MNRAS **000**, 1–16 (2017) Preprint 21 November 2018 Compiled using MNRAS LATEX style file v3.0

Uncertainty quantification for radio interferometric imaging: I. proximal MCMC methods

Xiaohao Cai[1⋆], Marcelo Pereyra[2⋆] and Jason D. McEwen[1⋆]
[1]*Mullard Space Science Laboratory, University College London (UCL), Surrey RH5 6NT, United Kingdom*
[2]*Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom*

Accepted —. Received —; in original form —

ABSTRACT

Uncertainty quantification is a critical missing component in radio interferometric imaging that will only become increasingly important as the big-data era of radio interferometry emerges. Since radio interferometric imaging requires solving a high-dimensional, ill-posed inverse problem, uncertainty quantification is difficult but also critical to the accurate scientific interpretation of radio observations. Statistical sampling approaches to perform Bayesian inference, like Markov Chain Monte Carlo (MCMC) sampling, can in principle recover the full posterior distribution of the image, from which uncertainties can then be quantified. However, traditional high-dimensional sampling methods are generally limited to smooth (*e.g.* Gaussian) priors and cannot be used with sparsity-promoting priors. Sparse priors, motivated by the theory of compressive sensing, have been shown to be highly effective for radio interferometric imaging. In this article proximal MCMC methods are developed for radio interferometric imaging, leveraging proximal calculus to support non-differential priors, such as sparse priors, in a Bayesian framework. Furthermore, three strategies to quantify uncertainties using the recovered posterior distribution are developed: (i) local (pixel-wise) credible intervals to provide error bars for each individual pixel; (ii) highest posterior density credible regions; and (iii) hypothesis testing of image structure. These forms of uncertainty quantification provide rich information for analysing radio interferometric observations in a statistically robust manner.

Key words: techniques: image processing – techniques: interferometric – methods: data analysis – methods: numerical – methods: statistical.

1 INTRODUCTION

Radio interferometric (RI) telescopes provide a wealth of valuable information for astrophysics and cosmology (Ryle & Vonberg 1946; Ryle & Hewish 1960; Thompson et al. 2017) since they allow observation of the radio emission of the sky with high angular resolution and sensitivity. The measured visibilities acquired by the telescope relate to Fourier measurements of the sky image of interest (the Fourier model may be modified to account for, *e.g.*, wide fields of view, co-planer baselines, and other directional dependent effects). Imaging observations made by radio telescopes requires solving an ill-posed linear inverse problem (Thompson et al. 2017), which is an important first step in many subsequent scientific analyses. Since the inverse problem is ill-posed (sometimes seriously), uncertainty information regarding reconstructed images (*e.g.* error estimates) is critical. Nevertheless, uncertainty information is currently lacking in all RI imaging techniques used in practice.

Classical imaging techniques were developed in the field to solve the RI reconstruction problem, such as CLEAN and its multi-scale variants (Högbom 1974; Bhatnagar & Corwnell 2004; Cornwell 2008; Stewart et al. 2011). In particular, CLEAN builds a model image by iteratively removing point source components from the residuals of the acquired data (at each iteration). CLEAN-based algorithms, however, are typically slow (generally requiring computationally demanding major cycles; *cf.* Clark CLEAN), requiring fine-tuning and supervision, while providing suboptimal imaging quality (see, *e.g.*, Li et al. 2011a; Carrillo et al. 2012). Another classical technique is the maximum entropy method (MEM) (Ables 1974; Gull & Daniell 1978), extended to RI imaging by Cornwell & Evans (1985). The MEM approach of Cornwell & Evans (1985) developed for RI imaging considers a regularisation problem consisting of a relative entropic prior, a (Gaussian) likelihood term and an additional flux constraint. In principle, MEM requires less fine-tuning and supervision compared to CLEAN and can therefore alleviate part of the shortcomings of CLEAN-based algorithms. However, an optimal metric – expressed as an entropy functional – is not known in advance and therefore needs to be cho-

⋆ E-mail: x.cai@ucl.ac.uk (XC); m.pereyra@hw.ac.uk (MP); jason.mcewen@ucl.ac.uk (JDM)

© 2017 The Authors

sen individually (Starck et al. 2001; Maisinger et al. 2004). Indeed, it is widely known that MEM fails to reconstruct sharp and smooth image features simultaneously. Recently, the theory of compressed sensing (CS) has suggested the use of sparse representation and regularisation approaches for the recovery of sparse signals from incomplete linear measurements (Donoho 2006; Candes & Wakin 2008; Candes et al. 2010), which has shown great success. CS techniques based on sparse regularisation were ushered into RI imaging for image reconstruction (Suksmono 2009; Wiaux et al. 2009a,b; Wenger et al. 2010; McEwen & Wiaux 2011; Li et al. 2011a,b; Carrillo et al. 2012, 2014; Wolz et al. 2013; Dabbech et al. 2015; Dabbech et al. 2017; Garsden et al. 2015; Onose et al. 2016, 2017; Pratley et al. 2018; Kartik et al. 2017) and have shown promising results and improvements compared to traditional approaches such as CLEAN-based methods and MEM. In general, such approaches can recover sharp and smooth image features simultaneously (*e.g.* Carrillo et al. 2012). While sparse approaches have been shown to be highly effective, the best approach to image different sources remains an open question. Algorithms have been developed to scale sparse approaches to big-data (Carrillo et al. 2014; Onose et al. 2016, 2017; Kartik et al. 2017; Cai et al. 2017a), such as that anticipated from the Square Kilometre Array (SKA[1]). However, CLEAN-based methods, MEM, and CS-based methods, unfortunately, do not provide any uncertainty quantification about the accuracy of recovered images.

Statistical sampling methods to perform Bayesian inference, like Markov chain Monte Carlo (MCMC) methods, which sample the full posterior distribution, have the ability to provide uncertainty information. However, this comes at a considerable computational cost. A proof of concept application of MCMC sampling to RI imaging was performed by Sutter et al. (2014), using Gibbs sampling with Gaussian process priors. Uncertainty information in the form of the posterior image variance was considered. However, an idealised telescope model was adopted and the technique has yet to be applied to real observational data. In general MCMC sampling techniques that scale to high-dimensional settings (like RI imaging), place restrictions on the priors that can be considered. Gibbs sampling, for example, requires the ability to draw from conditional distributions. Two of the most effective classes of MCMC methods for high-dimensional settings include Hamiltonian Monte Carlo (HMC) (Neal 2012) and the unadjusted Langevin algorithm (ULA) (Roberts & Tweedie 1996). When a Metropolis-Hasting (MH) accept-reject step is added to ULA, one obtains the Metropolis-adjusted Langevin algorithm (MALA) (Robert & Casella 2004). HMC, ULA and MALA exploit gradients to capture local properties of the target density in order to explore high-dimensional parameter spaces efficiently. However, a significant limitation of HMC, MALA and ULA is that the priors considered must be smooth, which prohibits their use for priors that promote sparseness. An alternative Bayesian approach to RI imaging using Information Field Theory (Enβlin et al. 2009) has been presented in the form of the RESOLVE algorithm (Junklewitz et al. 2016; Greiner et al. 2017). This approach assumes a log-normal prior and recovers a *maximum a posteriori* (MAP) estimate, proving uncertainty information in the form of an approximate posterior covariance. However, the method remains computationally demanding.

Uncertainty quantification is an important missing component in RI imaging for quantitative imaging, scientific inquiry, and decision-making. Moreover, since the RI imaging problem is often (severely) ill-posed, uncertainty quantification becomes increasingly important. No existing RI imaging techniques that are used in practice provide uncertainty quantification. Also, those approaches that do provide some form of uncertainty quantification in RI imaging cannot scale to big-data. Moreover, such approaches only support restrictive classes of priors (typically Gaussian or log-normal, which lead to poor reconstruction results relative to sparse priors). In summary, no existing approach can support the sparse priors that have been shown in practice to be highly effective for RI imaging (*e.g.* Pratley et al. 2018), while also providing uncertainty quantification, in a manner that can scale to big-data. We present new techniques that fulfil precisely these criteria.

In two companion articles, we present novel RI imaging techniques that support the sparsity-promoting priors that have been shown to be highly effective in practice, provide various forms of uncertainty quantification, and that scale to big-data. In the current article we show how to support uncertainty quantification for sparse priors via proximal MCMC methods. In the companion article (Cai et al. 2017b), we show how to scale uncertainty quantification with sparse priors to big-data.

In this article, two proximal MCMC methods, Moreau-Yosida ULA (MYULA) (Durmus et al. 2016) and proximal MALA (Px-MALA) (Pereyra 2016b), are introduced for RI imaging. These algorithms are direct extensions of ULA and MALA that exploit proximity mappings Moreau-Yosida envelopes, and Moreau approximations. Most importantly, due to the versatility of proximity mappings, these two algorithms are able to sample high-dimensional distributions with a variety of different types of priors, including the non-differentiable sparse priors that have been widely used in RI imaging but yet cannot be tackled by standard MCMC methods. Specifically, Px-MALA can sample the posterior distribution with high accuracy (formally, it is guaranteed to converge to the target distribution), but the MH accept-reject step embedded in it induces a high computation overhead. MYULA, on the other hand, eliminates the MH accept-reject step by introducing well-controlled approximations (formally, the bias introduced by such approximations can be made arbitrarily small), and thus has a lower computational overhead.

The uncertainty quantification strategy considered in this article proceeds as follows. Firstly, using Bayesian inference, two unconstrained inverse models – analysis and synthesis forms – with sparse priors are presented to address the RI imaging problem. Then, full posterior distributed samples corresponding to these two unconstrained models are generated by the sampling methods Px-MALA and MYULA. After that, three ways of quantifying uncertainty information for RI imaging are constructed, including: (i) local (pixel-wise) credible intervals (*cf.* error bars) computed from the generated posterior samples; (ii) highest posterior density (HPD) credible regions computed using the generated posterior samples; and (iii) hypothesis testing of image structure using the HPD credible regions. Moreover, comparisons between the performance of Px-MALA and MYULA, and between the analysis and synthesis models are presented.

The remainder of this article is organised as follows. In Section 2 we introduce the RI imaging problem, the Bayesian inference approach to imaging, and the regularisation approach to imaging, elaborating the relationship between various approaches and various algorithms (*e.g.* CLEAN and MEM). In Section 3 we discuss Bayesian inference for sparse priors by proximal MCMC methods and in Section 4 derive the detailed implementation of the proximal MCMC methods for RI imaging problems. Uncertainty quantification for RI imaging is formulated in Section 5. Numerical results

[1] http://www.skatelescope.org/

evaluating the performance of our uncertainty quantification methods are reported in Section 6. Finally, we conclude in Section 7 with a brief description of the main contributions, a discussion of planned extensions of this work, and elucidate connections with the companion article (Cai et al. 2017b).

2 RADIO INTERFEROMETRIC IMAGING

To start, we first recall the RI imaging problem and then review sparse representations, which are often exploited in modern approaches to solve this problem. We model the RI imaging problem from the perspective of Bayesian inference and, finally, elaborate the relationship between Bayesian inference and regularisation on which CLEAN, MEM, and CS approaches are based.

2.1 Radio interferometry

The sky intensity can be imaged by radio interferometric telescopes that measure the radio emission of the sky using an array of spatially separated antennas. When the baselines in an array are coplanar and the field of view is narrow, the visibility $\boldsymbol{y}$ can be measured by correlating the signals from pairs of antennas, separated by the baseline components $\boldsymbol{u} = (u, v)$. The general RI equation for obtaining $\boldsymbol{y}$ reads as (Thompson et al. 2017)

$$\boldsymbol{y}(\boldsymbol{u}) = \int A(\boldsymbol{l})\boldsymbol{x}(\boldsymbol{l})\mathrm{e}^{-2\pi i \boldsymbol{u}\cdot\boldsymbol{l}}\mathrm{d}^2\boldsymbol{l}, \tag{1}$$

where $\boldsymbol{x}$ represents the sky brightness distribution, described in coordinates $\boldsymbol{l} = (l, m)$ (the coordinates of the plane of the sky, centred on the pointing direction of the telescope), and $A(\boldsymbol{l})$ represents the primary beam of the telescope. While not considered further in this article, wide fields and other direction dependent effects can be incorporated (see *e.g.* McEwen & Scaife 2008; Cornwell et al. 2008; Bhatnagar et al. 2008; Wiaux et al. 2009b; McEwen & Wiaux 2011; Wolz et al. 2013; Offringa et al. 2014; Dabbech et al. 2017).

In RI imaging, the goal is to recover the sky intensity signal $\boldsymbol{x}$ from the measured visibilities $\boldsymbol{y}$ acquired according to (1). Precisely, we consider the estimation of a vector $\boldsymbol{x} \in \mathbb{R}^N$ representing a sampled image on a discrete grid of N points in real space, from a measurement vector $\boldsymbol{y} \in \mathbb{C}^M$ gathering the M visibilities observed in a complex vector space, related to $\boldsymbol{x}$ by the linear observation model

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is a linear measurement operator modelling the realistic acquisition of the sky brightness components and $\boldsymbol{n} \in \mathbb{C}^M$ is the instrumental noise. Without loss of generality, we assume independent and identically distributed (i.i.d.) Gaussian noise. The estimation of $\boldsymbol{x}$ is therefore a linear inverse problem, which is challenging because the operator $\boldsymbol{\Phi}$ is ill-posed and ill-conditioned, and because of the high dimensionality involved (Rau et al. 2009).

2.2 Sparse representation

RI imaging methods typically use prior knowledge about $\boldsymbol{x}$ to regularise the estimation problem and deliver more accurate estimation results. In particular, many new methods use the fact that natural signals and images in general, and RI images in particular, often exhibit a sparse representation in some bases (*e.g.* a point source basis or a multi-scale basis such as wavelets). Let

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{a} = \sum_i \boldsymbol{\Psi}_i a_i, \tag{3}$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{N\times L}$ is a dictionary (*e.g.* a wavelet basis or an overcomplete frame) and $\boldsymbol{a} = (a_1, \cdots, a_L)^\top$ is the vector of the synthesis coefficients of $\boldsymbol{x}$ under $\boldsymbol{\Psi}$. Then $\boldsymbol{x}$ is said to be sparse if $\boldsymbol{a}$ contains only K non-zero coefficients, *i.e.*, $\|\boldsymbol{a}\|_0 = K$ (recall $\|\boldsymbol{a}\|_0$ gives the number of non-zero components of $\boldsymbol{a}$), where $K \ll N$. Similarly, $\boldsymbol{x}$ is called compressible under $\boldsymbol{\Psi}$ if many coefficients of $\boldsymbol{a}$ are nearly zero, *i.e.*, its sorted coefficients a_i satisfy a power law decay. In practice, it is ubiquitous that natural signals and images $\boldsymbol{x}$ are sparse or compressible.

