Posterior sampling for inverse imaging problems on the sphere in seismology and cosmology

Augustin Marignier [1,2]★ Jason D. McEwen,[1] Ana M. G. Ferreira[2,3] and Thomas D. Kitching[1]

[1]*Mullard Space Science Laboratory, University College London, Dorking RH5 6NT, UK*
[2]*Department of Earth Sciences, University College London, London WC1E 6BT, UK*
[3]*CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon 1049-001, Portugal*

ABSTRACT

In this work, we describe a framework for solving spherical inverse imaging problems using posterior sampling for full uncertainty quantification. Inverse imaging problems defined on the sphere arise in many fields, including seismology and cosmology, where images are defined on the globe and the cosmic sphere, and are generally high-dimensional and computationally expensive. As a result, sampling the posterior distribution of spherical imaging problems is a challenging task. Our framework leverages a proximal Markov chain Monte Carlo (MCMC) algorithm to efficiently sample the high-dimensional space of spherical images with a sparsity-promoting wavelet prior. We detail the modifications needed for the algorithm to be applied to spherical problems, and give special consideration to the crucial forward modelling step, which contains computationally expensive spherical harmonic transforms. By sampling the posterior, our framework allows for full and flexible uncertainty quantification, something which is not possible with other methods based on, for example, convex optimization. We demonstrate our framework in practice on full-sky cosmological mass-mapping and to the construction of phase velocity maps in global seismic tomography. We find that our approach is potentially useful at moderate resolutions, such as those of interest in seismology. However, at high resolutions, such as those required for astrophysical applications, the poor scaling of the complexity of spherical harmonic transforms severely limits our method, which may be resolved with future GPU implementations. A new Python package, pxmcmc, containing the proximal MCMC sampler, measurement operators, wavelet transforms, and sparse priors are made publicly available.

Key words: Data Methods – Bayesian – Mass-Mapping – Seismic Tomography.

1 INTRODUCTION

Inverse problems on the sphere are common in many fields, from astrophysics (e.g. Planck Collaboration I 2016; Jeffrey et al. 2021), to geophysics (e.g. Ritsema et al. 2011; Chang et al. 2015) and more. These tend to be high-dimensional and computationally challenging imaging problems, increasingly so at high resolutions. Spherical inverse problems are difficult to solve using posterior sampling methods such as Markov chain Monte Carlo (MCMC), largely due to the generally large number of parameters to sample and the high computational cost of repeated evaluations of the forward problem on the sphere. MCMC methods have grown in popularity in recent decades (e.g. Mosegaard & Tarantola 1995; Lewis & Bridle 2002; Malinverno 2002; Corless, King & Clowe 2009; Bodin et al. 2012; Cai, Pereyra & McEwen 2018a), benefitting from their ability to sample the full posterior probability density function (PDF), which constitutes the solution of the inverse problem. This allows for flexible calculation of any measure of uncertainty in the solution. Furthermore, they can be used to solve non-linear inverse problems, which are commonplace in, for example, geophysics (Malinverno 2002; Bodin et al. 2012; Ferreira et al. 2020). However, sampling methods come at significant computational cost, for both solving

the inverse problem and for model comparison, which is often computationally infeasible. The simpler MCMC methods, such as the Metropolis–Hastings (MH) algorithm, are known to struggle in high-dimensional parameter spaces (e.g. Roberts & Rosenthal 1998; Neal 2012), such as those for spherical imaging problems. The exponential increase in volume of the parameter space makes MH unable to converge to a solution. Alternative gradient-based methods such as the Hamiltonian Monte Carlo (HMC) (Duane et al. 1987; Neal 2012) and the unadjusted Langevin algorithm (ULA) (Roberts & Tweedie 1996) scale favourably compared to the MH algorithm as the number of dimensions increases (Roberts & Rosenthal 1998; Neal 2012). The caveat here is that these methods can only be applied to smooth distributions, limiting the form of prior information that can be used as regularization. Conversely, nested sampling methods (Skilling 2006) can be used for non-smooth problems. However, most nested sampling approaches cannot scale to the high-dimensional settings of inverse problems without gradient information. Proximal MCMC methods have recently been developed (Pereyra 2016), leveraging proximal mappings to scale to high dimensions and allow for non-smooth distributions. Proximal mappings have also recently been used to scale nested sampling methods to high-dimensional imaging problems (Cai, McEwen & Pereyra 2022).

