence κ can prove difficult.

Many convergence inversion techniques have been considered (e.g. VanderPlas et al. 2011; Lanusse et al. 2016; Wallis et al. 2017; Chang et al. 2018; Jeffrey et al. 2018) though the simplest, most direct method in the planar setting is that of Kaiser–Squires (KS) inversion (Kaiser & Squires 1993). Though these methods often produce reliable estimates of κ, they all either lack principled statistical uncertainties on their reconstructions or make strong assumptions of Gaussianity (which heavily degrades the quality of non-Gaussian information in particular).

For example, Wiener filtering (see e.g. Horowitz, Seljak & Aslanyan 2019) directly adopts Gaussian priors that are more explicit assumptions of Gaussianity. Other approaches, such as the KS method, recover a noisy convergence estimate that is post-processed via convolution with a Gaussian kernel, which promotes Gaussianity.

On large scales, the lensing information is primarily Gaussian in nature, though on smaller scales (at higher resolutions) there occurs a non-negligible non-Gaussian contribution that encodes information about baryonic interactions and clustering among other non-linear effects. Analysis of such effects is expected (Munshi et al. 2008) to provide competitive and more importantly complementary constraints on cosmological parameters – in particular, parameters related closely to dark matter such as σ_8 and $\Omega_{\rm M}$. Consequently, mapping techniques that preserve the non-Gaussian information content are a crucial step forward for dark matter analysis via weak gravitational lensing.

★ E-mail: m.price.17@ucl.ac.uk

In previous work (Price et al. 2018), we presented a new sparse hierarchical Bayesian formalism for reconstructing the convergence field. This not only regularizes the ill-posed inverse problem but also allows us to explore the Bayesian posterior in order to recover principled uncertainties on our reconstruction. It is important to note here that this mathematical framework is entirely general and can be applied for any posterior that belongs to the set of log-concave functions – of which both sparsity enforcing Laplace-type priors and standard Gaussian priors are members.

Often, hierarchical Bayesian inference problems are solved by *Markov chain Monte Carlo* (MCMC) techniques (see e.g. Trotta 2017), which explicitly return a large number of samples from the full posterior distribution – from which one can construct true Bayesian uncertainties. Samples of the posterior via MCMC algorithms construct theoretically optimal estimates of the posterior (in the limit of a large number of samples), but in practice can be extremely computationally taxing to recover fully.

In fact, when the dimensionality becomes large these methods become infeasible – often referred to as *the curse of dimensionality*. In the context of lensing inverse problems, each pixel constitutes a dimension, and so for a resolution of 1024×1024 (which is typical) the dimension of the problem is $\mathcal{O}(10^6)$.

Recent advancements in probability density theory (Robert 2001) allow conservative approximations of Bayesian credible regions of the posterior from knowledge of the maximum a posteriori (MAP) solution alone (Pereyra 2017). The sparse Bayesian method presented in previous work (see Price et al. 2018) recasts the maximization of the posterior distribution as a convex optimization problem from which the MAP solution can be rapidly computed. Uncertainty quantification is then conducted utilizing the aforementioned approximate credible regions of the posterior. In Price et al. (2018), hypothesis testing (determining the statistical significance of a feature of the recovered convergence map) was introduced to the weak lensing setting as a form of uncertainty quantification.

In this article, we introduce a further uncertainty quantification technique called *local credible intervals* (cf. pixel-level error bars). Both hypothesis testing and local credible intervals were previously developed and applied to the radio interferometric setting (Cai, Pereyra & McEwen 2017a, b). We also remark that there are alternative ways of testing image structures (Repetti, Pereyra & Wiaux 2018). This paper serves as a benchmark comparison of our sparse hierarchical Bayesian formalism (see Price et al. 2018) to a bespoke MCMC algorithm, Px-MALA (Durmus, Moulines & Pereyra 2016; Pereyra 2016; Cai, Pereyra & McEwen 2017a, b). Px-MALA utilizes Moreau–Yosida envelopes and proximity operators (tools from convex analyses) to support non-differentiable terms in the prior or likelihood, making it one of the only somewhat efficient ways to support non-smooth sparsity-promoting priors (on which our sparse Bayesian mass-mapping framework is based) in high-dimensional settings.

The remainder of this article is structured as follows. We begin with Section 2 in which we review our sparse hierarchical Bayesian models for mass mapping and present a brief overview of the Px-MALA MCMC algorithm. We then cover the relevant mathematical background of approximate Bayesian uncertainty quantification in Section 3 before introducing the concept of *local credible intervals* – an additional form of uncertainty quantification. In Section 4, we conduct a series of mock scenarios to compare the uncertainties recovered by our MAP approach, and the full MCMC (Px-MALA) treatment. Finally, we draw conclusions and discuss future work in Section 5.

Section 2 relies on a strong understanding of Bayesian inference and MCMC techniques along with a moderate understanding of proximal calculus and compressed sensing. As such, for the reader interested only in the application and benchmarking, Section 4 onwards is relevant content.

2 HIERARCHICAL BAYESIAN INFERENCE FOR MASS MAPPING

Hierarchical Bayesian models provide a flexible, well-defined approach for dealing with uncertainties in a variety of problems. For an overview of Bayesian hierarchical modelling and MCMC techniques in the context of astrophysics, we refer the reader to Trotta (2017).

We begin by presenting an overview of the sparse hierarchical Bayesian approach developed in previous work (see Price et al. 2018), where we also review the weak lensing planar forward model. Following this, we make the MAP optimization problem explicit. We then review the Bayesian parameter inference hierarchy adopted in our sparse Bayesian mass-mapping algorithm (Price et al. 2018). Finally, we provide a short introduction to the Px-MALA and MYULA proximal MCMC algorithms (Durmus et al. 2016; Pereyra 2016).

2.1 Bayesian inference

Mathematically, let us begin by considering the *posterior distribution*, which by Bayes' theorem is given by

$$p(\kappa|\gamma) = \frac{p(\gamma|\kappa)p(\kappa)}{\int_{\mathbb{C}^N} p(\gamma|\kappa)p(\kappa)\mathrm{d}\kappa}. \tag{1}$$

Bayes' theorem relates the posterior distribution $p(\kappa|\gamma)$ to the product of some likelihood function $p(\gamma|\kappa)$ and some prior $p(\kappa)$. It is important to note here that a model is implicit which collectively defines the noise and the proposed relationship between observations γ and inferences κ – specifically, this term characterizes the noise model and the assumed mapping $\{\kappa \mapsto \gamma\}$. Note that the denominator in equation (1) is the model's marginal likelihood, which is unrelated to κ.

Suppose the discretized complex shear field $\gamma \in \mathbb{C}^M$ and the discretized complex convergence field $\kappa \in \mathbb{C}^N$ – where M represents the number of binned shear measurements and N represents the dimensionality of the convergence estimator – are related by a measurement operator $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ defined such that

$$\mathbf{\Phi} \in \mathbb{C}^{M \times N} : \kappa \in \mathbb{C}^N \mapsto \gamma \in \mathbb{C}^M. \tag{2}$$

Further, suppose a contaminating noise n is present. Measurements of γ are produced *via*

$$\gamma = \mathbf{\Phi}\kappa + n. \tag{3}$$

For the case considered within this paper, we take $n \sim \mathcal{N}(0, \sigma_n^2) \in \mathbb{C}^M$, i.e. i.i.d. (independent and identically distributed) additive Gaussian noise. For the purpose of this paper, we consider the simplest planar mapping,

$$\mathbf{\Phi} = \mathbf{F}^{-1}\mathbf{D}\mathbf{F}. \tag{4}$$

Here, $\mathbf{F}$ ($\mathbf{F}^{-1}$) is the forward (inverse) discrete fast Fourier transforms and $\mathbf{D}$ is the weak lensing planar forward model in Fourier space (e.g. Kaiser & Squires 1993),

$$\mathbf{D}_{k_x, k_y} = \frac{k_x^2 - k_y^2 + 2ik_xk_y}{k_x^2 + k_y^2}. \tag{5}$$

The measurement operator $\boldsymbol{\Phi}$ has also been extended to superresolution image recovery (Price et al. 2018), but that is beyond the scope of this paper.