2.3 Bayesian inference

The inverse problem presented in (2) can be addressed elegantly in the Bayesian statistical inference framework, which in addition to allowing one to derive estimates of $\boldsymbol{x}$ also provides tools to analyse and quantify the uncertainty in the solutions obtained. Let $p(\boldsymbol{y}|\boldsymbol{x})$ be the likelihood function of the statistical model associated with (2). In the case of i.i.d. Gaussian noise the likelihood function reads

$$p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp(-\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2), \tag{4}$$

where σ represents the standard deviation of the noise level.

As mentioned previously, recovering $\boldsymbol{x}$ solely from $\boldsymbol{y}$ is not possible because the problem is not well posed. Bayesian methods address this difficulty by exploiting prior knowledge – represented by a prior distribution $p(\boldsymbol{x})$ – to regularise the problem, reduce uncertainty, and improve estimation results. Typically priors of the form $p(\boldsymbol{x}) \propto \exp(-\phi(\mathcal{B}\boldsymbol{x}))$ are considered, for some linear operator $\mathcal{B}$ and potential function ϕ. Various forms for ϕ can be considered, for example: Tikhonov regularisation (Golub et al. 1999; Cai et al. 2013), used to promote smoothness, corresponds to the Gaussian prior of $p(\boldsymbol{x}) \propto \exp(-\mu\|\boldsymbol{x}\|_2^2)$; the entropic prior of $p(\boldsymbol{x}) \propto \exp(-\mu\boldsymbol{x}^\dagger \log \boldsymbol{x})$ (Ables 1974; Gull & Daniell 1978; Cornwell & Evans 1985); and the ℓ_p norm with $0 \leq p \leq 1$ used as a regulariser to promote sparseness (Chen et al. 2016; Cai et al. 2015; Wiaux et al. 2009a,b; McEwen & Wiaux 2011; Donoho 2006; Candes & Wakin 2008). Here $\mu > 0$ is a regularisation parameter. We refer to such priors as *analysis* priors because they operate on the canonical coordinate system of $\boldsymbol{x}$. Alternatively, it is also possible to adopt a so-called *synthesis* approach and use (3) to express the prior knowledge for $\boldsymbol{x}$ via a prior distribution $p(\boldsymbol{a})$ on the synthesis coefficients $\boldsymbol{a}$.

In this article we consider both analysis and synthesis formulations because they are both widely used in RI imaging. For analysis models we consider Laplace-type priors of the form

$$p(\boldsymbol{x}) \propto \exp(-\mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1), \tag{5}$$

where $\boldsymbol{\Psi}^\dagger$ denotes the adjoint of $\boldsymbol{\Psi}$, $\mu > 0$ is a regularisation parameter, and $\|\cdot\|_1$ is the ℓ_1 norm; while for synthesis models we consider the Laplace prior

$$p(\boldsymbol{a}) \propto \exp(-\mu\|\boldsymbol{a}\|_1). \tag{6}$$

Observe that both formulations are equivalent when $\boldsymbol{\Psi}$ is an orthogonal basis. However, for redundant dictionaries the approaches have very different properties. Further discussions about the analysis and synthesis forms can be found, for example, in Maisinger et al. (2004), Elad et al. (2007) and Cleju et al. (2012).

Prior and observed information can then be combined by using

Bayes' theorem to obtain the posterior distribution. For analysis formulations the posterior is given by

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p_{\mathrm{a}}(\boldsymbol{y})}, \tag{7}$$

which models our knowledge about $\boldsymbol{x}$ after observing $\boldsymbol{y}$, where $p_{\mathrm{a}}(\boldsymbol{y}) = \int_{\mathbb{R}^N} p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ is the marginal likelihood (or Bayesian evidence) of the analysis model. Similarly, for synthesis models the posterior reads

$$p(\boldsymbol{a}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{a})p(\boldsymbol{a})}{p_{\mathrm{s}}(\boldsymbol{y})}, \tag{8}$$

with $p(\boldsymbol{y}|\boldsymbol{a}) = p(\boldsymbol{y}|\boldsymbol{x})$ for $\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{a}$, where $p_{\mathrm{s}}(\boldsymbol{y}) = \int_{\mathbb{R}^N} p(\boldsymbol{y}|\boldsymbol{a})p(\boldsymbol{a})\mathrm{d}\boldsymbol{a}$ is the model's marginal likelihood.

Note that the denominators $p_{\mathrm{a}}(\boldsymbol{y})$ in (7) and $p_{\mathrm{s}}(\boldsymbol{y})$ in (8), *i.e.* the marginal likelihoods, are unrelated to $\boldsymbol{x}$ and $\boldsymbol{a}$, respectively, and therefore constants with respect to (w.r.t.) parameter inference. It follows that the unnormalised posterior distributions for the analysis and synthesis formulations read

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\Big\{ -\big(\mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1 + \|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2\big)\Big\} \tag{9}$$

and

$$p(\boldsymbol{a}|\boldsymbol{y}) \propto \exp\Big\{ -\big(\mu\|\boldsymbol{a}\|_1 + \|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2\big)\Big\}, \tag{10}$$

respectively, where the first terms (*i.e.* the ℓ_1 norm terms) in the exponentials of each equation correspond to the prior and the second (*i.e.* the ℓ_2 norm terms) correspond to the likelihood.

Drawing conclusions directly from $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{a}|\boldsymbol{y})$ can be difficult because of the high dimensionality involved. Instead, Bayesian methods often derive solutions by computing estimators that summarise $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{a}|\boldsymbol{y})$. In particular, it is often common practice to compute maximum-a-posteriori (MAP) estimators given by

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{map}} &= \operatorname*{argmax}_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y}) \\ &= \operatorname*{argmin}_{\boldsymbol{x}} \Big\{\mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1 + \|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2\Big\},\end{aligned} \tag{11}$$

for the analysis model, and

$$\begin{aligned}\hat{\boldsymbol{a}}_{\mathrm{map}} &= \operatorname*{argmax}_{\boldsymbol{a}} p(\boldsymbol{a}|\boldsymbol{y}) \\ &= \operatorname*{argmin}_{\boldsymbol{a}} \Big\{\mu\|\boldsymbol{a}\|_1 + \|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2\Big\},\end{aligned} \tag{12}$$

which is then mapped to canonical coordinates by using (3), for the synthesis model. A main computational advantage of the MAP estimators (11) and (12) is that they can be formulated as a convex optimisation problem that can be solved very efficiently, even in high dimensions, by using modern convex optimisation techniques (Green et al. 2015). Also, there is abundant empirical evidence that these estimators deliver accurate reconstruction results, and that they promote solutions that are sparse under $\boldsymbol{\Psi}$ in agreement with our prior knowledge about $\boldsymbol{x}$. See Pereyra (2016a) for a theoretical analysis of MAP estimation.

The regularisation parameter μ appearing in the analysis and synthesis formulations controls the balance between the likelihood and the prior information, and plays an important role in terms of image reconstruction quality. Typically, setting μ is performed by visual cross-validation. However, there exist more advanced Bayesian strategies to address the problem of unknown μ. For example, hierarchical Bayesian strategies allow estimating μ jointly with $\boldsymbol{x}$ (or $\boldsymbol{\alpha}$) from $\boldsymbol{y}$, or removing μ from the model by marginalisation followed by inference with the marginal model (see Pereyra et al. 2015 for details). Alternatively, empirical Bayesian approaches set regularisation parameters by marginal maximum likelihood estimation (Junklewitz et al. 2016; Fernandez Vidal & Pereyra 2018) or by MCMC sampling (Sutter et al. 2014). The selection of a regularisation parameter was also studied by Skilling & Gull (1991) in the context of maximum entropy methods, where the marginal distribution of the regularisation parameter is again maximised.

To compute other Bayesian estimators or quantifies of interest beyond MAP estimators it is typically necessary to use more advanced Bayesian computation tools, such as MCMC sampling methods. These methods compute probabilities and expectations w.r.t. $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{a}|\boldsymbol{y})$ and can be used to calculate moments and Bayesian confidence regions useful for uncertainty quantification. This is the main purpose of this article and thus will be detailed subsequently.

2.4 Connections with alternative approaches

It is worth noticing that many RI imaging techniques can be seen as regularisation techniques and many of them can be viewed as MAP estimation for appropriate priors. While this interpretation is not always precise, the resulting approximate unifying Bayesian framework is useful to aid intuition.

2.4.1 *Compressive sensing and ℓ_1-regularised regression*

The theory of CS (compressive sensing) led to an important breakthrough in the recovery of sparse signals from incomplete linear measurements (Donoho 2006; Candes & Wakin 2008; Candes et al. 2010). CS goes beyond the traditional Nyquist sampling paradigm, where its acquisition approaches can save a huge amount of time and memory thanks to the fact that natural signals often exhibit a sparse representation in multi-scale bases. CS can be implemented for signal reconstruction by regularising the resulting ill-posed inverse problem through a sparsity-promoting prior, resulting in a convex optimisation problem that can be solved by leveraging techniques from the field of convex optimisation. Briefly speaking, the theoretical framework of CS motivates sparse regularisation approaches such as the ones used in (11) and (12). In fact, the MAP estimators (11) and (12) are equivalent to the ℓ_1 regularised least-squares estimators used extensively in CS. In the literature and henceforth, the discussion of CS-based methods for RI imaging typically refers to sparse regularisation approaches, even though RI imaging models such as (11) and (12) may not satisfy the idealised CS setting.

2.4.2 *CLEAN*

CLEAN, the most well-known and standard RI image reconstruction algorithm, is a non-linear deconvolution method based on local iterative beam removal. In general, it can be operated iteratively in two steps, *i.e.* major and minor cycles. Let $\chi^2 = \|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2$ and denote the gradient of χ^2 at iteration t by $\boldsymbol{r}^{(t)} = \boldsymbol{\Phi}^\dagger(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^{(t)})$. The major cycle of CLEAN computes the residual image $\boldsymbol{r}^{(t)}$, followed by the minor cycle of deconvolving the brightest sources in $\boldsymbol{r}^{(t)}$, represented by $\mathcal{T}(\boldsymbol{r}^{(t)})$, yielding the iterative form

$$\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} + \mathcal{T}(\boldsymbol{r}^{(t)}) \tag{13}$$

to reconstruct an image $\boldsymbol{x}$.

Extensions of CLEAN have also been considered to achieve

better reconstruction. For example, multi-scale versions of CLEAN: MS-CLEAN (Cornwell 2008); and ASP-CLEAN (Bhatnagar & Corwnell 2004). For further variants of CLEAN, please refer to Rau et al. (2009) and references therein.

CLEAN implicitly involves a sparse prior on the original signal in real space. Moreover, a close connection has been shown between CLEAN and the well-known Matching Pursuit algorithm in the CS literature (Cornwell 1988; Wiaux et al. 2009a; Rau et al. 2009); in other words, CLEAN is essentially ℓ_0 regularisation with a point source basis. The performance of CLEAN, however, is empirically found to be similar to ℓ_1 regularisation with a point source basis (Wiaux et al. 2009a). As a proxy for CLEAN, ℓ_1 regularisation with a point source basis is equivalent to MAP estimation involving a Laplace prior.

2.4.3 *Maximum entropy method (MEM)*

Another important method for RI imaging is MEM, which is, mildly speaking, a special case of the MAP method. The MEM approach for RI imaging (Cornwell & Evans 1985) differs to the original MEM formulation (Ables 1974; Gull & Daniell 1978), in that not only does the regularisation problem considered consist of a relative entropic prior and a (Gaussian) likelihood, but an an additional flux constraint is also incorporated. In particular, an entropic prior, $\exp(-\mu \boldsymbol{x}^\dagger \log \boldsymbol{x})$, on the image is adopted.

2.4.4 *Constrained regularisation*

In addition to the unconstrained optimisation problems of (11) and (12), many CS-based approaches consider constrained forms of the analysis and synthesis models, which are, respectively, given by

$$\min_{\boldsymbol{x}} \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1, \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \leq \epsilon \tag{14}$$

and

$$\min_{\boldsymbol{a}} \|\boldsymbol{a}\|_1, \quad \text{s.t.} \quad \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 \leq \epsilon, \tag{15}$$

where ϵ is an upper-bound related to the noise level present in $\boldsymbol{y}$. CS approaches based on constrained optimisation problems, solved via convex optimisation techniques, have been applied broadly in RI imaging (Wiaux et al. 2009a,b; McEwen & Wiaux 2011; Li et al. 2011a,b; Carrillo et al. 2012, 2014; Onose et al. 2016; Pratley et al. 2018). These techniques have shown promising results, with improvements in terms of image fidelity and flexibility compared to traditional approaches such as CLEAN-based methods and MEM. For these constrained regularisation approaches, parallel implementation structures have also been explored (Carrillo et al. 2014; Onose et al. 2016). Compared with the unconstrained analysis and synthesis models, constrained approaches are parameterised by ϵ (related to noise level) which controls the error of the reconstruction explicitly; in contrast, unconstrained models use regularisation parameter μ to impose a tradeoff between the prior and data fidelity. The constrained approach therefore avoids the problem of unknown regularisation parameter μ, replacing it with the problem of estimating the noise bound ϵ. The latter can be performed in a principled manner by noting that for Gaussian noise the ℓ_2 norm data fidelity term follows a χ^2 distribution with $2M$ degrees of freedom (see, *e.g.*, Carrillo et al. 2012). While constrained problems do not afford a straightforward Bayesian interpretation, the constrained and unconstrained models are closely related (Nikolova 2016).