Compressed sensing (Donoho 2006; Candès et al. 2011) has demonstrated that sparse signals can be accurately recovered from

★ E-mail: augustin.marignier.14@ucl.ac.uk

incomplete data, recovering both sharp and smooth image features simultaneously from underdetermined systems (Loris et al. 2007; Wallis, Wiaux & Mcewen 2017). As a result, sparse priors have been widely adopted for solving inverse imaging problems, e.g. for radio interferometry applications (Wiaux et al. 2009; Carrillo, McEwen & Wiaux 2014; Pratley et al. 2017; Cai et al. 2018a; Cai, Pereyra & McEwen 2018b), cosmological mass-mapping (Lanusse et al. 2016; Price et al. 2020b), and have received some attention in seismic tomography (Loris et al. 2007; Charléty et al. 2013). The prior PDF (i.e. the distribution representing *a prior* beliefs to constrain the inverse problem) often used to promote sparsity is the Laplace distribution, which is non-differentiable and so cannot be used with gradient-based sampling algorithms.

Current approaches to uncertainty quantification for high-dimensional inverse problems with sparse a priori constraints exploit results from information theory to derive approximate credible regions from a single point estimate solution of the inverse problem (Pereyra 2017; Cai et al. 2018b; Price et al. 2020b). The point estimate solution can be obtained using convex optimisation (Cai et al. 2018b; Price et al. 2020b), and thus this approach is much faster than posterior sampling, particularly for high-resolution spherical problems. However, this information theory-based approach can only approximate credible regions. A sampling method, on the other hand, will have the flexibility to calculate any measure of uncertainty, as the samples represent a full probability distribution.

In this work, we provide a framework for solving inverse imaging problems on the sphere using posterior sampling with sparsity-promoting priors. For this, we leverage a proximal MCMC algorithm (Pereyra 2016), which is a sampling scheme that uses proximal mappings to efficiently sample high-dimensional parameter spaces. Proximal mappings can be viewed as more general gradient operators that can be used even on non-differentiable functions, allowing us to use a Laplace prior to promote sparsity. In our framework, we outline the modifications needed to the proximal MCMC algorithm for spherical problems, with particular consideration given to the spherical parameterization and forward operator. We then demonstrate our framework in practice first on a common problem from the field of global seismic tomography using both synthetic and real data, and then on a low-resolution full-sky cosmological mass-mapping example from simulation data.

This paper is structured as follows. Section 2 gives the necessary mathematical background for Bayesian inference and representations of spherical images. Section 3 outlines our framework for posterior sampling of inverse problems on the sphere, with details of the proximal MCMC algorithm that we use. Sections 4 and 5 contain our illustrative examples from global seismic tomography and cosmological mass-mapping, respectively, and we conclude in Section 6.

2 MATHEMATICAL BACKGROUND

In this section, we provide the necessary mathematical background for this work, including Bayesian inference for imaging, and harmonic and wavelet representations of spherical images.

2.1 Bayesian inference for imaging

Consider some observed data d and model parameters m that are related by some general, possibly non-linear, forward operator G, as

$$d = G(m) + n, \tag{1}$$

where n represents observational noise. The aim of the inverse problem is to infer m from d. This can be formulated in a Bayesian

statistical framework by Bayes Theorem as

$$p(m|d) \propto p(d|m)p(m). \tag{2}$$

Here $p(m|d)$ is known as the posterior PDF that encapsulates knowledge of the model parameters m given data d and represents the solution to the inverse problem (Tarantola and Valette 1982; Mosegaard and Tarantola 1995). The likelihood function $p(d|m)$, assuming independent and identically distributed Gaussian noise with known standard deviation σ in the data, is given by