In the majority of weak lensing surveys, $M < N$ (i.e. the shear field is a discrete undersampling of the underlying convergence field) and so inverting the forward model is typically ill-posed (often seriously). To regularize ill-posed inverse problems, a term encoding prior information is introduced – this is referred to as either the prior or regularization term.

We choose a prior that reflects the quasi-philosophical notion of *Occam's Razor* – a prior that says if two solutions are equally viable, the one that makes the fewest assumptions (the fewest active variables – non-zero coefficients in a sparse domain) is more likely to be true. Mathematically, this is equivalent to imposing sparsity that minimizes the number of non-zero coefficients in a sparse representation (dictionary).

One could select any sparsifying domain, though a natural choice for most physical systems is wavelets. We choose to use wavelets as our sparsifying dictionary in this paper and in previous work.

The natural sparsity-promoting prior is the ℓ_0-norm $\|.\|_0$, often referred to as the *Hamming distance*, i.e. the total number of non-zero coefficients of a field. However, this function is non-differentiable and (perhaps more importantly) non-convex. As such, it cannot exploit the computational advantages provided by conventional convex optimization techniques.

Researchers therefore often select the next most natural sparsity-promoting prior, the ℓ_1-norm $\|.\|_1$, which is convex and can be shown to share the same MAP solution as if one were to use the ℓ_0-norm in certain cases (see e.g. Donoho 2006; Candès & Wakin 2008, on *convex relaxation*).

We now define the likelihood function (data fidelity term) as a multivariate Gaussian with diagonal covariance $\Sigma = \sigma_n^2 \mathbb{I}$ such that

$$p(\gamma|\kappa) \propto \exp\left(\frac{-\|\boldsymbol{\Phi}\kappa - \gamma\|_2^2}{2\sigma_n^2}\right), \tag{6}$$

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

$$p(\kappa) \propto \exp\left(-\mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1\right), \tag{7}$$

where $\boldsymbol{\Psi}$ is an appropriately selected sparsifying dictionary (such as a wavelet dictionary) in which the signal is assumed to be sparse, and $\mu \in \mathbb{R}_+$ is a regularization parameter. Substituting $p(\gamma|\kappa)$ and $p(\kappa)$ into equation (1) yields

$$p(\kappa|\gamma) \propto \exp\left\{-\left(\mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1 + \frac{\|\boldsymbol{\Phi}\kappa - \gamma\|_2^2}{2\sigma_n^2}\right)\right\}. \tag{8}$$

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 2019, for alternate iterative Weiner filtering approaches).

2.2 Sparse MAP estimator

Drawing conclusions directly from $p(\kappa|\gamma)$ can be difficult because of the high dimensionality involved, which will be detailed in the next section. As an alternative, Bayesian methods often derive solutions by computing estimators that summarize $p(\kappa|\gamma)$, such as maximizing the probability of the recovered κ conditional on the data γ. Such a solution is referred to as the MAP solution. From the

monotonicity of the logarithm function, it is evident that

$$\begin{aligned}
\kappa^{\mathrm{map}} &= \underset{\kappa}{\mathrm{argmax}}\left\{p(\kappa|\gamma)\right\} \\
&= \underset{\kappa}{\mathrm{argmin}}\left\{-\log(p(\kappa|\gamma))\right\} \\
&= \underset{\kappa}{\mathrm{argmin}}\left\{\underbrace{\mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1}_{f(\kappa)} + \underbrace{\|\boldsymbol{\Phi}\kappa - \gamma\|_2^2/2\sigma_n^2}_{g(\kappa)}\right\},
\end{aligned} \tag{9}$$

which is a convex minimization problem and can therefore be computed in a highly computationally efficient manner.

To solve the convex minimization problem given in equation (9), we implement an adapted forward–backward splitting algorithm (Combettes & Pesquet 2009). A complete description of the steps adopted when solving this optimization problem, and the full details of the sparse hierarchical Bayesian formalism are outlined in previous work (Cai et al. 2017b; Price et al. 2018).

2.3 Sparse dictionary and regularization parameter

Here we provide a concise overview of the parameter selection aspect of our sparse Bayesian mass-mapping algorithm, which was developed and presented in previous work – for a complete description, see Price et al. (2018) and Pereyra, Bioucas-Dias & Figueiredo (2015).

The prior term in equation (9) promotes the a priori knowledge that the signal of interest κ is likely to be sparse in a given dictionary $\boldsymbol{\Psi}$. A function $f(x)$ is sparse in a given dictionary $\boldsymbol{\Psi}$ if the number of non-zero coefficients is small compared to the total size of the dictionary domain. Wavelets form a general set of naturally sparsifying dictionaries for a wide range of physical problems, and have recently been shown to work well in the weak lensing setting (Lanusse et al. 2016; Peel, Lanusse & Starck 2017; Jeffrey et al. 2018; Price et al. 2018). For the purpose of this paper, we restrict ourselves to Daubechies 8 (DB8) wavelets (with eight wavelet levels) for simplicity, though a wide variety of wavelets could be considered (e.g. Pires, Starck & Amara 2009; Carrillo, McEwen & Wiaux 2012; Starck, Murtagh & Fadili 2015). Note that the exact choice of wavelet representation (and prior) is independent from the results of this benchmarking paper, thus a discussion of dictionary (and prior) optimality is not of primary concern.

An issue in these types of regularized optimization problems is the setting of regularization parameter μ – several approaches have been presented (Paykari et al. 2014; Lanusse et al. 2016; Peel et al. 2017; Jeffrey et al. 2018). For uncertainties on reconstructed κ maps to be truly principled, μ must be computed in a well-defined, statistically principled way. In Price et al. (2018), a hierarchical Bayesian inference approach to compute the theoretically optimal μ was adopted, which we outline in Appendix A.

2.4 Proximal MCMC sampling

Sampling a full posterior distribution is very challenging in high-dimensional settings, particularly when the prior $p(\kappa)$ considered is non-differentiable – like the sparsity-promoting prior given in equation (7). In the following, we recall two proximal MCMC methods developed in Durmus et al. (2016) and Pereyra (2016) – MYULA and Px-MALA – which can be applied to sample the full posterior density $p(\kappa|\gamma)$ for mass mapping. After a set of samples has been obtained, various kinds of analysis can be performed, such as summary estimators of κ, and a range of uncertainty quantification techniques, as presented in Cai et al. (2017a, b).

For a probability density $p \in \mathcal{C}^1$ with Lipschitz gradient, the Markov chain of the unadjusted Langevin algorithm (ULA) to generate a set of samples $\{\mathbf{l}^{(m)}\}_{\in \mathbb{C}^N}$ based on a forward Euler–Maruyama approximation with a step size $\delta > 0$ has the form

$$\mathbf{l}^{(m+1)} = \mathbf{l}^{(m)} + \frac{\delta}{2} \nabla \log p[\mathbf{l}^{(m)}] + \sqrt{\delta} \mathbf{w}^{(m+1)}, \qquad (10)$$

where $\mathbf{w}^{(m+1)} \sim \mathcal{N}(0, \mathbb{1}_N)$ (a N-sequence of standard Gaussian random variables).

However, the chain generated by ULA given above converges to p with asymptotic bias. This kind of bias can be corrected at the expense of some additional estimation variance (Roberts & Tweedie 1996) after involving a Metropolis–Hasting (MH) accept–reject step in ULA, which results in the MALA algorithm (Metropolis-adjusted Langevin Algorithm). However, the convergence of ULA and MALA is limited to a continuously differentiable $\log p$ with Lipschitz gradient, which prohibits their application to our focus on mass mapping with non-differentiable sparsity-promoting prior in equation (7).