3 BAYESIAN INFERENCE WITH SPARSE PRIORS BY PROXIMAL MCMC SAMPLING

Sparse regularisation, motivated by CS, has been shown to be a powerful framework for solving inverse problems and has been used to deal with the recovery of sparse signals from incomplete linear measurements (*e.g.*, Donoho 2006). It has been demonstrated that sparse signals can be recovered accurately from incomplete data under some conditions. Sparse priors have also been ushered into RI imaging for image reconstruction (*e.g.* Wiaux et al. 2009a; McEwen & Wiaux 2011), and have shown promising results on real RI data (Pratley et al. 2018). Unfortunately, CS-based techniques do not provide any uncertainty information regarding their point estimates. This is also a limitation of CLEAN-based methods and MEM.

From an inferential viewpoint, the lack of uncertainty quantification is problematic, particularly because RI problems are ill-posed and hence solutions have significant intrinsic uncertainty. As explained previously, in this article we apply recent developments in Bayesian methodologies to analyse uncertainty in RI imaging. Precisely, we use new MCMC Bayesian computation algorithms to compute probabilities and expectations w.r.t. the posterior distribution of interest, *i.e.*, $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{a}|\boldsymbol{y})$ given by (7) and (8), depending on whether an analysis or a synthesis formulation is used. This involves constructing a Markov chain that generates samples from the distribution of interest, and then using the samples to approximate probabilities and expectations by Monte Carlo integration (Robert & Casella 2004). Computing such Markov chains in large-scale settings is computationally challenging, and we address this difficulty by using state-of-the-art MCMC methods tailored for these types of problems (Pereyra 2016b; Durmus et al. 2016). In this section we introduce these MCMC algorithms. To ease presentation, all symbols and dimensions specified here corresponds to the analysis model (11), however these can be straightforwardly adapted to the synthesis model (12).

3.1 Preliminaries

A function $g : \mathbb{C}^N \to (-\infty, \infty]$ is said to be lower semicontinuous (l.s.c.) if for all $M \in \mathbb{R}$, $\{g < M\}$ is a closed subset of $\mathbb{C}^N$. Let $\mathcal{C}^1(\mathbb{C}^N)$ be the class of continuously differentiable functions on $\mathbb{C}^N$. If $g \in \mathcal{C}^1(\mathbb{C}^N)$, denote by ∇g the gradient of g. Also, ∇g is said to be Lipchitz continuous with constant $\beta_{\text{Lip}} \in (0, \infty)$ if

$$\|\nabla g(\hat{\boldsymbol{z}}) - \nabla g(\bar{\boldsymbol{z}})\| \leq \beta_{\text{Lip}} \|\hat{\boldsymbol{z}} - \bar{\boldsymbol{z}}\|, \quad \forall (\hat{\boldsymbol{z}}, \bar{\boldsymbol{z}}) \in \mathbb{C}^N \times \mathbb{C}^N. \tag{16}$$

Moreover, let $h : \mathbb{C}^N \to (-\infty, \infty]$ be a convex l.s.c. function and $\lambda > 0$. The λ-Moreau-Yosida envelope of h is a carefully regularised approximation of h given by

$$h^\lambda(\boldsymbol{z}) \equiv \min_{\boldsymbol{u} \in \mathbb{R}^N} \left\{ h(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{z}\|^2 / 2\lambda \right\}. \tag{17}$$

The approximation h^λ can be made arbitrarily close to h by adjusting λ, *i.e.*, $\lim_{\lambda \to 0} h^\lambda(\boldsymbol{z}) = h(\boldsymbol{z})$ (see Parikh & Boyd 2014). Also, by construction $h^\lambda \in \mathcal{C}^1$, with λ-Lipchitz gradient given by

$$\nabla h^\lambda(\boldsymbol{z}) = \left(\boldsymbol{z} - \text{prox}_h^\lambda(\boldsymbol{z}) \right) / \lambda, \tag{18}$$

where $\text{prox}_h^\lambda(\boldsymbol{z})$ is the *proximity operator* of h at $\boldsymbol{z}$ defined as

$$\text{prox}_h^\lambda(\boldsymbol{z}) \equiv \underset{\boldsymbol{u} \in \mathbb{R}^N}{\text{argmin}} \left\{ h(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{z}\|^2 / 2\lambda \right\}. \tag{19}$$

It can be verified easily that $\text{prox}_h^\lambda(\boldsymbol{z}) = \text{prox}_{\lambda h}(\boldsymbol{z})$. For simplicity, we represent $\text{prox}_h^1(\boldsymbol{z})$ by $\text{prox}_h(\boldsymbol{z})$. This operator generalises

the projection operator defined as

$$\mathcal{P}_C(\boldsymbol{z}) \equiv \underset{\boldsymbol{u}\in\mathbb{R}^N}{\operatorname{argmin}} \left\{ \iota_C(\boldsymbol{u}) + \|\boldsymbol{u}-\boldsymbol{z}\|^2/2 \right\}, \tag{20}$$

where ι_C is the characteristic function for the convex set C defined by $\iota_C(\boldsymbol{u}) = \infty$ if $\boldsymbol{u} \notin C$ and 0 otherwise.

3.2 Langevin MCMC

Let π be a probability density (or a user-specified target density), such as the posteriors $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{a}|\boldsymbol{y})$. When π is defined on $\mathbb{C}^n$ and assume $\pi \in \mathcal{C}^1$ with Lipchitz gradient, the Langevin diffusion on $\mathbb{C}^n$ associated with π is a stochastic process defined as

$$\mathrm{d}\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi[\mathcal{L}(t)]\mathrm{d}t + \mathrm{d}\mathcal{W}(t), \tag{21}$$

where $\mathcal{W}$ is the Brownian motion on $\mathbb{C}^n$. This process converges to π as t increases, and is therefore useful for generating samples from π. Unfortunately, simulating $\mathcal{L}(t)$ in continuous time is generally not possible, so instead we use discrete-time approximations. In particular, ULA (unadjusted Langevin algorithm) is based on a forward Euler-Maruyama approximation with step-size $\delta > 0$, resulting in the Markov chain

$$\boldsymbol{l}^{(m+1)} = \boldsymbol{l}^{(m)} + \frac{\delta}{2}\nabla \log \pi[\boldsymbol{l}^{(m)}] + \sqrt{\delta}\boldsymbol{w}^{(m+1)}, \tag{22}$$

where $\boldsymbol{w}^{(m+1)} \sim \mathcal{N}(0, \mathbb{1}_N)$ (an N-sequence of standard Gaussian random variables). Under appropriate regularity conditions, the chain generated by ULA converges to an ergodic measure which is close to π. In MALA (Metropolis-adjusted Langevin Algorithm), this approximation error is corrected by complementing ULA with an MH (Metropolis-Hasting) accept-reject step targeting π, which removes the asymptotic bias due to the discretisation at the expense of some additional estimation variance (Roberts & Tweedie 1996). Theoretical and empirical results show that ULA and MALA scale very efficiently to high dimensions.

However, a main limitation of ULA and MALA (and generally MCMC methods based on gradients) is the requirement that $\log \pi$ is continuously differentiable with Lipchitz gradient, otherwise the Markov chain (22) fails to converge. As explained previously, this prohibits their application to image processing models with non-smooth densities, *e.g*, involving the term $\phi(\cdot) = \|\cdot\|_1$. In Pereyra (2016b), this limitation of ULA and MALA is addressed by using the Moreau-Yosida envelope of $\log \pi$ to regularise the diffusion process to handle non-smoothness, *e.g.* sparse priors.

3.3 Moreau-Yosida regularised ULA (MYULA)

We consider models of the form $\pi(\boldsymbol{x}) \propto \exp\{-f(\boldsymbol{x}) - g(\boldsymbol{x})\}$, where $f \notin \mathcal{C}^1$ is l.s.c. convex with operator $\operatorname{prox}_f^\lambda(\boldsymbol{z})$ tractable $\forall \boldsymbol{z} \in \mathbb{C}^N$, and $g \in \mathcal{C}^1$ is l.s.c. convex with ∇g and β_{Lip}-Lipchitz continuous. Typically f corresponds to the log-prior and g to the log-likelihood.

We wish to use the Langevin diffusion (21) to generate samples from π but this is not directly possible since f is not smooth, *i.e.* $f \notin \mathcal{C}^1$. The key idea underpinning proximal ULA and MALA is to carefully regularise f to guarantee that (21) and its discrete-time approximation (22) have good convergence properties (Pereyra 2016b). This is achieved by defining an approximation

$$\pi_\lambda(\boldsymbol{x}) = \frac{\exp\{-f^\lambda(\boldsymbol{x}) - g(\boldsymbol{x})\}}{\int \exp\{-f^\lambda(\boldsymbol{x}) - g(\boldsymbol{x})\}\mathrm{d}\boldsymbol{x}}, \tag{23}$$

where the non-smooth term f is replaced by its Moreau-Yosida envelope f^λ. Since $\nabla \log \pi_\lambda = -\nabla f^\lambda - \nabla g$ is Lipchitz continuous, the Langevin diffusion associated with π_λ is well posed and leads to a Markov chain (22) with good convergence properties. Precisely, the MYULA chain is defined by

$$\begin{aligned}\boldsymbol{l}^{(m+1)} = &\left(1 - \frac{\delta}{\lambda}\right)\boldsymbol{l}^{(m)} + \frac{\delta}{\lambda}\operatorname{prox}_f^\lambda(\boldsymbol{l}^{(m)}) - \delta\nabla g(\boldsymbol{l}^{(m)}) \\ &+ \sqrt{2\delta}\boldsymbol{w}^{(m)},\end{aligned} \tag{24}$$

where we have noted that $\nabla f^\lambda(\boldsymbol{z}) = \left(\boldsymbol{z} - \operatorname{prox}_f^\lambda(\boldsymbol{z})\right)/\lambda$.

The MYULA chain (24) scales well in high dimensions and efficiently delivers samples that are approximately distributed according to π. The approximation error involved can be made arbitrarily small by reducing the value of λ and by increasing the number of iterations (Durmus et al. 2016).

Finally, in our experiments we implement (24) with $f(\boldsymbol{x}) = \mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1$, $g(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2$ for the analysis model (11) (the setting for the synthesis model (12) is analogous), and by setting $\lambda = 2/\beta_{\mathrm{Lip}}$ and $\delta \in [1/5\beta_{\mathrm{Lip}}, 1/2\beta_{\mathrm{Lip}}]$, as suggested by Durmus et al. (2016).

3.4 Proximal MALA (Px-MALA)

In a manner akin to MALA, the Px-MALA combines MYULA with an MH step targeting the desired density π which is not differentiable (Pereyra 2016b). At each iteration of the algorithm a new candidate $\boldsymbol{l}^*$ is generated by using one MYULA iteration as proposal mechanism. The candidate is then accepted with probability

$$\rho = \min\left\{1, \frac{q(\boldsymbol{l}^{(m)}|\boldsymbol{l}^*)\pi(\boldsymbol{l}^*)}{q(\boldsymbol{l}^*|\boldsymbol{l}^{(m)})\pi(\boldsymbol{l}^{(m)})}\right\}, \tag{25}$$

where $q(\cdot|\cdot)$ is the MYULA transition kernel defined by (Pereyra et al. 2016)

$$q(\boldsymbol{l}^*|\boldsymbol{l}^{(m)}) \sim \exp\left(-\frac{\left(\boldsymbol{l}^* - \boldsymbol{l}^{(m)} - \frac{\delta}{2}\nabla \log \pi(\boldsymbol{l}^{(m)})\right)^2}{2\delta}\right). \tag{26}$$

Regarding computational efficiency, for the models considered here Px-MALA inherits the good convergence properties of MYULA and scales efficiently in high dimensions. However, note that the MH correction removes the asymptotic estimation bias at the expense of increasing the correlation of the Markov chain and hence the estimation variance (this is observed clearly in the experiments reported in Section 6). Also note that Px-MALA iterations are more expensive than MYULA iterations because of the computational overhead associated with the MH step.

Finally, in our experiments, following the setting in Pereyra (2016b), we implement Px-MALA with $f(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2 + \mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1$, $g(\boldsymbol{x}) = 0$ for the analysis model (11) (the setting for the synthesis model (12) is analogous), and by setting $\lambda = 2/\beta_{\mathrm{Lip}}$ and adjusting δ for an acceptance probability of approximately 0.5. Other settings w.r.t. the definitions of f and g, *e.g.* as used in MYULA, could also be considered. Also note that the efficient computation of $\operatorname{prox}_f^\lambda$ often involves some approximations, which we also correct with the MH step. We discuss such approximations for the analysis and synthesis models in Section 4.

4 PROXIMAL MCMC METHODS FOR RI IMAGING

This section presents the implementation details of MYULA and Px-MALA for the analysis model (11) and the synthesis model (12). We first consider the computation of the proximity operator of f, for different forms of f. Computing the proximity operator of f requires solving an optimisation problem, which must be performed efficiently since it needs to be computed to generate each sample by (24). We then summarise the sampling procedures for the two proximal MCMC methods. Note that computing the gradient of g in (24) is straightforward since it is differentiable. For clarity, we henceforth use the label $\bar{}$ for symbols related to the analysis model, and $\hat{}$ for symbols related to the synthesis model. Although not essential, we also assume $\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} = \mathbf{I}$ (where $\mathbf{I}$ is the identity matrix), unless otherwise stated.