$$p(d|m) \propto \exp\left(-\frac{\|d - G(m)\|_2^2}{2\sigma^2}\right). \tag{3}$$

Inverse problems are typically ill-posed and often require some form of regularization in the form of prior knowledge about the model parameters to further constrain the inverse problem. In the Bayesian framework, this is represented by the prior PDF $p(m)$, which can take many forms. Common examples include the Uniform prior (Jaynes 2003), $p(m) \propto$ constant; the Gaussian prior (Tarantola & Valette 1982; Golub, Hansen & O'Leary 1999), $p(m) \propto \exp(-\mu\|m\|_2^2)$; and the Laplace prior (Donoho 2006; Candès et al. 2011), $p(m) \propto \exp(-\mu\|m\|_1)$, to promote sparsity. In these priors, $\|\cdot\|_p$ is the l_p-norm and μ is a parameter that captures the width of the distribution. Henceforth, we use the Laplace prior as we wish to promote sparsity. The constant of proportionality in equation 2 is known as the Bayesian evidence. This is useful for comparing the physical model (G) to alternative hypotheses, but may be ignored for inference about model parameters m.

Solving the inverse problem by probabilistic sampling involves sampling the posterior PDF $p(m|d)$, most commonly using MCMC algorithms. The simplest is the MH algorithm. MH proposes a new sample from a proposal distribution, which has a certain probability of being accepted. More sophisticated and better suited to high-dimensional problems than MH are gradient-based methods, such as the HMC and Langevin algorithms mentioned previously, which use the gradients of the target PDF to more efficiently guide the exploration of the parameter space. While the set of samples from the posterior PDF represents the full solution of the inverse problem, results are generally reported in terms of summary statistics, such as the mean of the samples. The *maximum a posteriori* (MAP) solution,

$$m^{(\text{MAP})} = \underset{m}{\operatorname{argmax}}\, p(m|d) \tag{4}$$

is also a common choice; however, due to the randomness of MCMC there is no guarantee of the MAP being found. The highest posterior probability sample of the MCMC provides the closest estimate of the MAP, but can still be relatively far from the true MAP, particularly in higher dimensional spaces. Additionally, some measure of uncertainty on individual model parameters can be calculated from the posterior samples. We give further details of this in the next section.

In inverse imaging problems, the model parameters m represent a 2D (or possibly 3D) image. MCMC can be used to sample either the direct pixel values x of the image, or the coefficients α representing the image in a particular basis. The forward operator G in equation 1 can be described as a (possibly non-linear) measurement operator Φ acting on the image x. In many physical situations, the image x has a sparse representation in some basis, which can be exploited to regularize the inverse problem. For some sparsifying basis Ψ, we have $x = \Psi\alpha$, where α is the set of coefficients representing x in the basis encoded in Ψ. The model parameters m sampled by MCMC are the coefficients α, and as such the forward operator is given by

$$G(\alpha) = \Phi(\Psi\alpha). \tag{5}$$

For non-linear inverse problems, i.e. where $\mathbf{\Phi}$ is a non-linear measurement operator, a common choice is to linearize via a Taylor expansion around an initial guess model. While this usually leads to acceptable results in weakly non-linear problems, linearization errors and the need for a good initial guess model can lead to artefacts in the recovered model. If the forward problem (equation 5) is computationally fast, sampling methods are typically best for strongly non-linear problems.

2.2 Harmonic representations of spherical images

We consider functions $f(\theta, \phi)$ defined on the sphere $\mathbb{S}^2$ with colatitude $\theta \in [0, \pi]$ and longitude $\phi \in [0, 2\pi]$. The canonical orthogonal basis functions for $\mathbb{S}^2$ are the spherical harmonics

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2\ell + 1}{4\pi} \frac{(\ell - m)!}{(\ell + m)!}} P_\ell^m(\cos\theta) e^{im\phi} \qquad (6)$$

for non-negative integers ℓ and integers $m \leq |\ell|$, denoting angular degree and angular order, respectively. As per the Condon-Shortley phase convention, the associated Legendre functions $P_\ell^m(x)$ include a $(-1)^m$ phase factor, ensuring the conjugate symmetry relation

$$Y_{\ell, -m}(\theta, \phi) = (-1)^m Y_{\ell m}^*(\theta, \phi), \qquad (7)$$

where * denotes complex conjugation. Any square integrable scalar function on $\mathbb{S}^2$ can be represented in the frequency domain by its harmonic coefficients $f_{lm} \in \mathbb{C}$, obtained by projecting f onto the spherical harmonic basis functions using the inner product (forward spherical harmonic transform)

$$f_{lm} = \langle f, Y_{\ell m} \rangle = \int_{\mathbb{S}^2} f(\theta, \phi) Y_{\ell m}^*(\theta, \phi) \sin\theta \mathrm{d}\theta \mathrm{d}\phi. \qquad (8)$$