Proximal MCMC methods – such as MYULA and Px-MALA – can be used to address non-differentiable sparsity-promoting priors (Durmus et al. 2016; Pereyra 2016). Without loss of generality, consider a log-concave posterior that is of the exponential family

$$p(\kappa|\gamma) \propto \exp\{-f(\kappa) - g(\kappa)\}, \qquad (11)$$

for lower semicontinuous convex and Lipschitz differentiable log-likelihood $g(x) \in \mathcal{C}^1$ and lower semicontinuous convex log-prior $f(x) \notin \mathcal{C}^1$. It is worth noting that this is precisely the setting adopted within this paper, where from (8)

$$f(\kappa) = \mu \|\mathbf{\Psi}^\dagger \kappa\|_1, \quad \text{and} \quad g(\kappa) = \|\mathbf{\Phi}\kappa - \gamma\|_2^2 / 2\sigma_n^2. \qquad (12)$$

To sample this posterior, the gradient $\nabla \log p$ is required; however, $f(x)$ is not Lipschitz differentiable. To account for the non-differentiability of $f(x)$, let us now define the smooth approximation $p_\lambda(\kappa|\gamma) \propto \exp\{-f^\lambda(\kappa) - g(\kappa)\}$, where

$$f^\lambda(\kappa) \equiv \min_{\hat{\kappa} \in \mathbb{C}^N} \left\{ f(\hat{\kappa}) + \|\hat{\kappa} - \kappa\|^2 / 2\lambda \right\} \qquad (13)$$

is the λ-Moreau–Yosida envelope of f, which can be made arbitrarily close to f by letting $\lambda \to 0$ (see Parikh & Boyd 2014). Then we have $\lim_{\lambda \to 0} p_\lambda(\kappa|\gamma) = p(\kappa|\gamma)$, and more importantly that, for any $\lambda > 0$, the total-variation distance between the distributions p_λ and p is bounded by $\|p_\lambda - p\|_{TV} \leq \lambda \mu N$, providing an explicit bound on the estimation errors involved in using p_λ instead of p (see Durmus et al. 2016, for details). Also, the gradient $\nabla \log p_\lambda = -\nabla f^\lambda - \nabla g$ is always Lipschitz continuous, with $\nabla f^\lambda(\kappa) = (\kappa - \text{prox}_f^\lambda(\kappa))/\lambda$, where $\text{prox}_f^\lambda(\kappa)$ is the *proximity operator* of f at κ defined as

$$\text{prox}_f^\lambda(\kappa) \equiv \operatorname*{argmin}_{\hat{\kappa} \in \mathbb{C}^N} \left\{ f(\hat{\kappa}) + \|\hat{\kappa} - \kappa\|^2 / 2\lambda \right\}. \qquad (14)$$

Replacing $\nabla \log p$ by $\nabla \log p_\lambda$ in the Markov chain of ULA and MALA given in (10) yields

$$\begin{aligned}\mathbf{l}^{(m+1)} =&\ \left(1 - \frac{\delta}{\lambda}\right) \mathbf{l}^{(m)} + \frac{\delta}{\lambda} \text{prox}_f^\lambda(\mathbf{l}^{(m)}) - \delta \nabla g(\mathbf{l}^{(m)}) \\ &+ \sqrt{2\delta} \mathbf{w}^{(m)},\end{aligned} \qquad (15)$$

which is named the MYULA algorithm (Moreau–Yosida regularized ULA). The MYULA chain (15), with small λ, efficiently delivers samples that are approximately distributed according to the posterior $p(\kappa|\gamma)$. By analogy with the process used to obtain MALA from ULA, we create the Px-MALA (proximal MALA)

after involving an MH (Metropolis–Hasting) accept–reject step in MYULA.

Essentially, the main difference of the two proximal MCMC methods (MYULA and Px-MALA) is that Px-MALA includes a Metropolis–Hastings step, which is used to correct the bias present in MYULA. Therefore, Px-MALA can provide results with more accuracy, at the expense of a higher computational cost and slower convergence (Pereyra 2016). Note, however, that these MCMC methods (as with any MCMC method) will suffer when scaling to high-dimensional data. Refer to, e.g. Durmus et al. (2016), Pereyra (2016), and Cai et al. (2017a) for a more detailed description of the proximal MCMC methods.

In this article, akin to the experiments performed in Cai et al. (2017b), we use the proximal MCMC method Px-MALA as a benchmark in the subsequent numerical tests presented in this work.

3 APPROXIMATE BAYESIAN UNCERTAINTY QUANTIFICATION

Though MAP solutions are theoretically optimal (most probable, given the data), one is often interested in the posterior distribution about this MAP point estimate – a necessity if one wishes to be confident in one's result. As described in Section 2.4, we can recover this posterior distribution completely using proximal MCMC techniques such as Px-MALA. However, these approaches are highly computationally demanding. They are feasible in the planar setting at a resolution of 256×256, where computation is of $\mathcal{O}(30 \text{ hours})$, but quickly become unrealistic for high resolutions.

More fundamentally, if we extend mass mapping from the planar setting to the spherical setting (Wallis et al. 2017), the wavelet and measurement operators become more complex – fast Fourier transforms are replaced with full spherical harmonic transforms – and the recovery of the posterior *via* MCMC techniques becomes highly computationally challenging at high resolutions.

In stark contrast to traditional MCMC techniques, recent advances in probability density theory have paved the way for efficient calculation of theoretically conservative approximate Bayesian credible regions of the posterior (Pereyra 2017). This approach allows us to extract useful information from the posterior without explicitly having to sample the full posterior. Crucially, this approach is shown to be many orders of magnitude less computationally demanding than state-of-the-art MCMC methods (Cai et al. 2017a) and can be parallelized and distributed.

In the following section, we formally define the concept of a Bayesian credible region of the posterior. We discuss limitations of computing these credible regions and highlight recently proposed approximations to Bayesian credible region. Finally, we outline recently developed computationally efficient uncertainty quantification techniques, which can easily scale to high-dimensional data. Specifically, we introduce the concept of *local credible intervals* (cf. pixel level error bars) presented first in Cai et al. (2017b) to the weak lensing setting.

3.1 Highest posterior density

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

$$p(\kappa \in C_\alpha|\gamma) = \int_{\kappa \in \mathbb{C}^N} p(\kappa|\gamma) \mathbb{I}_{C_\alpha} d\kappa = 1 - \alpha. \qquad (16)$$

Generally, there are many regions that satisfy this constraint. The minimum volume, and thus decision-theoretical optimal (Robert

2001), region is the *highest posterior density* (HPD) credible region, defined to be

$$C_\alpha := \{\kappa : f(\kappa) + g(\kappa) \leq \epsilon_\alpha\}, \tag{17}$$

where $f(\kappa)$ is the prior and $g(\kappa)$ is the data fidelity (likelihood) term. In the above equation, ϵ_α is an isocontour (i.e. level-set) of the log-posterior set such that the integral constraint in equation (16) is satisfied. In practice, the dimension N of the problem is large and the calculation of the true HPD credible region is difficult to compute.

Recently, a conservative approximation of C_α has been derived (Pereyra 2017), which can be used to tightly constrain the HPD credible region without having to explicitly calculate the integral in equation (16):

$$C'_\alpha := \{\kappa : f(\kappa) + g(\kappa) \leq \epsilon'_\alpha\}. \tag{18}$$

By construction, this approximate credible region is conservative, which is to say that $\{C_\alpha \subset C'_\alpha\}$. Importantly, this means that if a κ map does not belong to C'_α then it necessarily cannot belong to C_α. The approximate level-set threshold ϵ'_α at a confidence of $100(1 - \alpha)$ per cent is given by

$$\epsilon'_\alpha = f(\kappa^{\mathrm{map}}) + g(\kappa^{\mathrm{map}}) + \tau_\alpha\sqrt{N} + N, \tag{19}$$

where we recall N is the dimension of κ. The constant $\tau_\alpha = \sqrt{16\log(3/\alpha)}$ quantifies the envelope required such that the HPD credible region is a subset of the approximate HPD credible region. There exists an upper bound on the error introduced through this approximation, which is given by

$$0 \leq \epsilon'_\alpha - \epsilon_\alpha \leq \eta_\alpha\sqrt{N} + N, \tag{20}$$

where the factor $\eta_\alpha = \sqrt{16\log(3/\alpha)} + \sqrt{1/\alpha}$. This approximation error scales at most linearly with N. As will be shown in this paper, this upper bound is typically extremely conservative in practice, and the error is small.