4.1 Computing proximity operators

Algorithm 1: Sample generation by MYULA

Input: visibility $\boldsymbol{y} \in \mathbb{C}^M$, $\boldsymbol{x}^{(0)} \in \mathbb{R}^N$, $\boldsymbol{a}^{(0)} \in \mathbb{C}^L$, K, K_{gap}, K_{burn}, $P_{\text{type}} \in \{\texttt{analysis}, \texttt{synthesis}\}$, and $m = 0, j = 1$
Output: K samples $\{\boldsymbol{x}^{(j)}\}_{j=1}^K$ or $\{\boldsymbol{\Psi}\boldsymbol{a}^{(j)}\}_{j=1}^K$
do
  if $P_{\text{type}} == \texttt{analysis}$
    compute $\boldsymbol{x}^{(m+1)}$
    $= \boldsymbol{x}^{(m)} - \delta\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x}^{(m)} - \boldsymbol{y})/2\sigma^2 + \sqrt{\delta}\bar{\boldsymbol{w}}^{(m)}$
      $+\frac{\delta}{\lambda}\boldsymbol{\Psi}\left(\text{soft}_{\lambda\mu/2}(\boldsymbol{\Psi}^\dagger\boldsymbol{x}^{(m)}) - \boldsymbol{\Psi}^\dagger\boldsymbol{x}^{(m)})\right)$
    set $\boldsymbol{z} = \boldsymbol{x}^{(m+1)}$
  elseif $P_{\text{type}} == \texttt{synthesis}$
    compute $\boldsymbol{a}^{(m+1)}$
      $= (1 - \frac{\delta}{\lambda})\boldsymbol{a}^{(m)} + \frac{\delta}{\lambda}\text{soft}_{\lambda\mu/2}(\boldsymbol{a}^{(m)})$
      $-\delta\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/2\sigma^2 + \sqrt{\delta}\hat{\boldsymbol{w}}^{(m)}$
    set $\boldsymbol{z} = \boldsymbol{a}^{(m+1)}$
  endif
  if m satisfies (45)
    if $P_{\text{type}} == \texttt{analysis}$
      set $\boldsymbol{x}^{(j)} = \boldsymbol{z}$
    elseif $P_{\text{type}} == \texttt{synthesis}$
      set $\boldsymbol{a}^{(j)} = \boldsymbol{z}$
    endif
    $j = j + 1$
  endif
  $m = m + 1$
while $j \leq K$;

Before considering the computation of various proximity operators for the analysis and synthesis forms, define, $\forall \boldsymbol{z} \in \mathbb{R}^L$, the soft-thresholding operator with threshold β_{th} as

$$\text{soft}_{\beta_{\text{th}}}(\boldsymbol{z}) = \left(\text{soft}_{\beta_{\text{th}}}(z_1), \cdots, \text{soft}_{\beta_{\text{th}}}(z_L)\right), \tag{27}$$

where for $i = 1, \ldots, L$,

$$\text{soft}_{\beta_{\text{th}}}(z_i) = \begin{cases} 0, & \text{if } |z_i| \leq \beta_{\text{th}}, \\ z_i(|z_i| - \beta_{\text{th}})/|z_i|, & \text{otherwise.} \end{cases} \tag{28}$$

Algorithm 2: Sample generation by Px-MALA

Input: visibility $\boldsymbol{y} \in \mathbb{C}^M$, $\boldsymbol{x}^{(0)} \in \mathbb{R}^N$, $\boldsymbol{a}^{(0)} \in \mathbb{C}^L$, K, K_{gap}, K_{burn}, $P_{\text{type}} \in \{\texttt{analysis}, \texttt{synthesis}\}$, and $m = 0, j = 1$
Output: K samples $\{\boldsymbol{x}^{(j)}\}_{j=1}^K$ or $\{\boldsymbol{\Psi}\boldsymbol{a}^{(j)}\}_{j=1}^K$
do
  if $P_{\text{type}} == \texttt{analysis}$
    compute $\boldsymbol{x}^{(m+1)} = \text{prox}_{\bar{f}}^{\delta/2}(\boldsymbol{x}^{(m)}) + \sqrt{\delta}\bar{\boldsymbol{w}}^{(m)}$
    set $\boldsymbol{z} = \boldsymbol{x}^{(m+1)}$, $\boldsymbol{z}' = \boldsymbol{x}^{(j-1)}$
  elseif $P_{\text{type}} == \texttt{synthesis}$
    compute $\boldsymbol{a}^{(m+1)} = \text{prox}_{\hat{f}}^{\delta/2}(\boldsymbol{a}^{(m)}) + \sqrt{\delta}\hat{\boldsymbol{w}}^{(m)}$
    set $\boldsymbol{z} = \boldsymbol{a}^{(m+1)}$, $\boldsymbol{z}' = \boldsymbol{a}^{(j-1)}$
  endif
  if m satisfies (45)
    if $\texttt{MH}(\boldsymbol{z}, \boldsymbol{z}') == 1$ // Metropolis-Hasting step
      if $P_{\text{type}} == \texttt{analysis}$
        set $\boldsymbol{x}^{(j)} = \boldsymbol{z}$
      elseif $P_{\text{type}} == \texttt{synthesis}$
        set $\boldsymbol{a}^{(j)} = \boldsymbol{z}$
      endif
      $j = j + 1$
    endif
  endif
  $m = m + 1$
while $j \leq K$;

function $\texttt{MH}(\boldsymbol{l}^*, \boldsymbol{l})$
  Compute the acceptance probability
    $\rho = \min\left\{1, \frac{q(\boldsymbol{l}|\boldsymbol{l}^*)\pi(\boldsymbol{l}^*)}{q(\boldsymbol{l}^*|\boldsymbol{l})\pi(\boldsymbol{l})}\right\}$
  Generate a threshold $u \sim \mathcal{U}(0, 1)$
  if $u \leq \rho$
    return 1 // Accept the candidate
  elseif
    return 0 // Reject the candidate
  endif
end function

4.1.1 *Analysis form: MYULA*

To implement MYULA for the analysis model (11), we set $\bar{f}(\boldsymbol{x}) = \mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1$ and $\bar{g}(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2$. Then, to compute the iteration (24) it is necessary to evaluate $\text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x})$ and $\nabla\bar{g}(\boldsymbol{x})$.

To evaluate $\text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x})$ we use the closed-form representation (Combettes & Pesquet 2010, see Table 1),

$$\begin{aligned} \text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x}) &= \underset{\boldsymbol{u}\in\mathbb{R}^N}{\text{argmin}}\ \lambda\mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{u}\|_1 + \|\boldsymbol{u} - \boldsymbol{x}\|^2/2 \\ &= \boldsymbol{x} + \boldsymbol{\Psi}\left(\text{prox}_{\mu\|\cdot\|_1}^{\lambda}(\boldsymbol{\Psi}^\dagger\boldsymbol{x}) - \boldsymbol{\Psi}^\dagger\boldsymbol{x}\right) \\ &= \boldsymbol{x} + \boldsymbol{\Psi}\left(\text{soft}_{\lambda\mu}(\boldsymbol{\Psi}^\dagger\boldsymbol{x}) - \boldsymbol{\Psi}^\dagger\boldsymbol{x}\right). \end{aligned} \tag{29}$$

Moreover,

$$\nabla\bar{g}(\boldsymbol{x}) = \nabla(\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2) = \boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/\sigma^2. \tag{30}$$

Remark 4.1. If $\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} \neq \mathbf{I}$, the case where $\boldsymbol{\Psi}$ is overcomplete,

$\text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x})$ can be computed in an iterative manner:

$$\boldsymbol{u}^{(t+\frac{1}{2})} = \lambda_{\text{ite}}^{(t)}(\mathbf{1} - \text{prox}^{\lambda}_{\|\cdot\|_1/\lambda_{\text{ite}}^{(t)}})\left(\frac{\boldsymbol{u}^{(t-\frac{1}{2})}}{\lambda_{\text{ite}}^{(t)}} + \boldsymbol{\Psi}^{\dagger}\boldsymbol{u}^{(t)}\right), \quad (31)$$

$$\boldsymbol{u}^{(t+1)} = \boldsymbol{x} - \boldsymbol{\Psi}\boldsymbol{u}^{(t+\frac{1}{2})}, \quad (32)$$

where $\lambda_{\text{ite}}^{(t)} \in (0, 2/\beta_{\text{Par}})$ (β_{Par} is a constant satisfying $\|\boldsymbol{\Psi}\boldsymbol{z}\|^2 \leq \beta_{\text{Par}}\|\boldsymbol{z}\|^2, \forall \boldsymbol{z} \in \mathbb{R}^L$) is a predefined step size and $\boldsymbol{u}^{(t)} \to \text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x})$; refer to Fadili & Starck (2009) and Jacques et al. (2011) for details.

4.1.2 *Analysis form: Px-MALA*

To implement Px-MALA for the analysis model (11), we set $\bar{f}(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2 + \mu\|\boldsymbol{\Psi}^{\dagger}\boldsymbol{x}\|_1$ and $\bar{g}(\boldsymbol{x}) = 0$. Therefore, at each iteration of the algorithm it is necessary to evaluate

$$\text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x}) = \underset{\boldsymbol{u}\in\mathbb{R}^N}{\text{argmin}}\left\{\mu\|\boldsymbol{\Psi}^{\dagger}\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{x}\|_2^2}{2\lambda}\right\}. \quad (33)$$

By the Taylor expansion of $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2$ at point $\boldsymbol{x}$,

$$\begin{aligned}\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2^2 \approx& \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + (\boldsymbol{u} - \boldsymbol{x})^{\top}\nabla\left(\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2\right) \\ =& \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + 2(\boldsymbol{u} - \boldsymbol{x})^{\top}\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y}),\end{aligned} \quad (34)$$

and we obtain the following approximation of $\text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x})$,

$$\begin{aligned}&\underset{\boldsymbol{u}\in\mathbb{R}^N}{\text{argmin}}\Big\{\mu\|\boldsymbol{\Psi}^{\dagger}\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{u} - \boldsymbol{x}\|_2^2}{2\lambda} + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2}{2\sigma^2} \\ &\qquad + (\boldsymbol{u} - \boldsymbol{x})^{\top}\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/\sigma^2\Big\} \\ \approx& \underset{\boldsymbol{u}\in\mathbb{R}^N}{\text{argmin}}\left\{\mu\|\boldsymbol{\Psi}^{\dagger}\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{u} - \boldsymbol{x} + \delta\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/2\sigma^2\|_2^2}{2\lambda}\right\} \\ =& \text{prox}^{\lambda}_{\mu\|\boldsymbol{\Psi}^{\dagger}\cdot\|_1}\left(\boldsymbol{x} - \lambda\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/\sigma^2\right).\end{aligned} \quad (35)$$

Let $\bar{\boldsymbol{v}} = \boldsymbol{x} - \lambda\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/\sigma^2$, using (29), we have

$$\text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x}) \approx \bar{\boldsymbol{v}} + \boldsymbol{\Psi}\left(\text{soft}_{\mu\lambda}(\boldsymbol{\Psi}^{\dagger}\bar{\boldsymbol{v}}) - \boldsymbol{\Psi}^{\dagger}\bar{\boldsymbol{v}}\right). \quad (36)$$

Note that $\text{prox}_{\bar{f}}^{\lambda}(\boldsymbol{x})$ here can be computed in the same manner as the one mentioned in remark 4.1 if $\boldsymbol{\Psi}^{\dagger}\boldsymbol{\Psi} \neq \mathbf{I}$.

Remark 4.2. The approximation shown in (36) can be regarded as one iteration of the forward-backward algorithm (Combettes & Pesquet 2010) minimising objective function $\bar{f} + \bar{g}$. The Taylor approximation performed above makes the assumptions in performing a single forward-backward iteration explicit.

4.1.3 *Synthesis form: MYULA*

To implement MYULA for the synthesis model (12), we set $\hat{f}(\boldsymbol{a}) = \mu\|\boldsymbol{a}\|_1$ and $\hat{g}(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2$. Then, to compute the iteration (24) it is necessary to evaluate

$$\begin{aligned}\text{prox}^{\lambda}_{\mu\|\cdot\|_1}(\boldsymbol{a}) &= \underset{\boldsymbol{u}\in\mathbb{R}^L}{\text{argmin}}\left\{\mu\|\boldsymbol{u}\|_1 + \|\boldsymbol{u} - \boldsymbol{a}\|^2/2\lambda\right\}, \\ &= \text{soft}_{\lambda\mu}(\boldsymbol{a}),\end{aligned} \quad (37)$$

and

$$\nabla\hat{g}(\boldsymbol{a}) = \nabla(\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2) = \boldsymbol{\Psi}^{\dagger}\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/\sigma^2. \quad (38)$$

4.1.4 *Synthesis form: Px-MALA*

To implement Px-MALA for the synthesis model (12), we set $\hat{f}(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2 + \mu\|\boldsymbol{a}\|_1$ and $\hat{g}(\boldsymbol{a}) = 0$. Therefore, at each iteration of the algorithm it is necessary to evaluate

$$\text{prox}_{\hat{f}}^{\lambda}(\boldsymbol{a}) = \underset{\boldsymbol{u}\in\mathbb{R}^L}{\text{argmin}}\left\{\mu\|\boldsymbol{u}\|_1 + \frac{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{u}\|_2^2}{2\sigma^2} + \frac{\|\boldsymbol{u} - \boldsymbol{a}\|_2^2}{2\lambda}\right\}. \quad (39)$$

By proceeding similarly to (36) we obtain

$$\begin{aligned}\text{prox}_{\hat{f}}^{\lambda}(\boldsymbol{a}) &\approx \text{prox}^{\lambda}_{\mu\|\cdot\|_1}\left(\boldsymbol{a} - \lambda\boldsymbol{\Psi}^{\dagger}\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/\sigma^2\right) \\ &\approx \text{soft}_{\mu\lambda}\left(\boldsymbol{a} - \lambda\boldsymbol{\Psi}^{\dagger}\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/\sigma^2\right),\end{aligned} \quad (40)$$

where the first line of (40) follows by (37).

Remark 4.3. Similar to Remark 4.2, the approximation shown in (40) can be regarded as one iteration of the forward-backward algorithm (Combettes & Pesquet 2010) minimising $\hat{f} + \hat{g}$. Again, the above derivations make the corresponding assumptions explicit.