The function $f(\theta, \phi)$ can be recovered exactly from its spherical harmonic coefficients (inverse spherical harmonic transform) by

$$f(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{\ell m} Y_{\ell m}(\theta, \phi). \qquad (9)$$

This formalism can be generalized to spin-s fields (Newman & Penrose 1966; Goldberg et al. 1967), common to astrophysical spherical signals (McEwen et al. 2015a; Wallis et al. 2022), which are characterized by local rotations $\chi \in [0, 2\pi]$ in the tangent plane at a point (θ, ϕ),

$$_s f'(\theta, \phi) = e^{-is\chi} {}_s f(\theta, \phi) \qquad (10)$$

The spin-weight s of a function can be increased or decreased by applying the spin-raising or spin-lowering operators, $\eth$, $\bar{\eth}$, respectively.

$$\eth \equiv -\sin^s\theta \left(\frac{\partial}{\partial\theta} + \frac{i}{\sin\theta} \frac{\partial}{\partial\phi} \right) \sin^{-s}\theta, \qquad (11)$$

$$\bar{\eth} \equiv -\sin^{-s}\theta \left(\frac{\partial}{\partial\theta} - \frac{i}{\sin\theta} \frac{\partial}{\partial\phi} \right) \sin^s\theta. \qquad (12)$$

Note how these operators are defined based on the spin of the function to which they are applied. Using these operators, one can obtain the spin-weighted spherical harmonics from the original scalar (spin-0) spherical harmonics, which form an orthogonal basis for spin functions on the sphere

$$_s Y_{\ell m}(\theta, \phi) = \begin{cases} 0, & \ell < |s| \\ \sqrt{\frac{(\ell-s)!}{(\ell+s)!}} \eth^s Y_{\ell m}(\theta, \phi), & 0 \leq s \leq \ell \\ \sqrt{\frac{(\ell+s)!}{(\ell-s)!}} (-1)^s \bar{\eth}^{-s} Y_{\ell m}(\theta, \phi), & -\ell \leq s \leq 0 \end{cases} \qquad (13)$$

As such, spin-s fields can easily be decomposed and reconstructed in terms of spin spherical harmonics in a manner similar to equations 8 and 9.

In this work, we consider bandlimited signals, that is signals such that $_s f_{\ell m} = 0, \forall \ell \geq L$ for some bandlimit L. We discretize spherical signals according to the McEwen-Wiaux (MW) sampling theorem (McEwen & Wiaux 2011). This equiangular sampling theorem has theoretically exact and efficient spherical harmonic transforms and requires fewer discrete points on the sphere than other sampling theorems (e.g. Driscoll & Healy 1994; Górski et al. 2005). It also exploits a relationship between spin spherical harmonics and Wigner functions (Goldberg et al. 1967; McEwen & Wiaux 2011) to avoid repeated applications of spin-raising/lowering operators for fast spin spherical harmonic transforms.

2.3 Wavelet representations of spherical images

For our sparsifying basis, we consider scale-discretized axisymmetric wavelets (Wiaux et al. 2008; Leistedt et al. 2013; McEwen, Durastanti & Wiaux 2018), i.e. the wavelets are azimuthally symmetric when placed at the poles. There exist directional wavelets which vary azimuthally (e.g. Wiaux et al. 2008; McEwen et al. 2015a, 2018) and may be desirable in certain applications. However, such wavelets require transforms on the rotation group SO(3), which are more computationally expensive than transforms on $\mathbb{S}^2$ (McEwen et al. 2015b). The spherical wavelet transform is defined as the convolution of f and the wavelets $\Psi^j(\theta, \phi)$. Convolution on the sphere is defined as the inner product of f with wavelets that have been rotated over the surface of the sphere by some operator $\mathcal{R}_{(\theta, \phi)}$, analogous to standard convolution on the plane where the wavelet is laterally translated over the image. As such, the wavelet coefficients $W^{\Psi^j} \in \mathbb{S}^2$ are given by

$$W^{\Psi^j}(\theta, \phi) = \langle f, \mathcal{R}_{(\theta, \phi)} \Psi^j \rangle. \qquad (14)$$