We now introduce a recently proposed strategy for uncertainty quantification, building on the concept of approximate HPD credible regions. For further details on the strategy, we recommend the reader to see related work (Cai et al. 2017b).
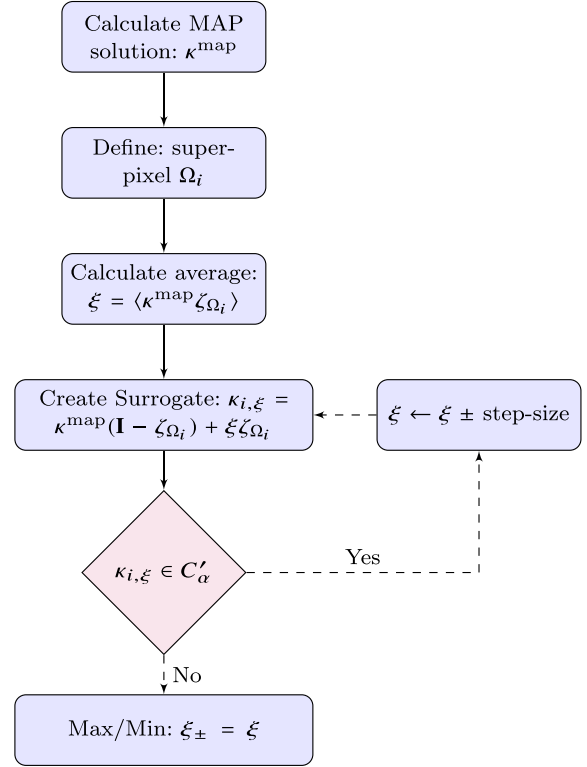
3.2 Local credible intervals

Local credible intervals can be interpreted as error bars on individual pixels or superpixel regions (collection of pixels) of a reconstructed κ map. This concept can be applied to any method for which the HPD credible region (and thus the approximate HPD credible region) can be computed. Mathematically, local credible intervals can be computed as follows (Cai et al. 2017b).

Select a partition of the κ domain $\Omega = \cup_i\Omega_i$ such that superpixels Ω_i (e.g. an 8×8 block of pixels) are independent subsets of the κ domain $\Omega_i \cap \Omega_j = \varnothing$, $\forall \{i \neq j\}$. Clearly, provided the superpixels Ω_i completely tessellate Ω they can be of arbitrary dimension. We define indexing notation on the superpixels Ω_i via the index operator ζ_{Ω_i}, which satisfy analogous relations to the standard set indicator function given in equation (A5), i.e. $\zeta_{\Omega_i} = 1$ if the pixel of the convergence map κ belongs to Ω_i and 0 otherwise.

For a given superpixel region Ω_i, we quantify the uncertainty by finding the upper and lower bounds ξ_{+,Ω_i} and ξ_{-,Ω_i}, respectively, which raise the objective function above the approximate level-set threshold ϵ'_α (or colloquially, 'saturate the HPD credible region C'_α'). In a mathematical sense, these bounds are defined by

$$\xi_{+,\Omega_i} = \max_\xi \left\{\xi | f(\kappa_{i,\xi}) + g(\kappa_{i,\xi}) \leq \epsilon'_\alpha, \ \forall \xi \in \mathbb{R}\right\} \tag{21}$$

Figure 1. Schematic of the process to construct local credible intervals. At each iterative step, the superpixel region is uniformly increased (decreased) by a step size. Once the level-set threshold ϵ'_α is saturated, the iteration is terminated. Note that this diagram does not represent the bisection method that is adopted in this article, which is a little more involved, but just a simple iterative scheme for conceptualization.

and

$$\xi_{-,\Omega_i} = \min_\xi \left\{\xi | f(\kappa_{i,\xi}) + g(\kappa_{i,\xi}) \leq \epsilon'_\alpha, \ \forall \xi \in \mathbb{R}\right\}, \tag{22}$$

where $\kappa_{i,\xi} = \kappa^{\mathrm{map}}(\mathbf{I} - \zeta_{\Omega_i}) + \xi\zeta_{\Omega_i}$ is a surrogate solution where the superpixel region has been replaced by a uniform intensity ξ. We then construct the difference image $\sum_i(\xi_{+,\Omega_i} - \xi_{-,\Omega_i})$, which represents the length of the local credible intervals (cf. error bars) on given superpixel regions at a confidence of $100(1 - \alpha)$ per cent.

In this paper, we locate $\xi_\pm$ iteratively *via* bisection, though faster converging algorithms could be used to further increase computational efficiency. A schematic diagram for constructing local credible intervals is found in Fig. 1. Conceptually, this is finding the maximum and minimum constant values that a superpixel region could take, at $100(1 - \alpha)$ per cent confidence – which is effectively Bayesian error bars on the convergence map.

In plain English, starting from the MAP convergence solution κ^{map} – at which all pixels are in positions that minimize the objective function – we then select a subset of the pixels (e.g. an 8×8 block of pixels). We start by averaging the pixels in the block that is selected. We then set the pixels within this block to the average value. Following this, we iteratively raise/lower the now uniform value of the pixels within this block whilst keeping the rest of the image fixed. After each iteration, we check if the surrogate solution (κ^{map} with the block of interest replaced by some constant value) is an acceptable solution (i.e. the objective function is below the threshold ϵ'_α). We find the values (upper and lower bounds) at which objective function is equal to the threshold ϵ'_α. We then take

Bolshoi-8 κ Bolshoi-7 κ
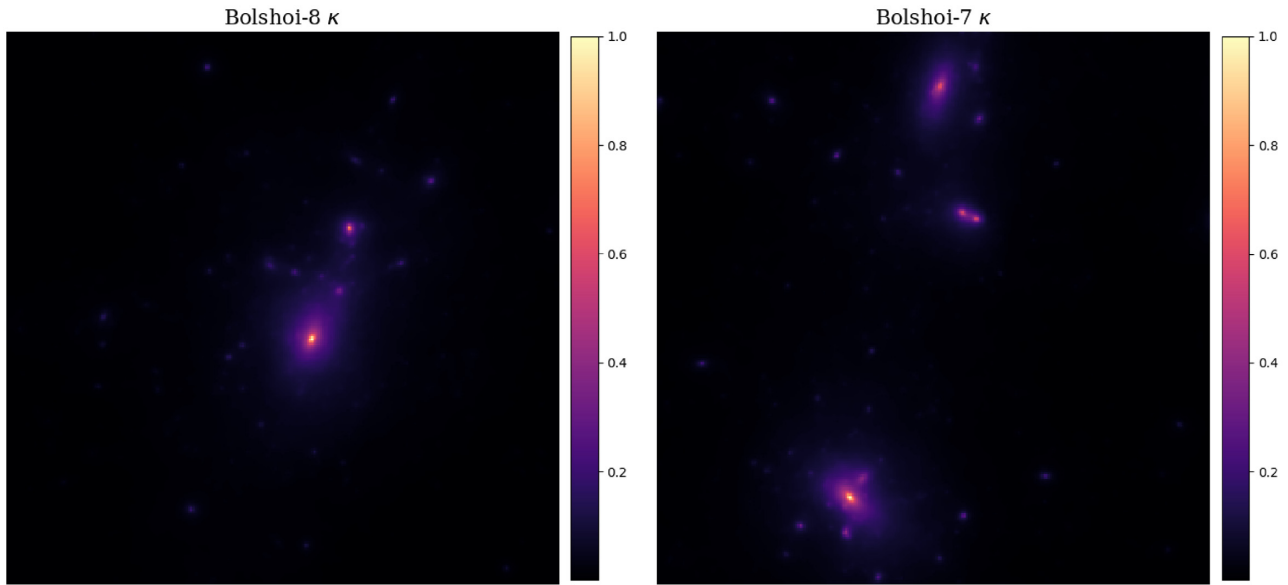
Figure 2. Two of the largest clusters extracted from the Bolshoi simulation data base, labelled as Bolshoi 7 and 8 somewhat arbitrarily. In both cases, at least one massive subhalo is located within the FoF (friends of friends) subcatalogue, as can be clearly seen.

the difference between these bounds, which is the *local credible interval* for a given 'block of interest' (superpixel region).

4 EVALUATION ON SIMULATIONS

For computing Bayesian inference problems, one would ideally adopt an MCMC approach as they are (assuming convergence) guaranteed to produce optimal results; however, these approaches are computationally demanding and can often be computationally infeasible. Therefore, it is beneficial to adopt approximate but significantly computationally cheaper methods, such as the MAP estimation approach reviewed in this article – first presented in Price et al. (2018).