4.2 Sampling by proximal MCMC methods

Using formulas (30) and (38) which compute gradient operators, formulas (29) and (37) which compute proximity operators according to sparse regularisations, and the MYULA iterative formula (24), a set of full posterior samples for the analysis model (11) and synthesis model (12) can be generated by

$$\begin{aligned}\boldsymbol{x}^{(m+1)} =& \boldsymbol{x}^{(m)} + \frac{\delta}{\lambda}\boldsymbol{\Psi}\left(\text{soft}_{\lambda\mu/2}(\boldsymbol{\Psi}^{\dagger}\boldsymbol{x}^{(m)}) - \boldsymbol{\Psi}^{\dagger}\boldsymbol{x}^{(m)})\right) \\ &- \delta\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{x}^{(m)} - \boldsymbol{y})/2\sigma^2 + \sqrt{\delta}\bar{\boldsymbol{w}}^{(m)}\end{aligned} \quad (41)$$

and

$$\begin{aligned}\boldsymbol{a}^{(m+1)} =& (1 - \frac{\delta}{\lambda})\boldsymbol{a}^{(m)} + \frac{\delta}{\lambda}\text{soft}_{\lambda\mu/2}(\boldsymbol{a}^{(m)}) \\ &- \delta\boldsymbol{\Psi}^{\dagger}\boldsymbol{\Phi}^{\dagger}(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/2\sigma^2 + \sqrt{\delta}\hat{\boldsymbol{w}}^{(m)},\end{aligned} \quad (42)$$

respectively, where $\bar{\boldsymbol{w}}^{(m)} \in \mathbb{R}^N \sim \mathcal{N}(0, \mathbb{1}_N)$ and $\hat{\boldsymbol{w}}^{(m)} \in \mathbb{R}^L \sim \mathcal{N}(0, \mathbb{1}_L)$.

Analogously, using formulas (36) and (40), the Px-MALA iterative forms generating samples as to the analysis and synthesis models can be written as

$$\boldsymbol{x}^{(m+1)} = \text{prox}_{\bar{f}}^{\delta/2}(\boldsymbol{x}^{(m)}) + \sqrt{\delta}\bar{\boldsymbol{w}}^{(m)}, \quad (43)$$

and

$$\boldsymbol{a}^{(m+1)} = \text{prox}_{\hat{f}}^{\delta/2}(\boldsymbol{a}^{(m)}) + \sqrt{\delta}\hat{\boldsymbol{w}}^{(m)}, \quad (44)$$

respectively. After a proper candidate generated by (43) or (44), Px-MALA includes an MH accept-reject step with an acceptance probability ρ, specified by (25), to ensure the sequence converges to the target distribution.

To generate K samples using the proximal MCMC methods proposed, two parameters controlling sample candidates should be assigned: (i) the number of initial or *burn-in* iterations, $K_{\text{burn}} \in \mathbb{Z}$ (denotes the previous number of iterations that are discarded); and (ii) the chain's *thinning* factor or number of intermediate iterations between samples, $K_{\text{gap}} \in \mathbb{Z}$ (denotes the intermediate number of iterations that are discarded; used to reduce correlations between samples and the algorithm's memory footprint). Because of memory limitations we do not store all samples (generated by (41), (42), (43) or (44)), and only store 1-in-K_{gap} samples if

$$m > K_{\text{burn}} \quad \text{and} \quad \texttt{mod}(m - K_{\text{burn}}, K_{\text{gap}}) = 0, \quad (45)$$

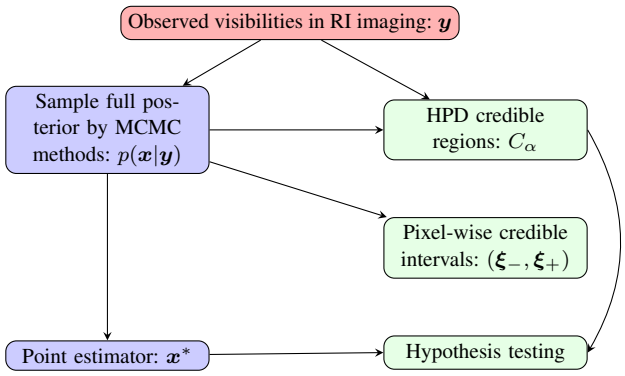


Figure 1. Our proposed uncertainty quantification procedure for RI imaging based on proximal MCMC sampling. The light green areas on the right show the types of uncertainty quantification developed. Firstly, the full posterior distribution of the image is sampled by MCMC methods, such as MYULA and Px-MALA. Then, various forms of uncertainty quantification are performed. Pixel-wise credible intervals (*cf.* error bars) are computed using the posterior samples. Global Bayesian credible regions are computed, again using the posterior samples, and are then used to perform hypothesis testing of image structure to test whether a structure of interest is either physical or an artefact.

where $\texttt{mod}(\cdot,\cdot)$ represents the modulus after division.

We conclude this section by summarising the MYULA and Px-MALA implementations for RI imaging in Algorithms 1 and 2, respectively. Note that symbol $P_{\text{type}} \in \{\texttt{analysis}, \texttt{synthesis}\}$ specifies the problem type considered. Moreover, after obtaining the sets of samples corresponding to the analysis and synthesis models using Algorithms 1 and 2, the posterior mean (or median) of each set of samples can be computed as a point estimator to represent the recovered sky image of interest and thus address the original ill-posed reconstruction problem.

5 BAYESIAN UNCERTAINTY QUANTIFICATION: PROXIMAL MCMC METHODS

In this section we describe a range of uncertainty quantification analyses that are of interest for RI imaging. The analyses require calculating summary statistics w.r.t. the posterior $p(\boldsymbol{x}|\boldsymbol{y})$, which we compute using the samples $\{\boldsymbol{x}^{(j)}\}_{j=1}^K$ generated by MYULA or Px-MALA (in the case of synthesis we generate samples $\{\boldsymbol{a}^{(j)}\}_{j=1}^K$ from $p(\boldsymbol{a}|\boldsymbol{y})$ and map them to the image space by using $\boldsymbol{\Psi}$).

The diagram in Figure 1 shows the main components of our proposed uncertainty quantification methodology based on (proximal) MCMC methods. As is shown, firstly, the full posterior distribution of the image is sampled by MCMC methods, such as MYULA and Px-MALA as adopted in this article. Then, various forms of uncertainty quantification are performed. Firstly, pixel-wise credible intervals are computed using the posterior samples. After that, global Bayesian credible regions are computed, and are then used to perform hypothesis testing of image structure to test whether a structure of interest is either physical or an artefact.

5.1 Pixel-wise credible intervals

The first analyse we consider is the set of marginal credible intervals of each image pixel, denoted by $[\xi_{i-}, \xi_{i+}]$ for pixel x_i. These intervals specify the range of values that the image pixels take with probability $(1-\alpha)$, *i.e.*,

$$p(x_i \in [\xi_{i-}, \xi_{i+}]|\boldsymbol{y}) = 1-\alpha\,, \quad i = 1, \ldots, N\,. \tag{46}$$

Pixel-wise intervals are useful for analysing local information relevant to small image structures and for identifying regions of the image with high uncertainty. For example, these can be conveniently visualised by constructing an image with the quantities $\{\xi_{i+} - \xi_{i-}\}_{i=1}^N$ related to the length of the intervals.

To compute the marginal credible interval we simply calculate:

$$(\bar{\xi}_{i-}, \bar{\xi}_{i+}) = \texttt{quantile}\left(\{x_i{}^{(j)}\}_{j=1}^K, \left\{\frac{\alpha}{2}, 1-\frac{\alpha}{2}\right\}\right), \tag{47}$$

$$(\hat{\xi}_{i-}, \hat{\xi}_{i+}) = \texttt{quantile}\left(\{(\boldsymbol{\Psi}\boldsymbol{a}^{(j)})_i\}_{j=1}^K, \left\{\frac{\alpha}{2}, 1-\frac{\alpha}{2}\right\}\right), \tag{48}$$

depending on whether an analysis or a synthesis formulation is used, respectively; we have used the fact that samples can be marginalised implicitly by projection.

Remark 5.1. Function $\texttt{quantile}(\cdot,\cdot)$ is a standard function built into many programming languages, which, *e.g.*, in (47) computes the quantile thresholds $\bar{\xi}_{i-}$ and $\bar{\xi}_{i+}$ at probabilities $\alpha/2$ and $(1-\alpha/2)$, respectively. In detail, $\bar{\xi}_{i-}$ and $\bar{\xi}_{i+}$ can be computed respectively from the following definitions:

$$\begin{aligned} \bar{\xi}_{i-} &= \inf\{\xi_{i-} : p(z_i \leq \xi_{i-}|\boldsymbol{y}) \geq \alpha/2\}\,, \\ \bar{\xi}_{i+} &= \inf\{\xi_{i+} : p(z_i \leq \xi_{i+}|\boldsymbol{y}) \geq 1-\alpha/2\}\,, \end{aligned} \tag{49}$$

where z_i denotes i-th image pixel in the canonical coordinate system. Refer to, *e.g.*, Koenker & Bassett (1978) for more details about computing quantile thresholds.

5.2 Highest posterior density (HPD) credibility regions

Pixel-wise intervals are useful for analysing local image structures. To perform more sophisticated analyses it is more convenient to compute credible regions that operate at an image level. Precisely, in Bayesian decision theory, a set $C_\alpha \subset \mathbb{R}^N$ with $\alpha \in (0,1)$ is a posterior credible region with confidence level $100(1-\alpha)\%$ if

$$p(\boldsymbol{x} \in C_\alpha|\,\boldsymbol{y}) = \int_{\mathbb{R}^N} p(\boldsymbol{x}|\boldsymbol{y}) \mathbb{1}_{C_\alpha}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 1-\alpha, \tag{50}$$

where $\mathbb{1}_C$ is the indicator function for the set C defined by $\mathbb{1}_C(\boldsymbol{u}) = 1$ if $\boldsymbol{u} \in C$ and 0 otherwise.

There are infinitely many regions C_α that satisfy the above property. The optimal region, in the sense of compactness, is the so-called highest posterior density (HPD) region

$$C_\alpha = \{\boldsymbol{x} : f(\boldsymbol{x}) + g(\boldsymbol{x}) \leq \gamma_\alpha\}, \tag{51}$$

where the threshold γ_α is set such that (50) holds, and we recall that $p(\boldsymbol{x}|\,\boldsymbol{y}) \propto \exp\{-f(\boldsymbol{x}) - g(\boldsymbol{x})\}$. The threshold γ_α defines an isocontour or level-set of the log-posterior. This region is decision-theoretically optimal in the sense of minimum volume (Robert 2007).

The value of γ_α such that (50) and (51) holds is easily estimated from the MCMC samples. Precisely, let $\bar{C}_\alpha$ and $\hat{C}_\alpha$ represent the HPD regions associated with the set of samples $\{\boldsymbol{x}^{(j)}\}_{j=1}^K$ and $\{\boldsymbol{a}^{(j)}\}_{j=1}^K$ generated with MYULA or Px-MALA for the analysis and synthesis models, respectively. To calculate the thresholds $\bar{\gamma}_\alpha$ and $\hat{\gamma}_\alpha$ we use the estimators:

$$\begin{aligned} \bar{\gamma}_\alpha &= \texttt{quantile}\left(\{(\bar{f}+\bar{g})(\boldsymbol{x}^{(j)})\}_{j=1}^K, 1-\alpha\right), \\ \hat{\gamma}_\alpha &= \texttt{quantile}\left(\{(\hat{f}+\hat{g})(\boldsymbol{a}^{(j)})\}_{j=1}^K, 1-\alpha\right). \end{aligned} \tag{52}$$

Notice that C_α is a joint credible region operating at the image level (as opposed to the pixel level), and therefore we use it to analyse larger image structures. In addition, we use C_α for posterior checks to analyse the degree of confidence in specific structure observed in reconstructions, as discussed in the following section.

5.3 Hypothesis testing of image structure

We now describe a *knock-out* posterior check to assess specific areas or structures of interest in reconstructed images. The rationale for this test is that if the data supports a specific feature that we observe in a reconstructed image, *e.g.* $\boldsymbol{x}_{\text{map}}$, then removing this feature from the image is likely to lead to a point that is outside the HPD credible region. Precisely, we use a segmentation-inpainting procedure to carefully replace the feature of interest with background information (although alternative procedures can certainly be considered). If the segmented-inpainted image lies outside of the HPD region this indicates that the likelihood strongly disagrees with the modification, and hence that the data support the feature or structure under consideration. Conversely, if the segmented-inpainted image is within the HPD region, this suggests that the likelihood is not too sensitive to the modification, and therefore that the data does not strongly support the feature or structure being scrutinised.

Algorithmically, the first step of this two-step procedure is to generate a meaningful surrogate test image $\boldsymbol{x}^{*,\text{sgt}}$. We achieve this by taking a point estimator $\boldsymbol{x}^*$ (*e.g.*, the posterior mean $\bar{\boldsymbol{x}}^* = \sum_{j=1}^K \boldsymbol{x}^{(j)}/K$, or $\hat{\boldsymbol{x}}^* = \sum_{j=1}^K \boldsymbol{\Psi}\boldsymbol{a}^{(j)}/K$ if a synthesis model is used) and masking out the structure of interest. This region of the image is then filled by inpainting with background information. Here we use a classical inpaiting approach (Cai et al. 2008) based on a recursive wavelet filter

$$\boldsymbol{x}^{(m+1),\text{sgt}} = \boldsymbol{x}^* \mathbb{1}_{\Omega-\Omega_D} + \boldsymbol{\Lambda}^\dagger \text{soft}_{\lambda_{\text{th}}}(\boldsymbol{\Lambda}\boldsymbol{x}^{(m),\text{sgt}})\mathbb{1}_{\Omega_D}, \qquad (53)$$

where Ω is the image domain, Ω_D is the masked region, Λ is a wavelet filter operator, λ_{th} is a prefixed threshold, and $\boldsymbol{x}^{(m+1),\text{sgt}}$ is the inpainted result obtained at iteration m (generally 100 iterations suffice to achieve convergence). The second step of the procedure is simply to check if $\bar{\boldsymbol{x}}^{*,\text{sgt}} \notin \bar{C}_\alpha$ by using (51) and (52), *i.e.* by evaluating $\bar{f}(\bar{\boldsymbol{x}}^{*,\text{sgt}}) + \bar{g}(\bar{\boldsymbol{x}}^{*,\text{sgt}})$ and comparing to $\bar{\gamma}_\alpha$ (or to check if $\hat{\boldsymbol{x}}^{*,\text{sgt}} \notin \hat{C}_\alpha$ in the synthesis setting).