The wavelets Ψ^j cover a range of scales $J_0 \leq j \leq J$, which extract the highest frequency (high ℓ information of f). The lowest frequency information (low ℓ) is extracted by a scaling function $\Upsilon(\theta, \phi)$ in a similar manner:

$$W^{\Upsilon}(\theta, \phi) = \langle f, \mathcal{R}_{(\theta, \phi)} \Upsilon \rangle. \qquad (15)$$

The wavelets and scaling function are defined on the harmonic line such that the wavelets can be seen as narrow pass-band filters and the scaling function is a low-pass filter that includes the zero frequency component. For more details on choosing the width of the wavelets on the harmonic line, we refer the interested reader to Leistedt et al. (2013). The wavelets and scaling function are defined such that they satisfy an admissibility condition, which allows f to be decomposed and recovered exactly from its wavelet coefficients. The reconstruction is given by

$$f(\theta, \phi) = \sum_{\Gamma} \int_{\mathbb{S}^2} W^{\Gamma}(\theta', \phi') (\mathcal{R}_{(\theta', \phi')} \Gamma)(\theta, \phi) \sin\theta' \mathrm{d}\theta' \mathrm{d}\phi', \qquad (16)$$

where $\Gamma \in \{\Upsilon, \Psi^j\}$. Note that in practice, the spherical wavelet transform is computed in harmonic space (Leistedt et al. 2013), meaning the forward and inverse wavelet transforms (equations 14–16) implicitly involve both a forward and an inverse spherical harmonic transform (Wallis et al. 2017), which dominate the computational complexity. The spherical harmonic and axisymmetric

wavelet transforms used in this work are implemented in the packages PYSSHT[1] and PYS2LET,[2] respectively.

3 FRAMEWORK FOR POSTERIOR SAMPLING ON THE SPHERE

In this section, we outline our framework for sampling the posterior of spherical inverse problems. We give the details of the high-dimensional sampling algorithm we use and the modifications necessary for the sphere. Extra consideration is given to the forward operator, and finally we define a measure of uncertainty that can be calculated from the posterior samples.

3.1 Proximal MCMC

3.1.1 Moreau–Yosida envelope and proximal mapping

The λ-Moreau–Yosida envelope of some concave function $h(x)$ is given by (Moreau 1962; Bauschke & Combettes 2017)

$$h^\lambda(x) = \min_u \{h(u) + \|u - x\|_2^2/(2\lambda)\}, \tag{17}$$

where $\lambda > 0$. This is a smooth approximation of h with many desirable properties. First, h^λ can be made arbitrarily close to h by making λ small. Secondly, the minimizers of h^λ are the same as the minimizers of h. Thirdly, h^λ is continuously differentiable even if h is not. The gradient of h^λ is given by

$$\nabla h^\lambda(x) = [(x - \text{prox}_h^\lambda(x))]/\lambda, \tag{18}$$

where $\text{prox}_h^\lambda(x)$ is the proximal mapping of h (Moreau 1962)

$$\text{prox}_h^\lambda(x) = \underset{u}{\text{argmin}}\{h(u) + \|u - x\|_2^2/(2\lambda)\}. \tag{19}$$

Rewriting equation (18) as

$$\text{prox}_h^\lambda(x) = x - \lambda \nabla h^\lambda(x) \tag{20}$$

is reminiscent of the standard first-order finite differences approximation $h(x_0 + a) \approx h(x_0) + ah'(x_0)$, which shows that prox_h^λ may be interpreted as a gradient step in h^λ with step size λ. The key property here is that the proximal map can be used to minimise h^λ and, by extension, the non-differentiable h since they share the same minimizers (Parikh & Boyd 2014). For an excellent introduction to proximal mappings and algorithms, see Parikh & Boyd (2014) and Combettes & Pesquet (2011).