However, the approximation error introduced through these approximate methods must be ascertained. Therefore, in this section, we benchmark the uncertainties reconstructed *via* our MAP algorithm to those recovered by the state-of-the-art proximal MCMC algorithm, Px-MALA (Durmus et al. 2016; Pereyra 2016). Additionally, we compare the computational efficiencies of both approaches, highlighting the computational advantages provided by approximate methods.

For simplicity and brevity, throughout we will refer to any uncertainties recovered *via* our aforementioned MAP reconstruction method as 'MAP uncertainties'. Additionally, we will refer to the MAP reconstruction method discussed throughout this paper as 'MAP algorithm'.

4.1 Data sets

We select four test convergence fields: two large-scale Buzzard *N*-body simulation (DeRose, Wechsler & Rykoff, in preparation; Wechsler, in preparation) planar patches selected at random, and two of the largest dark matter haloes from the Bolshoi *N*-body simulation (Klypin, Trujillo-Gomez & Primack 2011). This selection is chosen such as to provide illustrative examples of the uncertainty quantification techniques in both cluster and wider field weak lensing settings.

4.1.1 Bolshoi N-body

The Bolshoi cluster convergence maps used were produced from two of the largest haloes in the Bolshoi *N*-body simulation. These clusters were selected for their large total mass and the complexity of their substructure, as can be seen in Fig. 2.

Raw particle data were extracted from the Bolshoi simulation using CosmoSim,[1] and was then gridded into 1024×1024 images. These images inherently contain shot noise and so were passed through a multiscale Poisson denoising algorithm before being re-gridded to 256×256.

The denoising algorithm consisted of a forward Anscombe transform (to Gaussianize the noise), several TV-norm (total-variation) denoising optimizations of different scales, before finally inverse Anscombe transforming. Finally, the images were re-scaled on to [0, 1] – a similar denoising approach for Bolshoi *N*-body simulations was adopted in related articles (Lanusse et al. 2016).
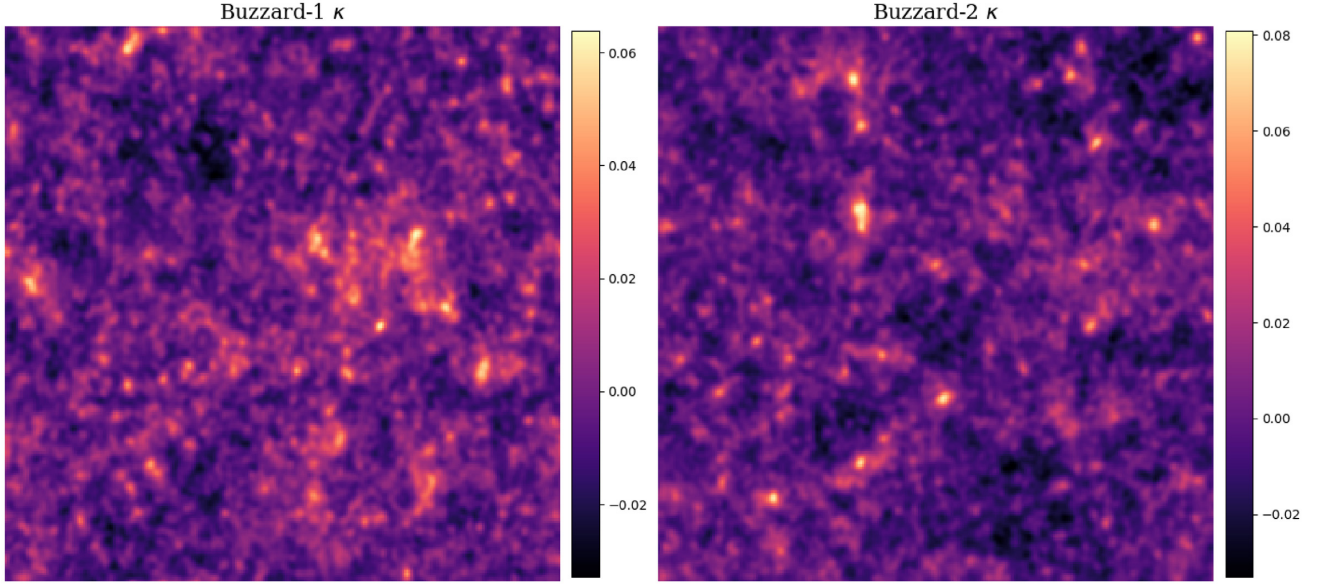
4.1.2 Buzzard N-body

The Buzzard v-1.6 shear catalogues are extracted by ray tracing from a full end-to-end *N*-body simulation. The origin for tracing is positioned in the corner of the simulation box and so the catalogue has 25 per cent sky coverage. Access to the Buzzard simulation catalogues was provided by the LSST-DESC collaboration.[2]

In the context of this paper, we restrict ourselves to working on the plane, and as such we extracted smaller planar patches. To do so, we first project the shear catalogue into a coarse HEALPix[3] (Gorski et al. 2005) gridding (with N_{side} of 16). Inside each HEALPix pixel, we tessellate the largest possible square region, on to which we rotate and project the shear catalogue. Here HEALPix pixelization is solely used for its equal area pixel properties.

---

[1] https://www.cosmosim.org
[2] http://lsst-desc.org
[3] http://healpix.sourceforge.net/documentation.php

Figure 3. Two ~1.2 deg^2 planar random extractions from the Buzzard V-1.6 N-body simulation catalogue, each containing $\mathcal{O}(10^6)$ galaxies.

After following the above procedure, the Buzzard v-1.6 shear catalogue reduces to ~3×10^3 planar patches of angular size ~1.2 deg^2, with ~4×10^6 galaxies per patch. In previous work (Price et al. 2018), we utilized 60 of these realizations, but for the purpose of this paper we select at random two planar regions to study, which we grid at a 256×256 resolution. These plots can be seen in Fig. 3.

4.2 Methodology

To draw comparisons between our MAP uncertainties and those recovered *via* Px-MALA, we conduct the following set of tests on the aforementioned data sets (see Section 4.1).

Initially, we transform the ground truth convergence κ^{in} into a clean shear field γ^{in} by

$$\gamma^{\text{in}} = \mathbf{\Phi}\kappa^{\text{in}}. \tag{23}$$

This clean set of shear measurements is then contaminated with a noise term n to produce mock noisy observations γ such that

$$\gamma = \gamma^{\text{in}} + n. \tag{24}$$

For simplicity, we choose the noise to be zero mean i.i.d. Gaussian noise of variance σ_n^2, i.e. $n \sim \mathcal{N}(0, \sigma_n^2)$. In this setting, σ_n is calculated such that the signal-to-noise ratio (SNR) is 20 dB (decibels) where

$$\sigma_n = \sqrt{\frac{\|\mathbf{\Phi}\kappa\|_2^2}{N}} \times 10^{-\frac{\text{SNR}}{20}}. \tag{25}$$

Throughout this uncertainty benchmarking, we use a fiducial noise level of 20 dB. For further details on how a noise level in dB maps to quantities such as galaxy number density and pixel size, see Price et al. (2018). In particular, we draw the reader's attention to Price et al. (table C1 2018). The noise level of 20 dB considered here is somewhat optimistic (corresponding to between 30 and 100 galaxies per square arcmin for a band limit of ~400), which is appropriate for the purposes of benchmarking against MCMC simulations, which is the focus of the current article (less optimistic simulations would simply increase the absolute level of the quantified uncertainties but not their relative level).

We then apply our entire reconstruction pipeline (Price et al. 2018), as briefly outlined in Section 2.1, to recover κ^{map}, along with the objective function – with regularization parameter μ and noise variance σ_n^2. Using these quantities, and the Bayesian framework outlined in Sections 2 and 3, we conduct uncertainty quantifications on κ^{map}.

To benchmark the MAP reconstructed uncertainties, we first construct an array of *local credible interval* maps described in Section 3 for superpixel regions of sizes [4, 8, 16] at 99 per cent confidence. These local credible interval maps are then compared to those recovered from the full MCMC analysis of the posterior.