Finally, note that if the test involves a large structure then the choice of the point estimator used to construct $\boldsymbol{x}^{*,\text{sgt}}$ is usually not important. However, for small structures we recommend using the posterior median as it is closer to the boundaries of C_α than the posterior mean and the MAP estimates.

6 EXPERIMENTAL RESULTS

In this section we demonstrate MYULA and Px-MALA on a range of experiments with simulated RI observations. The generated samples are then used to compute Bayesian point estimators and to perform various forms of uncertainty quantification.

6.1 Simulations

The following four images are used in our experiments: the HI region of the M31 galaxy (size 256×256 pixels) shown in Figure 3 (a); the Cygnus A radio galaxy (size 256×512 pixels) shown in Figure 4 (a, top); the W28 supernova remnant (size 256×256 pixels) shown in Figure 4 (a, middle); and the 3C288 radio galaxy (size 256×256 pixels) shown in Figure 4 (a, bottom). The hardware used to perform these simulations and subsequent numerical experiments is a workstation with 24 CPU cores, x86_64 architecture, and 256 GB memory. All the codes are run on MATLAB R2015b.

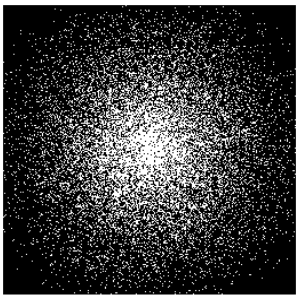
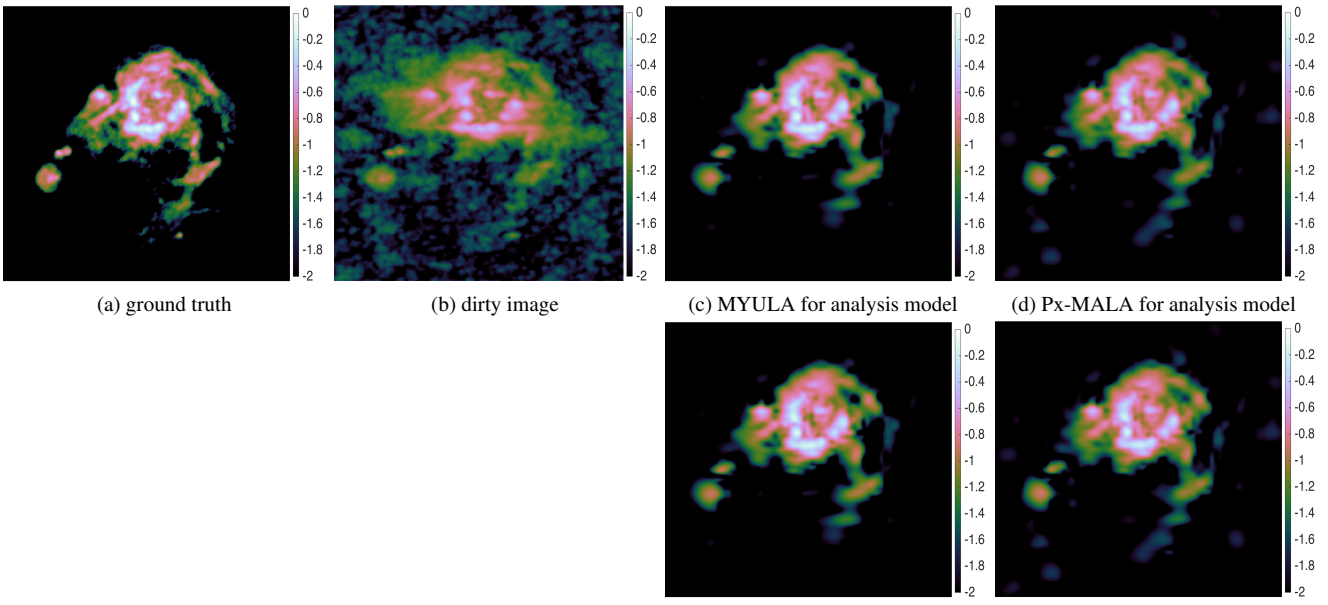


Figure 2. A randomly generated visibility coverage (10% of Fourier coefficients) with size of 256×256.

To generate visibilities, a u-v coverage is generated randomly through the variable density sampling profile (Puy et al. 2011) in half the Fourier plane with 10% of Fourier coefficients of each ground truth image; see Figure 2 for an example of the sampling profile. The visibilities are then corrupted by zero mean complex Gaussian noise with standard deviation σ computed by $\sigma = \|f\|_\infty 10^{-\text{SNR}/20}$, where $\|\cdot\|_\infty$ is the infinity norm (the maximum absolute value of components of f), and SNR (signal to noise ratio) is set to 30 dB for all simulations.

The dictionary $\boldsymbol{\Psi}$ in the analysis and synthesis models (11) and (12) is set to Daubechies 8 wavelets (therefore, we do not expect appreciable difference between the results of the analysis and synthesis models), which is implemented by using the MATLAB built-in function `wavedec2`; complex wavelets or their hybrids, such as those with overcomplete bases, are suggested for better reconstruction. The ℓ_1 regularisation parameter μ in the analysis and synthesis models is fixed to 10^4 by visual cross-validation. Note that, in practice, parameter μ generally needs to be selected carefully either manually or automatically according to some appropriate criterion (see the discussion in Section 2.3). This is beyond the scope of the current article but application of the hierarchical Bayesian strategies developed by Pereyra et al. (2015) will be considered in future work.

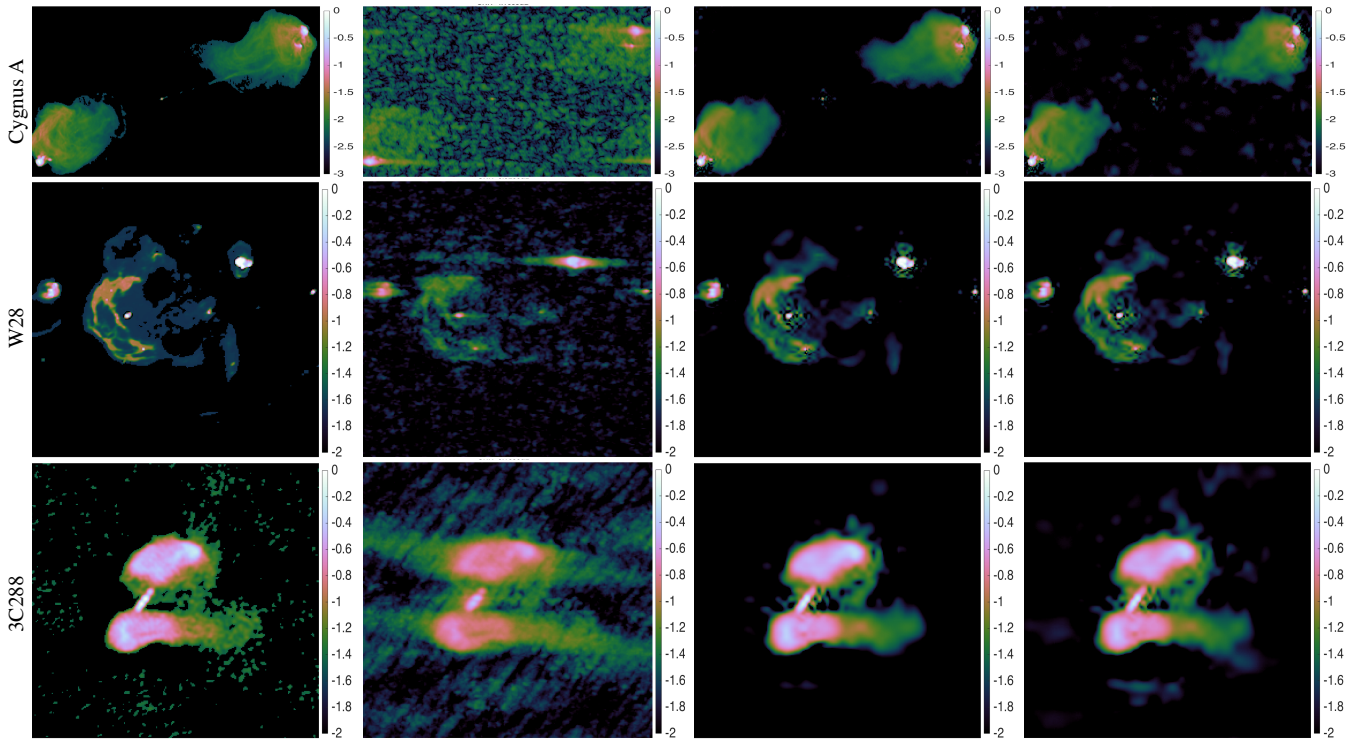
In all experiments MYULA and Px-MALA are implemented using the same algorithm parameters. Precisely, we use each algorithm to generate 10^3 samples from the posterior distributions (7) and (8), with 10^5 burn-in iterations (these iterations correspond to the chains' transient period and are discarded), and a thinning factor of 10^3 iterations between samples (with these settings each algorithm runs for 1.1×10^6 iterations to produce 10^3 samples). We have used these settings to simplify comparisons between MYULA and Px-MALA, however in all our experiments MYULA converged very quickly and could have been implemented with a significantly lower numbers of iterations. The other parameters are set as follows: the maximum iteration number used in (53) for segmented-inpainting is set to 200; the range of values of α in (50) is fixed to [0.01, 0.99]; the credible intervals (47) are computed at level 95% with $\alpha = 0.05$; and α is set to 0.01 (corresponding to the 99% confidence level) in (52) for hypothesis testing.



(a) ground truth (b) dirty image (c) MYULA for analysis model (d) Px-MALA for analysis model

(e) MYULA for synthesis model (f) Px-MALA for synthesis model

Figure 3. Image reconstructions for M31 (size 256×256). All images are shown in $\texttt{log}_{10}$ scale (*i.e.* the numeric labels on the colour bar are the logarithms of the image intensity). Panel (a): ground truth; (b): dirty image (reconstructed by inverse Fourier transform); (c) and (d): point estimators recovered from the mean of the samples generated by MYULA and Px-MALA for the analysis model (11), respectively; (e) and (f): the same as (c) and (d) but for the synthesis model (12). Clearly, consistent results between MYULA and Px-MALA, and between the analysis and synthesis models, are obtained. See further discussion in main text.



(a) ground truth (b) dirty image (c) MYULA for analysis model (d) Px-MALA for analysis model

Figure 4. Image reconstructions for Cygnus A (size 256×512), W28 (size 256×256), and 3C288 (size 256×256) (first to third rows). All images are shown in $\texttt{log}_{10}$ scale. First column: (a) ground truth. Second to forth columns: (b) dirty images, (c) and (d) point estimators for the analysis model (11) using samples generated by MYULA and Px-MALA, respectively. Clearly, consistent results between MYULA and Px-MALA are obtained. See further discussion in main text.

Table 1. CPU time in minutes for MYULA and Px-MALA, for the M31, Cygnus A, W28 and 3C288 experiments, with respect to the analysis and synthesis models (11) and (12). The results show that MYULA is much more economical than Px-MALA, requiring approximately half the computation time of Px-MALA. However, by including an MH (Metropolis-Hastings) accept-reject step Px-MALA removes asymptotic bias.

Images	Methods	CPU time (min)	
		Analysis	Synthesis
M31 (Fig. 3)	MYULA	618	581
	Px-MALA	1307	944
Cygnus A (Fig. 4)	MYULA	1056	942
	Px-MALA	2274	1762
W28 (Fig. 4)	MYULA	646	598
	Px-MALA	1122	879
3C288 (Fig. 4)	MYULA	607	538
	Px-MALA	1144	881

6.2 Image reconstruction

In our first experiment we apply MYULA and Px-MALA to the M31 data and use the samples generated to compute the posterior mean for the synthesis and the analysis models. For comparison, we also report the dirty reconstruction obtained directly via inverse Fourier transform of the visibilities $\boldsymbol{y}$. The dirty image is shown in Figure 3 (b) and compares poorly with the ground truth in Figure 3 (a). The posterior means associated with the models (7) and (8) obtained with MYULA and Px-MALA are displayed in panels (c)–(f). All of these results demonstrate accurate and similar reconstruction performance. In detail, MYULA provides slightly superior reconstruction quality. Moreover, as we can see from Figure 3, the difference between the results with respect to the analysis and synthesis models is negligible (due to an orthogonal basis $\boldsymbol{\Psi}$ being used). Figure 4 shows the results obtained for the Cygnus A, W28, and 3C288 data with the analysis model, observing that these results support the conclusions obtained from the M31 data presented in Figure 3 (results for the synthesis model are not reported here to avoid redundancy because the results are very similar to those of the analysis model).

In summary, both MYULA and Px-MALA perform well for image reconstruction and produce accurate point estimation results. MYULA provides slightly superior reconstruction performance. This is related to the fact that while Px-MALA has more accurate asymptotic properties than MYULA, the superior convergence properties of MYULA mean that it performs better in practice for a fixed number of samples. Furthermore, to generate the same number of samples, MYULA requires approximately half the computation time of Px-MALA; see Table 1 for the CPU time cost in detail.

6.3 Pixel-wise credible intervals

Figure 5 reports the length of the pixel-wise credible intervals (47) for the M31, Cygnus A, W28, and 3C288 data, computed with MYULA and Px-MALA, and for the analysis and the synthesis models (7) and (8). We observe that in this case MYULA delivers significantly better results than Px-MALA; the difference in the estimates illustrates clearly the bias-variance tradeoff related to the MH step in Px-MALA. Precisely, MYULA produces stable smooth estimates with low estimation variance, but which suffer from some estimation bias and overestimates uncertainties as a result. If necessary, this bias can be reduced by decreasing the value of λ. Conversely, the estimates obtained with Px-MALA are unstable and suffer from high estimation variance; however, they do not exhibit a noticeable bias as this is corrected by the MH step. Note that the amount of bias and variance observed are not universal properties of the MYULA and Px-MALA chains. They depend on the quantities that are estimated, and this is why they are visible in the marginal quantiles but not on the posterior means reported in Figure 4.