3.1.2 Proximal Langevin algorithm

The proximal MCMC method developed by Pereyra (2016) is based on Langevin MCMC (Roberts & Tweedie 1996), a gradient-based sampling method, which we describe here. Consider a Langevin diffusion process $Y(t)$ for $0 \leq t \leq T$ associated with a stationary PDF π. This process is given by the stochastic differential equation

$$dY(t) = \frac{1}{2}\nabla \log \pi[Y(t)]dt + dW(t) \tag{21}$$

for some Brownian motion W. We use the forward Euler discrete time approximation with step size δ

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

[1]https://github.com/astro-informatics/ssht
[2]https://github.com/astro-informatics/s2let

where m is the discretized Langevin diffusion and $w^{(i)} \sim \mathcal{N}(0, \mathbb{I})$. This is the ULA (Roberts & Tweedie 1996). Under certain regularity conditions, ULA produces samples that converge to an ergodic measure close to π. A MH acceptance step can be added to remove the approximation error, giving the Metropolis-Adjusted Langevin Algorithm (Roberts & Tweedie 1996).

A well-known limitation of the Langevin and other gradient-based sampling algorithms is that they require differentiable PDFs, which is not the case for some popular priors such as the Laplace prior. A proposed solution to this is to apply a Moreau-Yosida approximation to the non-differentiable terms in the posterior PDFs (Pereyra 2016; Cai et al. 2018a; Pereyra, Mieles & Zygalakis 2020). The posterior PDF for our inverse problem (equation 2) is of the form $\pi(m) \propto \exp(-g(m) - f(m))$, where $g(m) = \frac{1}{2\sigma^2}\|d - G(m)\|_2^2$ is our Gaussian data fidelity and $f(m) = \mu\|m\|_1$ is our non-differentiable Laplacian prior. Applying a λ-Moreau-Yosida approximation to the prior, it then follows that the chain step for the Moreau-Yosida ULA (MYULA) is given by

$$m^{(i+1)} = \left(1 - \frac{\delta}{\lambda}\right)m^{(i)} + \frac{\delta}{\lambda}\text{prox}_f^\lambda[m^{(i)}]$$
$$- \delta\nabla g[m^{(i)}] + \sqrt{\delta}w^{(i)}. \tag{23}$$

The tuning parameter δ must be small for the forward Euler approximation (equation 22) to converge, and Pereyra (2016) argued that the optimal value for λ is $\lambda = \delta/2$. We describe how to calculate the proximal mapping of our prior on the sphere (second term equation 23) in the following subsection. If G is linear, the gradient of the data fidelity is straightforward to calculate using the adjoint

$$\nabla g = G^\dagger(Gm - d)/\sigma^2. \tag{24}$$

It is straightforward to modify equation 24 if the data errors are characterized by a covariance matrix C rather than a single standard deviation σ. It is important to note here that for spherical inverse problems, the forward operator G may include spherical harmonic transforms, and thus the adjoint transforms are also needed for equation 24. Further discussion about G in the spherical setting is given later in this section. Algorithm 1 outlines the use of the MYULA chain in practice, highlighting the key steps and equations. Again, an MH acceptance step can be added (Pereyra 2016; Cai et al. 2018a); however, in our experiments there was little improvement for the additional computational cost.

Algorithm 1 MYULA on the sphere

INPUTS: observed data d, data errors σ, initial sample $m^{(0)}$, $i = 0$, N, N_{thin}, N_{burn}, quadrature weights q, δ, λ, μ
OUTPUTS: chain $\{m^{(i)} : i = 1, ..., N\}$
while $i < N \times N_{\text{thin}} + N_{\text{burn}}$ **do**
   Calculate gradient of data fidelity (eq~**??**)
   Calculate proximity map of prior (eq~**??**)
   Calculate $m^{(i+1)}$ (eq~**??**)
   if $i > N_{\text{burn}}$ **then**
      if $\text{mod}(i, N_{\text{thin}}) = 0$ **then**
         Save $m^{(i+1)}$ to chain
      end if
   end if
   $i \mathrel{+}= 1$
end while