We adopt two basic statistical measures to compare each set of recovered local credible interval maps: the Pearson correlation coefficient r and the recovered SNR. The Pearson correlation coefficient between our MAP local credible interval map $\xi^{\text{map}} \in \mathbb{R}^{N'}$ and the Px-MALA local credible interval map $\xi^{\text{px}} \in \mathbb{R}^{N'}$, where N' is the dimension of the superpixel space, is defined to be

$$r = \frac{\sum_{i=1}^{N'}(\xi^{\text{map}}(i) - \bar{\xi}^{\text{map}})(\xi^{\text{px}}(i) - \bar{\xi}^{\text{px}})}{\sqrt{\sum_{i=1}^{N'}(\xi^{\text{map}}(i) - \bar{\xi}^{\text{map}})^2}\sqrt{\sum_{i=1}^{N'}(\xi^{\text{px}}(i) - \bar{\xi}^{\text{px}})^2}}, \tag{26}$$

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

The second of our two statistics is the recovered SNR, which is calculated between ξ^{map} and ξ^{px} to be

$$\text{SNR} = 20 \times \log_{10}\left(\frac{\|\xi^{\text{px}}\|_2}{\|\xi^{\text{px}} - \xi^{\text{map}}\|_2}\right), \tag{27}$$

where ξ^{px} recovered by Px-MALA is assumed to represent the ground truth Bayesian local credible interval, and $\|.\|_2$ is the ℓ_2-norm. The SNR is a measure of the absolute similarity of two maps – in this context, rather than the structural correlation that is encoded into r, the SNR is a proxy measure of the relative magnitudes of the two data sets. Additionally, we compute the root mean squared per cent error (RMSE),

$$\text{RMSE} = 100 \times \left(\frac{\|\xi^{\text{px}} - \xi^{\text{map}}\|_2}{\|\xi^{\text{px}}\|_2}\right) \text{ per cent.} \tag{28}$$
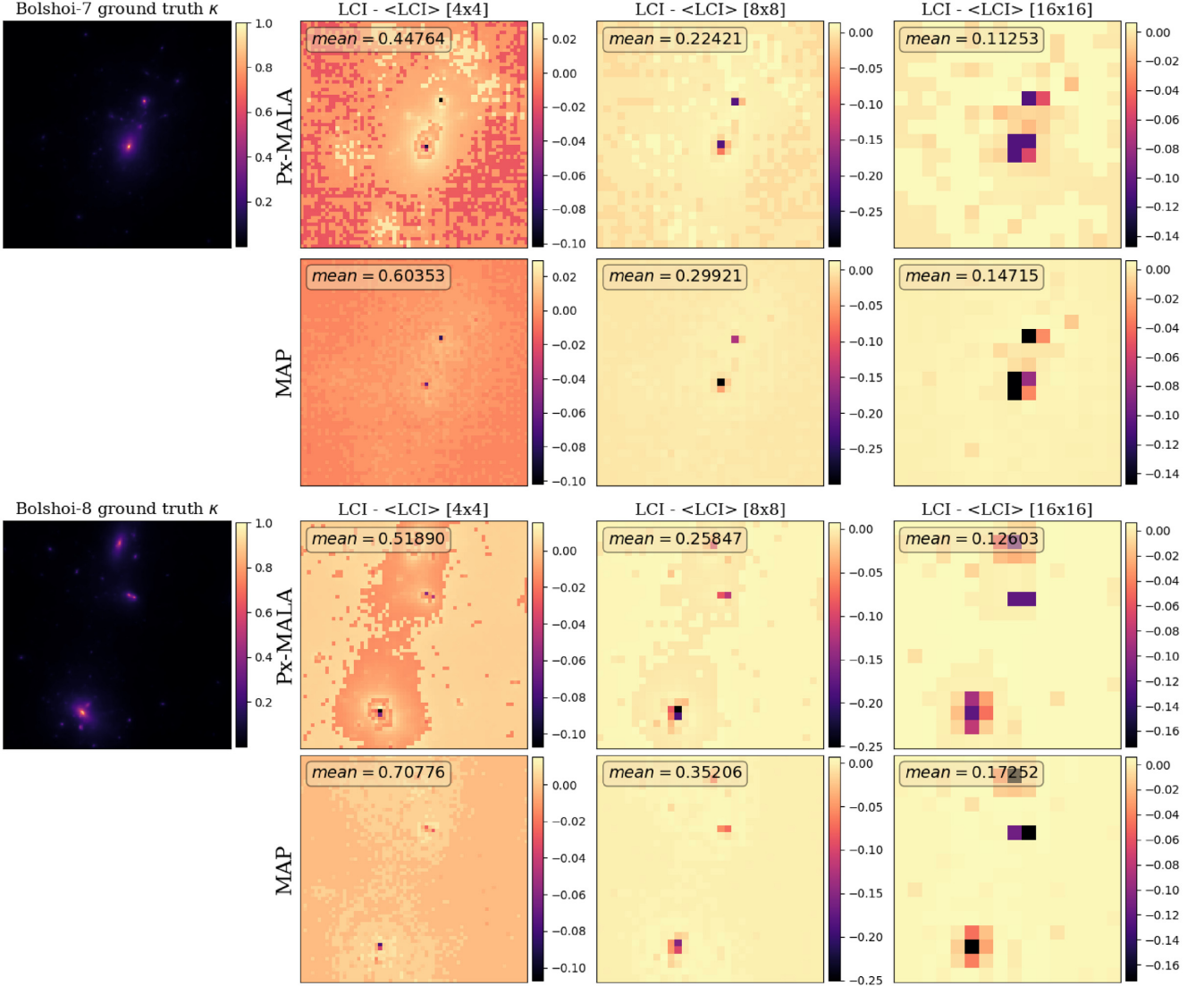
Figure 4. Local credible intervals (*cf. Bayesian error bars*) at 99 per cent confidence for the Bolshoi-7 (*top*) and Bolshoi-8 (*bottom*) cluster sparse reconstruction in both the Px-MALA setting (*top*) and MAP (*bottom*) for superpixel regions of dimension (4×4), (8×8), and (16×16) – left to right, respectively. Note that these plots display the variation about the mean of each set of LCIs, with the mean being given numerically in the subfigure legends – this is done to best display the topological similarity whilst also conveying the absolute difference in size between the methods. Notice that the mean of the MAP LCIs is in all cases larger than that of the corresponding Px-MALA LCIs. Further note that the smaller the dimension of the superpixel the larger the local credible interval, which is because adjusting fewer pixels raises the objective function by less, and so the smaller superpixels can be raised/lowered by more before saturating the level-set threshold. All numerical results are displayed in Table 1.

Conceptually, the SNR roughly compares the absolute magnitudes of recovered local credible intervals, and the Pearson correlation coefficient gives a rough measure of how geometrically similar the local credible intervals are. In this sense, the closer r is to 1 the more similar the recovered local credible intervals are, and the higher the SNR the smaller the approximation error given by equation (20). Thus, a positive result is quantified by both large correlation and large SNR.

4.3 Results

As can be seen in Figs 4 and 5, the local credible intervals recovered through our sparse hierarchical Bayesian formalism are at all times larger than those recovered via Px-MALA – confirming that the uncertainties are conservative, as proposed in Section 3.

Moreover, a strong correlation between the reconstructions can be seen.

The largest correlation coefficients r are observed for superpixel regions of dimension 16×16 in all cases ($\langle r \rangle \approx 0.9$), peaking as high as 0.98 for the Buzzard 1 extraction – which constitutes a near maximal correlation, and thus an outstanding topological match between the two recovered local credible intervals.

Additionally, in the majority of cases, the recovered SNR is ≥ 10 dB – in some situations rising as high as ≈ 13 dB (corresponding to ≈ 20 per cent RMSE per cent error) – which indicates that the recovered MAP uncertainties are close in magnitude to those recovered *via* Px-MALA.