Furthermore, by inspecting Figure 5 we observe that the pixels close to object boundaries have wider credible intervals than the pixels in homogenous regions. This is related to the fact that there is uncertainty about the high frequency components of the image because of the sampling profile (see Figure 2). Similarly, we observe regular oscillations related to frequencies that are not measured by the sampling profile. Finally, as expected, we note that the analysis and synthesis models produce similar results.

6.4 HPD credibility regions

Figure 6 shows the values of the HPD isocontour threshold γ_α ($\alpha \in [0.01, 0.99]$), defined in (51), computed with MYULA and Px-MALA using (52) for the synthesis and analysis models (red and blue colours are used to represent the results of the analysis and synthesis models, respectively). We observe that the MYULA and Px-MALA estimates are in agreement with each other. Similarly, the analysis and the synthesis models produce similar results. The minor differences in the estimates are again related to the bias-variance tradeoff of Px-MALA (MYULA produces estimates that are larger than Px-MALA but which are also more consistent, whereas Px-MALA estimates have less bias but are also less consistent because of a higher estimation variance).

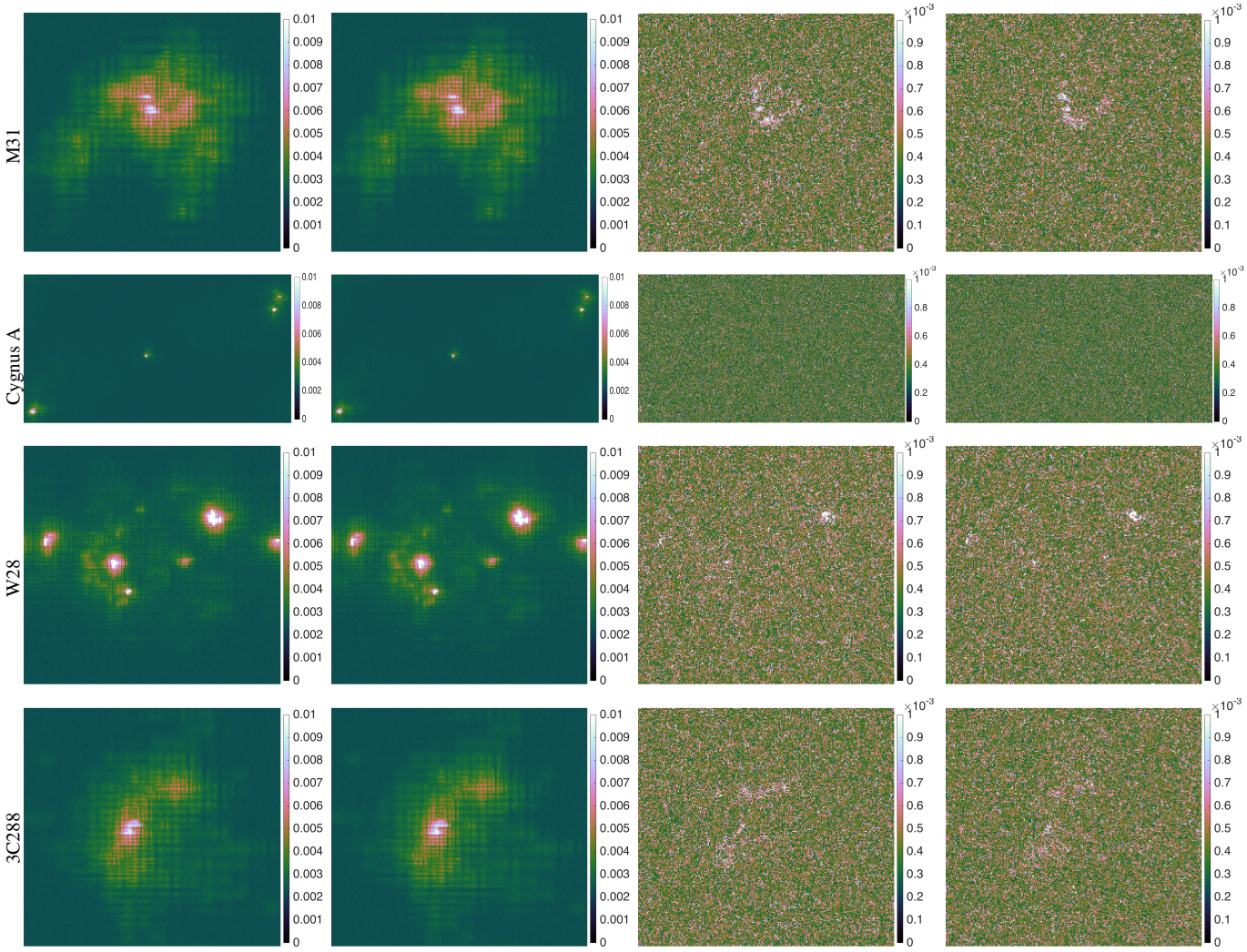
In the following section we use the HDP regions related to Figure 6 to perform uncertainty quantification analyses and posterior checks for specific image structures.

6.5 Hypothesis testing of image structure

We now illustrate our methodology for testing structure in reconstructed images. We consider the five structures depicted in yellow in the first column of Figure 7. All of these structures are physical (*i.e.* present in the ground truth images), while for structure 2 in 3C288 is a reconstruction artefact.

Recall that the methodology proceeds as follows. First, we construct a surrogate test image $\boldsymbol{x}^{*,\mathrm{sgt}}$ by modifying a point estimator (*e.g.*, the sample mean or sample media image) by removing the structure of interest via segmentation-inpaiting (*e.g.*, by using (53), but results are generally not sensitive to the exact method used). Second, we check if $\boldsymbol{x}^{*,\mathrm{sgt}} \notin C_\alpha$ to determine whether there is strong evidence in favour of the structure considered. Conclusions are generally not highly sensitive to the exact value of α; here we report results for $\alpha = 0.01$ related to a 99% credible level.

The results of these experiments are summarised in Table 2 and Table 3, which have been computed by using the posterior mean and the posterior median, respectively, to reconstruct $\hat{\boldsymbol{x}}^{*,\mathrm{sgt}}$. We observe that the same overall conclusions are largely obtained no matter which sampling method is used (MYULA or Px-MALA) or what model is applied (analysis model or synthesis model), indicating that the procedure is robust. Moreover, we observe that the three large physical structures are correctly classified and the reconstruction artefact is correctly highlighted as a structure for which there is lack of evidence. The structure in Cygnus A (see Figure 7)



(a) MYULA, analysis model (b) MYULA, synthesis model (c) Px-MALA, analysis model (d) Px-MALA, synthesis model

Figure 5. Length of pixel-wise credible intervals (95% credible level). First to fourth rows are results for the images M31, Cygnus A, W28, and 3C288, respectively. Columns (a) and (b) are results obtained with samples generated by MYULA using the analysis and synthesis models (11) and (12), respectively; columns (c) and (d) correspond to results obtained with Px-MALA. The results show that MYULA produces wider and smoother credible intervals, compared to those recovered by Px-MALA. See further discussion in main text.

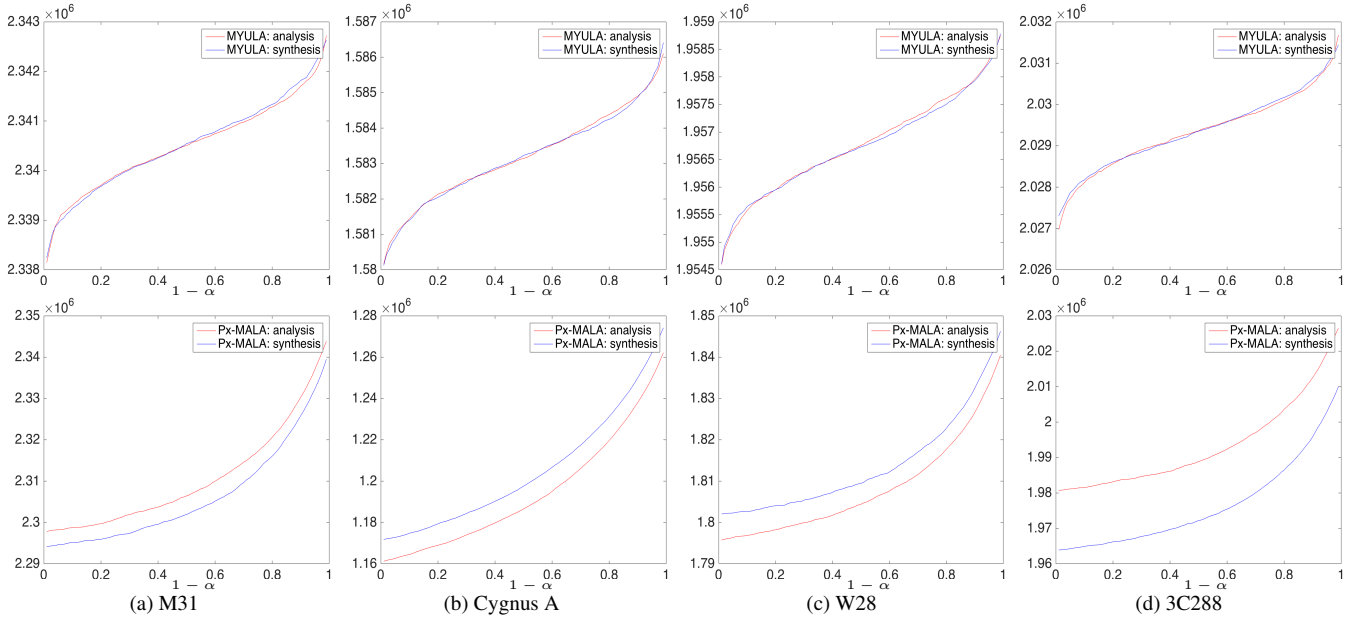
is very small, containing only a few bright pixels that can easily be confused as noise, and it is typically highlighted as potentially non-physical. The only difference between Table 2 and Table 3 is the result of MYULA for the structure of Cygnus A, where the structure is correctly classified as physical when using the posterior median. This is due to the fact that the posterior median is closer to the boundary of C_α and has better sensitivity to small structures as a result. Therefore, we recommend using the median sample for testing. In summary, the proposed methodology, coupled with efficient MCMC sampling by MYULA, provides a powerful framework to perform detailed uncertainty analyses.

To conclude, we emphasise again that the standard methods for RI imaging, such as CLEAN-based methods, MEM and CS-based methods, cannot provide error margins for their solutions, let alone support the detailed uncertainty quantification analyses presented in this article, which includes the calculation of local (pixel-wise) credible intervals, global HPD credible regions, and tests for image structure.

7 CONCLUSIONS

Uncertainty quantification is an important missing component in RI imaging that will only become increasingly important as the big-data era of radio interferometry emerges. No existing RI imaging techniques that are used in practice (*e.g.* CLEAN, MEM or CS approaches) provide uncertainty quantification. Recent techniques that do provide some form of uncertainty information only support restrictive classes of priors (typically Gaussian or log-normal) and do not scale to big-data. While sparsity-promoting priors have shown a great deal of promise for RI imaging (*e.g.* Pratley et al. 2018) and are receiving a great deal of attention, it has not previously been possible to quantify uncertainty information when adopting sparse priors. Traditional MCMC sampling approaches that provide uncertainty information and scale to high dimensional settings, such as RI imaging, often exploit gradient information and cannot support non-differentiable sparse priors. In the current article we solve precisely this problem.

We formulate the RI imaging problem in a Bayesian frame-



(a) M31 (b) Cygnus A (c) W28 (d) 3C288

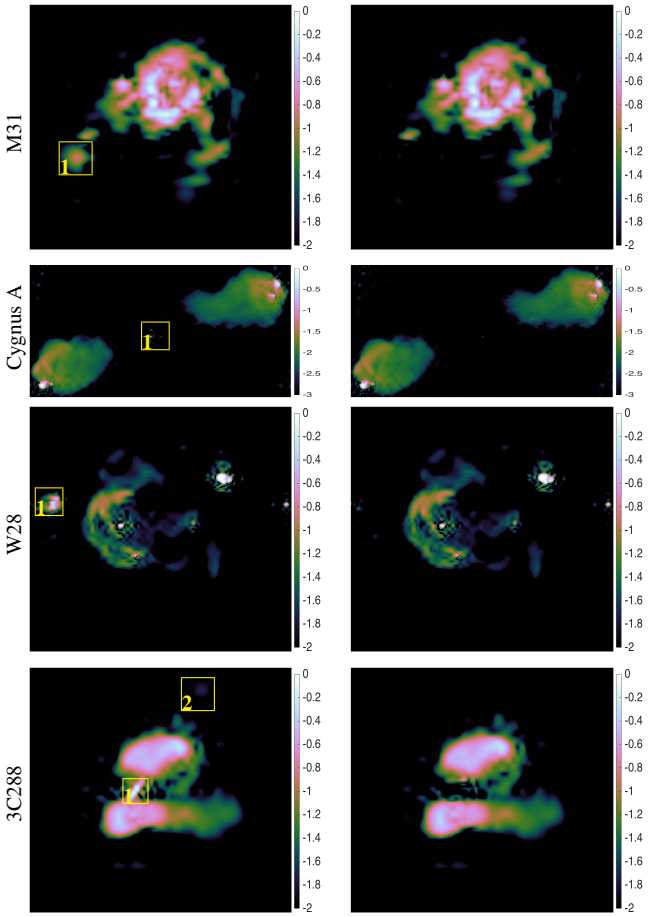
Figure 6. HPD credible region isocontour levels γ_α, computed by MYULA (first row) and Px-MALA (second row), for test images (a) M31, (b) Cygnus A, (c) W28, and (d) 3C288, for the analysis and synthesis models. Clearly, consistent results between Px-MALA and MYULA, and between the analysis and synthesis models, are obtained. Minor differences are discussed in the main text.

Table 2. Hypothesis test results for test structures shown in Figure 7 for M31, Cygnus A, W28, and 3C288. Note that γ_α represents the isocontour defining the HPD credible region at credible level $(1-\alpha)$, where here $\alpha = 0.01$, $\boldsymbol{x}^{*,\mathrm{sgt}}$ represents the surrogate of point estimator $\boldsymbol{x}^*$ (sample mean), and $(f+g)(\cdot)$ represents the objective function; symbols with labels $\bar{}$ and $\hat{}$ are related to the analysis model (11) and the synthesis model (12), respectively. Symbol ✗ indicates that the test area is artificial (and no strong statistical statement can be made as to the area), while ✓ indicates that the test area is physical. All values are in units 10^6. Clearly, MYULA and Px-MALA give convincing and consistent hypothesis test results.

Images	Test areas	Ground truth	Method	$(\bar{f}+\bar{g})(\bar{\boldsymbol{x}}^{*,\mathrm{sgt}})$	Isocontour $\bar{\gamma}_{0.01}$	$(\hat{f}+\hat{g})(\boldsymbol{\Psi}^\dagger\hat{\boldsymbol{x}}^{*,\mathrm{sgt}})$	Isocontour $\hat{\gamma}_{0.01}$	Hypothesis test
M31 (Fig. 7)	1	✓	MYULA	**2.20**	2.34	**2.20**	2.34	✓
			Px-MALA	**2.44**	2.34	**2.43**	2.34	✓
Cygnus A (Fig. 7)	1	✓	MYULA	1.09	**1.59**	1.09	**1.59**	✗
			Px-MALA	1.17	**1.26**	1.18	**1.27**	✗
W28 (Fig. 7)	1	✓	MYULA	**3.43**	1.96	**3.43**	1.96	✓
			Px-MALA	**3.38**	1.84	**3.37**	1.85	✓
3C288 (Fig. 7)	1	✓	MYULA	**3.02**	2.03	**3.02**	2.03	✓
			Px-MALA	**3.27**	2.02	**3.25**	2.01	✓
	2	✗	MYULA	1.752	**2.032**	1.752	**2.031**	✗
			Px-MALA	1.971	**2.027**	1.954	**2.010**	✗

Table 3. Same as Table 2 but based on the sample median instead of the sample mean (the mean is considered for Table 2). This table shows that hypothesis tests based on the median, when using MYULA to generate samples, are able to detect very small structure, such as the test region of Cygnus A.

Images	Test areas	Ground truth	Method	$(\bar{f}+\bar{g})(\bar{\boldsymbol{x}}^{*,\mathrm{sgt}})$	Isocontour $\bar{\gamma}_{0.01}$	$(\hat{f}+\hat{g})(\boldsymbol{\Psi}^\dagger\hat{\boldsymbol{x}}^{*,\mathrm{sgt}})$	Isocontour $\hat{\gamma}_{0.01}$	Hypothesis test
M31 (Fig. 7)	1	✓	MYULA	**2.47**	2.34	**2.48**	2.34	✓
			Px-MALA	**2.46**	2.34	**2.46**	2.34	✓
Cygnus A (Fig. 7)	1	✓	MYULA	**1.597**	1.586	**1.595**	1.586	✓
			Px-MALA	1.205	**1.262**	1.216	**1.274**	✗
W28 (Fig. 7)	1	✓	MYULA	**3.67**	1.96	**3.67**	1.96	✓
			Px-MALA	**3.41**	1.84	**3.39**	1.85	✓
3C288 (Fig. 7)	1	✓	MYULA	**3.30**	2.03	**3.30**	2.03	✓
			Px-MALA	**3.29**	2.02	**3.27**	2.01	✓
	2	✗	MYULA	2.026	**2.032**	2.027	**2.031**	✗
			Px-MALA	1.994	**2.027**	1.977	**2.010**	✗



(a) MYULA point estimators (b) inpainted surrogate

Figure 7. Hypothesis testing for M31, Cygnus A, W28, and 3C288. The five structures depicted in yellow are considered, all of which are physical (*i.e.* present in the ground truth images), except for structure 2 in 3C288, which is a reconstruction artefact. First column (a): point estimators obtained by MYULA for the analysis model (11) (shown in $\texttt{log}_{10}$ scale). Second column (b): segmented-inpainted surrogate test images with information in the yellow rectangular areas removed and replaced by inpainted background (shown in $\texttt{log}_{10}$ scale). Hypothesis testing is then performed to test whether the structure considered is physical by checking whether the surrogate test images shown in (b) fall outside of the HPD credible regions. Results of these hypothesis tests are specified in Table 2 and Table 3. Note that for the case shown in the last row the structures within areas 1 and 2 are tested independently.

work and consider two image models – the analysis and synthesis models – where sparse priors in a suitable signal representation (*e.g.* wavelet basis) are adopted. To perform Bayesian inference for models with sparse priors we consider two innovative MCMC sampling techniques, MYULA and Px-MALA, to sample the full, high-dimensional posterior image distribution. These so-called proximal MCMC techniques exploit proximal calculus to handle non-differentiable prior distributions in high dimensional settings.

Once the full posterior distribution is recovered, a single image is obtained from a point estimator and a variety of methods are presented to perform different types of uncertainty quantification. Pixel-wise credible intervals are computed from the posterior distribution to provide, essentially, error bars for each individual pixel of the recovered image. HPD credible regions are determined for the entire reconstruction, which are then used to perform hypothesis tests of image structure to determine whether the structure is physical or an artefact.

We evaluated our methods on several test images that are representative in RI imaging. Simple simulations of RI observations were performed and Px-MALA and MYULA were used to sample the full image posterior distribution, from which the uncertainty quantification techniques outlined above were applied. Accurate point estimates of recovered images and meaningful uncertainty information were obtained. While Px-MALA is guaranteed to converge to the target distribution, MYULA exhibits an asymptotic bias that can be made arbitrarily small. MYULA, however, does not involve an MH accept-reject step which slows convergence considerably for Px-MALA.

In summary, we develop proximal MCMC techniques to sample the full image posterior distribution for RI imaging for the sparse priors that have been shown in practice to be highly effective. From the posterior distribution a point estimate of the image can be computed and uncertainty information regarding the accuracy of the reconstructed image can be quantified in a variety of ways. These forms of uncertainty quantification provide rich information for analysing RI observations in a statistically robust manner.

In future work the techniques presented here will be extended to consider more complex models, for example with overcomplete dictionaries and for ℓ_p priors with $0 \leq p < 1$, which can provide a stronger sparsity constraint than the ℓ_1 prior. Furthermore, we will investigate optimal techniques for setting the regularisation parameter in a hierarchical Bayesian framework, applying the strategies developed by Pereyra et al. (2015). A more realistic measurement operator that better models real radio interferometry telescopes can be easily incorporated in our framework simply by replacing the measure operator $\mathbf{\Phi}$ adopted.

We have so far considered the telescope calibration parameters to be estimated *a priori* and then fixed. Similarly to μ, one can also consider hierarchical and empirical Bayesian approaches to fix or marginalise calibration parameters. In terms of uncertainty quantification, marginalisation has the advantage of integrating the uncertainty w.r.t. calibration parameters in the analyses, whereas methods that fix calibration parameters neglect this source of uncertainty. We emphasise at this point that performing RI imaging and calibration jointly is a challenging problem because of the dimensionality involved, and this difficulty also extends to uncertainty quantification. Consequently, we leave this problem for future consideration.

For massive data sizes, *e.g.* big-data, like those anticipated from the SKA, it will be difficult if not impossible to apply any MCMC technique due to its inherent computational cost. In the companion article (Cai et al. 2017b) we show how to scale the uncertainty quantification techniques presented in this article to big-data, exploiting recent developments in probability theory and again supporting the sparse priors that have been shown to be so effective in practice.

ACKNOWLEDGEMENTS

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) by grant EP/M011089/1, and Science and Technology Facilities Council (STFC) ST/M00113X/1. We also thank the editor and the anonymous reviewer for their constructive comments, which have significantly improved this manuscript.

REFERENCES

Ables J. G., 1974, A&AS, 15, 383
Bhatnagar S., Corwnell T. J., 2004, A&A, 426, 747
Bhatnagar S., Corwnell T. J., Golap K., Uson J. M., 2008, A&A, 487, 419
Cai J., Chan R., Shen Z., 2008, Appl. Comput. Harmon. Anal., 24, 131
Cai X., Chan R., Zeng T., 2013, SIAM Journal on Imaging Sciences, 6, 368
Cai X., Fitschen J., Nikolova M., Steidl G., Storath M., 2015, Information and Inference: A Journal of the IMA, 4, 43
Cai X., Pratley L., McEwen J. D., 2017a, preprint (arXiv:1712.04462)
Cai X., Pereyra M., McEwen J. D., 2017b, preprint (arXiv:1711.04819)
Candes E. J., Wakin M. B., 2008, IEEE Signal Process. Mag, 25, 21
Candes E. J., Eldar Y. C., Needell D., Randall P., 2010, preprint, (arXiv:1005.2613)
Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223
Carrillo R. E., McEwen J. D., Wiaux Y., 2014, MNRAS, 439, 3591
Chen F., Shen L., Suter B. W., 2016, IET Signal Processing, 10, 557
Cleju N., Jafari M. G., Plumbley M. D., 2012, in Signal Processing Conference (EUSIPCO). pp 869–873
Combettes P. L., Pesquet J. C., 2010, preprint (arXiv:0912.3522v4)
Cornwell T. J., 1988, A&A, 202, 316
Cornwell T. J., 2008, IEEE J. Sel. Topics Signal Process., 2, 793
Cornwell T. J., Evans K. F., 1985, A&A, 143, 77
Cornwell T. J., Golap K., Bhatnagar S., 2008, IEEE J. Sel. Topics Signal Process., 2, 647
Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, A&A, 576, A7
Dabbech A., Wolz L., Pratley L., McEwen J. D., Wiaux Y., 2017, preprint (arXiv:1702.05009)
Donoho D. L., 2006, IEEE Trans. Inf. Theory, 52, 1289
Durmus A., Moulines E., Pereyra M., 2016, preprint (arXiv:1612.07471v1)
Elad M., Milanfar P., Rubinstein R., 2007, Inv. Prob., 23, 947
Enβlin T. A., Frommert M., Kitaura F. S., 2009, Phys. Rev. D, 80
Fadili M. J., Starck J. L., 2009, in ICIP.
Fernandez Vidal A., Pereyra M., 2018, in ICIP.
Garsden H., et al., 2015, A&A, 575, A90
Golub G. H., Hansen P. C., O'Leary D. P., 1999, SIMAX, 21, 185
Green P. J., Łatuszyński K., Pereyra M., Robert C. P., 2015, Statistics and Computing, 25, 835
Greiner M., Vacca V., Junklewitz H., Enßlin T. A., 2017, preprint (arXiv:1605.04317v2)
Gull S. F., Daniell G. J., 1978, Nature, 272, 686
Högbom J. A., 1974, A&AS, 15, 417
Jacques L., Hammond D., Fadili M., 2011, IEEE Trans. Inf. Theory, 57, 559
Junklewitz H., Bell M. R., Selig M., Enßlin T. A., 2016, A&A, 586, A76
Kartik S. V., Carrillo R. E., Thiran J.-P., Y. W., 2017, Springer, New-York, 468, 2382
Koenker R., Bassett G., 1978, Econometrica, 46, 33
Li F., Cornwell T. J., de Hoog F., 2011a, A&A, 528, A31
Li F., Brown S., Cornwell T. J., de Hoog F., 2011b, A&A, 531, A126
Maisinger K., Hobson M. P., Lasenby A. N., 2004, MNRAS, 347, 339
McEwen J. D., Scaife A. M. M., 2008, MNRAS, 389, 1163
McEwen J. D., Wiaux Y., 2011, MNRAS, 413, 1318
Neal R., 2012, preprint (arXiv:1206.1901)
Nikolova M., 2016, Appl. Comput. Harmon. Anal., 41, 237
Offringa A. R., McKinley B., Hurley-Walker N., Briggs F. H., Wayth R. B., Kaplan D. L., et al. 2014, MNRAS, 444, 606
Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J. P., Pesquet J. C., Wiaux Y., 2016, MNRAS, 462, 4314
Onose A., Dabbech A., Wiaux Y., 2017, MNRAS, 469, 938
Parikh N., Boyd S., 2014, Foundations and Trends in Optimization, 1, 123
Pereyra M., 2016a, preprint, (arXiv:1612.06149)
Pereyra M., 2016b, Statistics and Computing, 26, 745
Pereyra M., Bioucas-Dias J., Figueiredo M., 2015, Signal Processing Conference (EUSIPCO)
Pereyra M., Schniter P., Chouzenoux E., Pesquet J., Tourneret J., Hero A., McLaughlin S., 2016, J. Sel. Topics Signal Process., 10, 224
Pratley L., McEwen J. D., d'Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2018, MNRAS, 473, 1038
Puy G., Vandergheynst P., Wiaux Y., 2011, IEEE Signal Processing Letters, 18, 595
Rau U., Bhatnagar S., Voronkov M. A., Cornwell T. J., 2009, Proc. IEEE, 97, 1472
Robert C. P., 2007, The Bayesian Choice. Springer-Verlag New York
Robert C. P., Casella G., 2004, Monte Carlo Statistical Methods. Springer-Verlag New York
Roberts G. O., Tweedie R. L., 1996, Bernoulli, 2, 341
Ryle M., Hewish A., 1960, MNRAS, 120, 220
Ryle M., Vonberg D. D., 1946, Nature, 158, 339
Skilling J., Gull S. F., 1991, Institute of Mathematical Statistics, 20, 341
Starck J. L., Murtagh F., Querre P., Bonnarel F., 2001, A&A, 368, 730
Stewart I. M., Fenech D. M., Muxlow T. W. B., 2011, A&A, 535, A81
Suksmono A. B., 2009, Electrical Engineering and Informatics, 1, 110
Sutter P. M., et al., 2014, MNRAS, 438, 768
Thompson A., Moran J., Swenson G., 2017, Interferometry and Synthesis in Radio Astronomy. Springer International Publishing
Wenger S., Magnor M., Pihlström Y., Bhatnagar S., Rau U., 2010, Electrical Engineering and Informatics, 122, 1367
Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vandergheynst P., 2009a, MNRAS, 395, 1733
Wiaux Y., Puy G., Boursier Y., Vandergheynst P., 2009b, MNRAS, 400, 1029
Wolz L., McEwen J. D., Abdalla F. B., Carrillo R. E., Wiaux Y., 2013, MNRAS, 436, 1993