However, for superpixels with dimension 4×4, the structural correlation between ξ^{map} and ξ^{px} becomes small – in one case becoming marginally negatively correlated. This is likely to be a
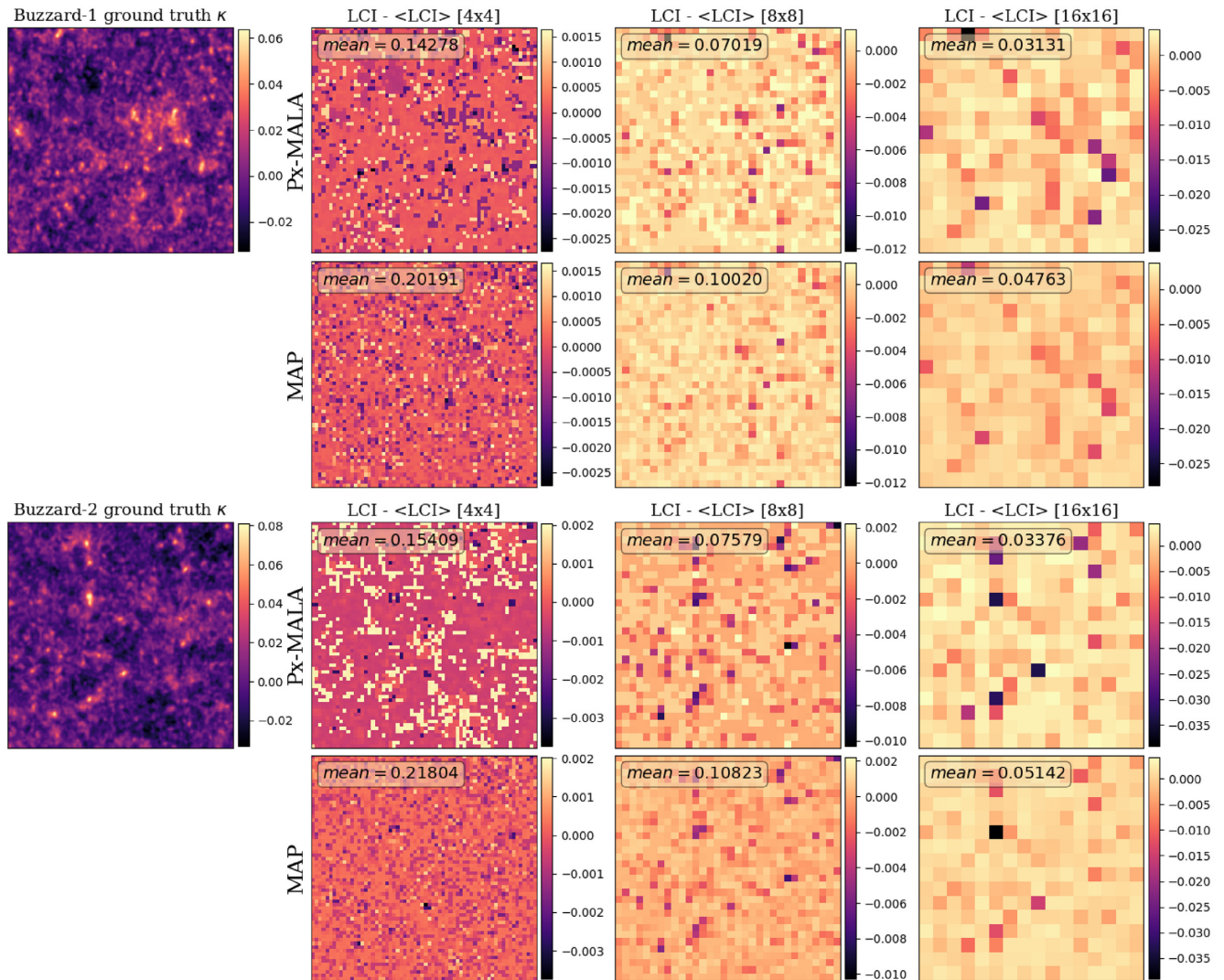
Figure 5. Local credible intervals (*cf. Bayesian error bars*) at 99 per cent confidence for the Buzzard-1 (*top*) and Buzzard-2 (*bottom*) cluster sparse reconstruction in both the Px-MALA setting (*top*) and MAP (*bottom*) for superpixel regions of dimension (4×4), (8×8), and (16×16) – left to right, respectively. Note that these plots display the variation about the mean of each set of LCIs, with the mean being given numerically in the subfigure legends – this is done to best display the topological similarity whilst also conveying the absolute difference in size between the methods. Notice that the mean of the MAP LCIs is in all cases larger than that of the corresponding Px-MALA LCIs. Further note that the smaller the dimension of the superpixel the larger the local credible interval, which is because adjusting fewer pixels raises the objective function by less, and so the smaller superpixels can be raised/lowered by more before saturating the level-set threshold. All numerical results are displayed in Table 1.

direct result of the error given by equation (20) inherited from the definition of the approximate HPD credible region – as this approximation has the side effect of smoothing the posterior hyper-volume, and for small superpixels the hyper-volume is typically not smooth, thus the correlation coefficient r decreases.

We conducted additional tests for large 32×32 dimension superpixels, which revealed a second feature of note. For particularly large superpixel regions (32×32 or larger), the SNR becomes small for both Buzzard maps. This is a result of the assumption that within a superpixel there exists a stable mean that is roughly uniform across the superpixel. Clearly, for Buzzard-type data, on large scales this breaks down and so the recovered local credible intervals deviate from those recovered *via* Px-MALA. It is important to stress that this is a breakdown of the assumptions made when constructing local credible intervals and not an error of the approximate HPD credible region.

The numerical results are summarized in Table 1. Typically, structures of interest in recovered convergence maps cover superpixel regions of roughly 8×8 to 16×16, and so for most realistic applications our MAP uncertainties match very well with those recovered through Px-MALA. In most situations, weak lensing data are gridded such that it best represents the features of interest, and so structures of interest (by construction) typically fall within 8×8 to 16×16 dimension superpixel regions for 256×256 gridded images – for higher resolution images the structures of interest, and corresponding optimal superpixels will follow a similar ratio.

Overall, we find a very close relation between the local credible intervals recovered through our MAP algorithm with those recovered *via* Px-MALA – a state-of-the-art MCMC algorithm. We find that MAP and Px-MALA local credible intervals are typically strongly topologically correlated (Pearson correlation coefficient ≈ 0.9) in addition to being physically tight (RMSE error

Table 1. Comparisons between the local credible interval maps recovered *via* MAP and those recovered *via* Px-MALA. Note that higher superpixels correspond to coarser resolutions whereas smaller superpixels lead to higher resolution. This is because the superpixel size is the size of the groups of pixels used to tile the original image – therefore, larger tiling components lead to fewer tiles, and therefore lower resolution.

Super Pixel	Pearson Correlation	SNR (dB)	RMSE Error
Bolshoi-7			
4×4	0.463	11.737	25.892 per cent
8×8	0.848	11.994	25.137 per cent
16×16	0.945	12.509	23.690 per cent
Bolshoi-8			
4×4	− 0.168	11.467	26.710 per cent
8×8	0.929	11.490	26.637 per cent
16×16	0.941	11.350	27.070 per cent
Buzzard-1			
4×4	0.164	10.666	29.289 per cent
8×8	0.916	10.473	29.948 per cent
16×16	0.984	9.262	34.427 per cent
Buzzard-2			
4×4	0.140	10.653	29.333 per cent
8×8	0.904	10.465	29.973 per cent
16×16	0.926	9.217	34.605 per cent

Table 2. Numerical comparison of computational time of Px-MALA and MAP. The MAP approach typically takes $\mathcal{O}(10^{-1})$ s, compared to Px-MALA's $\mathcal{O}(10^5)$ s. Therefore, for linear reconstructions, MAP is close to $\mathcal{O}(10^6)$ times faster.

Px-MALA Time (s)	MAP Time (s)	Ratio
Buzzard-1		
133 761	0.182	0.734×10^6
Buzzard-2		
141 857	0.175	0.811×10^6
Bolshoi-7		
95 339	0.153	0.623×10^6
Bolshoi-8		
92 929	0.143	0.650×10^6

of ≈20–30 per cent). Moreover, we find that the MAP local credible intervals are, everywhere, larger than the Px-MALA local credible intervals, corroborating the assertion that the approximate HPD level-set threshold ϵ'_α is in fact conservative.

We now compare the computational efficiency of our sparse Bayesian reconstruction algorithm against Px-MALA. It is worth noting that all Px-MALA computation was done on a high-performance workstation (with 24 CPU cores and 256 GB of memory), whereas all MAP reconstructions were done on a standard 2016 MacBook Air. The computation time for MAP estimation is found to be $\mathcal{O}(\text{seconds})$ whereas the computation time for Px-MALA is found to be $\mathcal{O}(\text{days})$. Specifically, we find that the MAP reconstruction algorithm is of $\mathcal{O}(10^6)$ (typically $\geq 8 \times 10^5$) times faster than the state-of-the-art Px-MALA MCMC algorithm, see Table 2 for numerical results. Moreover, the MAP reconstruction algorithm supports algorithmic structures that can be highly parallelized and distributed.

5 CONCLUSIONS

In this article, we introduce the concept of local credible intervals (cf. pixel-level error bars) – developed in previous work and applied in the radio-interferometric setting – to the weak lensing setting as an additional form of uncertainty quantification. Utilizing local credible intervals, we validate the sparse hierarchical Bayesian mass-mapping formalism presented in previous work (Price et al. 2018). Specifically, we compare the local credible intervals recovered *via* the MAP formalism and those recovered *via* a complete MCMC analysis – from which the true posterior is effectively recovered.

To compute the asymptotically exact posterior, we utilize Px-MALA – a state-of-the-art proximal MCMC algorithm. Using the local credible intervals, we benchmark the MAP uncertainty reconstructions against Px-MALA.

Quantitatively, we compute the Pearson correlation coefficient (r, as a measure of the correlation between hyper-volume topologies), and recovered SNR and the root mean squared percentage error (SNR and RMSE, both as measures of how tightly constrained is the absolute error).

We find that for a range of superpixel dimensions the MAP and Px-MALA uncertainties are strongly topologically correlated ($r \geq 0.9$). Moreover, we find the RMSE to typically be ∼20–30 per cent, which is tightly constrained when one considers this is a conservative approximation along each of at least $\mathcal{O}(10^3)$ dimensions.

Additionally, we compare the computational efficiency of Px-MALA and our MAP approach. In a 256×256 setting, the computation time of the MAP approach was $\mathcal{O}(\text{seconds})$ whereas the computation time for Px-MALA was $\mathcal{O}(\text{days})$. Overall, the MAP approach is shown to be $\mathcal{O}(10^6)$ times faster than the state-of-the-art Px-MALA algorithm.

A natural progression is to extend the planar sparse Bayesian algorithm to the sphere, which will be the aim of upcoming work – a necessity when dealing with wide-field stage IV surveys such as LSST[4] and *Euclid*.[5] Additionally, we will expand the set of uncertainty quantification techniques to help propagate principled Bayesian uncertainties into the set of higher order statistics typically computed on the convergence field.

[4]https://www.lsst.org
[5]http://euclid-ec.org

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

Cai X., Pereyra M., McEwen J. D., 2017a, MNRAS, 480, 4154

Cai X., Pereyra M., McEwen J. D., 2017b, MNRAS, 480, 4170

Candès E. J., Wakin M. B., 2008, IEEE Signal Process. Mag., 25, 21

Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223

Chang C. et al., 2018, MNRAS, 475, 3165

Coles P., Chiang L.-Y., 2000, Nature, 406, 376

Combettes P. L., Pesquet J.-C., 2011, Fixed-point algorithms for inverse problems in sciences and engineering. Springer, New York, NY, USA, p. 185

Donoho D. L., 2006, IEEE Trans. Inf. Theory, 52, 1289

Durmus A., Moulines E., Pereyra M., 2016, preprint (arXiv:1612.07471)

Gorski K. M., Hivon E., Banday A., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759

Heavens A., 2009, Nucl. Phys. B, 194, 76

Horowitz B., Seljak U., Aslanyan G., 2019, J. Cosmol. Astropart. Phys., 2019, 035

Jeffrey N. et al., 2018, MNRAS, 479, 2871

Kaiser N., Squires G., 1993, ApJ, 404, 441

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

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

Parikh N., Boyd S., 2014, Found. Trends Optim., 1, 123

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, Stat. Comput., 26, 745

Pereyra M., 2017, SIAM J. Imaging Sci., 10, 285

Pereyra M., Bioucas-Dias J., Figueiredo M., 2015, Maximum-a-posteriori Estimation with Unknown Regularisation Parameters. p. 230

Pires S., Starck J.-L., Amara A., 2009, MNRAS, 395, 1265

Price M. A., Cai X., McEwen J. D., Kitching T. D., Wallis C. G. R., 2018, MNRAS, submitted

Repetti A., Pereyra M., Wiaux Y., 2019, SIAM Journa on Imaging Sciences, 12, 87

Robert C.-P., 2001, The Bayesian Choice. Springer, New York, NY, USA

Roberts G. O., Tweedie R. L., 1996, Bernoulli, 2, 341

Starck J.-L., Murtagh F., Fadili J., 2015, Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis. Cambridge University Press, New York, NY, USA

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: REGULARIZATION MARGINALIZATION

A prior $f(\kappa)$ is k-homogeneous if $\exists\, k \in \mathbb{R}_+$ such that

$$f(\eta\kappa) = \eta^k f(\kappa), \ \forall \kappa \in \mathbb{R}^n, \ \forall \eta > 0. \tag{A1}$$

As all norms, composite norms and compositions of norms and linear operators (Pereyra et al. 2015), have homogeneity of 1, k in our setting is set to 1. If we wish to infer κ without a priori knowledge of μ (the regularization parameter) then we calculate the normalization factor of $p(\kappa|\mu)$,

$$C(\mu) = \int_{\mathbb{C}^N} \exp\{-\mu f(\kappa)\}\mathrm{d}\kappa. \tag{A2}$$

For the vast majority of cases of interest, calculating $C(\mu)$ is not feasible, due to the large dimensionality of the integral. However, it was recently shown (Pereyra et al. 2015) that if the prior term $f(\kappa)$ is k-homogeneous then

$$C(\mu) = D\mu^{-N/k}, \quad \text{where} \quad D \equiv C(1). \tag{A3}$$

A gamma-type hyper-prior is then selected (a typical choice for scale parameters) on μ such that

$$p(\mu) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} \mathrm{e}^{-\beta\mu} \mathbb{I}_{\mathbb{R}_+}(\mu), \tag{A4}$$

where the hyper-parameters (α, β) are very weakly dependent and can be set to 1 (as in Pereyra et al. 2015), and $\mathbb{I}_{C_\alpha}$ is an indicator function defined by

$$\mathbb{I}_{C_\alpha} = \begin{cases} 1 & \text{if,} \quad \kappa \in C_\alpha \\ 0 & \text{if,} \quad \kappa \notin C_\alpha. \end{cases} \tag{A5}$$

Now construct a joint Bayesian inference problem of $p(\kappa, \mu|\gamma)$ with MAP estimator $(\kappa^{\mathrm{map}}, \mu^{\mathrm{map}}) \in \mathbb{C}^N \times \mathbb{R}_+$. By definition, at this MAP estimator

$$\mathbf{0}_{N+1} \in \partial_{\kappa,\mu} \log p(\kappa^{\mathrm{map}}, \mu^{\mathrm{map}}|\gamma), \tag{A6}$$

where $\mathbf{0}_i$ is the i-dimensional null vector. This in turn implies both that

$$\mathbf{0}_N \in \partial_\kappa \log p(\kappa^{\mathrm{map}}, \mu^{\mathrm{map}}|\gamma), \tag{A7}$$

from which equation (9) follows naturally, and

$$\mathbf{0} \in \partial_\mu \log p(\kappa^{\mathrm{map}}, \mu^{\mathrm{map}}|\gamma). \tag{A8}$$

Using equations (A3), (A4), and (A8), it can be shown (Pereyra et al. 2015) that

$$\mu^{\mathrm{map}} = \frac{\frac{N}{k} + \alpha - 1}{f(\kappa^{\mathrm{map}}) + \beta}. \tag{A9}$$

Hereafter, we drop the map superscript on μ^{map} for simplicity. In order to compute the MAP μ preliminary iterations are performed as follows:

$$\kappa^{(t)} = \underset{\kappa}{\mathrm{argmin}} \left\{ f(\kappa; \mu^{(t)}) + g(\kappa) \right\}, \tag{A10}$$

$$\mu^{(t+1)} = \frac{\frac{N}{k} + \alpha - 1}{f(\kappa^{(t)}) + \beta}, \tag{A11}$$

where α and β are (weakly dependent) hyper-parameters from a gamma-type hyper-prior, N is the dimension of the reconstructed space, and the sufficient statistic $f(\kappa)$ is k-homogeneous. Typically, the MAP solution of μ converges within $\sim$5–10 iterations, after which μ is fixed and the optimization in equation (9) is computed.

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