3.2 Modification for the sphere

We sample a set of spherical wavelet coefficients that are defined at each point on the sphere (equation 14). The MW sampling theorem is equiangular, and as such, we need to account for an overdensity of sampling points near the poles when we calculate the proximal mapping of our sparsity-promoting prior. Proximal operators are generally calculated by a small convex optimization problem (Parikh & Boyd 2014). Fortunately, there exist closed-form representations for the proximal mapping of many common functions (Combettes & Pesquet 2011), including the ℓ_1-norm we use in our prior. This is crucial for MCMC methods where the proximal mapping needs to be repeatedly computed. It can be shown that for $f(\boldsymbol{m}) = \mu \|\boldsymbol{m}\|_1$,

$$\text{prox}_f^\lambda(\boldsymbol{m}) = \text{soft}_{\lambda\mu}(\boldsymbol{m}) \tag{25}$$

where $\text{soft}_{\lambda\mu}$ is the *soft thresholding* operator with threshold $\lambda\mu$ defined as (Combettes & Pesquet 2011)

$$\text{soft}_{\lambda\mu}(m_i) = \begin{cases} 0 & \text{if } m_i \leq \lambda\mu, \\ m_i - \lambda\mu\,\text{sgn}(m_i) & \text{if } m_i > \lambda\mu \end{cases} \tag{26}$$

where $\text{sgn}(x)$ is the sign of x, and can be applied component-wise to the components m_i of $\boldsymbol{m}$. To account for the spherical sampling overdensity, a weighting must be applied to parameters $\boldsymbol{m}$, which takes the form of a diagonal matrix $\boldsymbol{W}$ of quadrature weights (McEwen & Wiaux 2011) that vary with colatitude and bandlimit. This can be seen as applying a component-wise prior on each of the elements of $\boldsymbol{m}$. The proximal mapping changes only in that the threshold in equation 26 becomes $\lambda\mu w_i$, where w_i is the ith diagonal element of $\boldsymbol{W}$.

3.3 The forward operator

A key consideration for any MCMC sampling algorithm is the speed of the forward modelling step, i.e. making predictions of the data $\boldsymbol{d}$ given an MCMC sample $\boldsymbol{m}$. This is particularly important for spherical problems where the forward operator $\boldsymbol{G}$ may contain spherical harmonic transforms, which are, unfortunately, slow and typically scale in complexity as $\mathcal{O}(L^3)$ (McEwen & Wiaux 2011). As discussed in Section 2, the spherical harmonic transforms are implicit in the spherical wavelet transforms which form our sparsifying basis. Expanding equation (5) to see this, to sample the spherical wavelet coefficients in pixel space $\boldsymbol{\alpha}$ we have in practice

$$G(\boldsymbol{\alpha}) = \boldsymbol{\Phi}\boldsymbol{S}^{-1}\boldsymbol{W}\boldsymbol{S}\boldsymbol{\alpha}, \tag{27}$$

where $\boldsymbol{S}$ and $\boldsymbol{S}^{-1}$ are the forward and inverse spherical harmonic transforms (equations 8 and 9), respectively, and $\boldsymbol{W}$ is the spherical wavelet transform in harmonic space (Leistedt et al. 2013; Wallis et al. 2017). This is further complicated by the need for the adjoints of these transforms (equation 24). These four spherical harmonic transforms at each MCMC step immediately limits the bandlimits L for which sampling the posterior may be feasible. For reference, in our experiments, we performed dummy inversions at $L = 64$ and 128 with an identity measurement operator ($\boldsymbol{\Phi} = \boldsymbol{I}$) which took around 3 and 17 d, respectively, for 10^6 samples on a 2.5-GHz Intel Xeon Platinum 8180M processor. While this may not be an issue for applications where information exists at relatively low degrees, for example in seismic tomography where maximum bandlimits are typically around $L = 40$ (e.g. Ritsema et al. 2011; Chang et al. 2015), it will be impractical for applications where the bandlimits of interest are higher (e.g. Price et al. 2020b). To avoid spherical harmonic transforms in equation 27, one could instead sample the spherical harmonic coefficients of $\boldsymbol{\alpha}$, provided one can reformulate

the measurement operator $\boldsymbol{\Phi}$ appropriately. We note, however, that it is possible that, as shown in the following section, the harmonic measurement operator may be slower than the pixel space operator even with spherical wavelet transforms. Further, it is conceivable that, depending on how the measurement operator scales, the choice of harmonic or pixel space will depend on the bandlimit L. In either case, at least one spherical harmonic transform will be needed for the prior, as there is no reason to expect the harmonic representation of the wavelet coefficients to also be sparse. As such, spherical harmonic transforms are unavoidable, and it is crucial that these are computed as efficiently as possible, for example via repeated exploitation of fast Fourier transforms as in McEwen & Wiaux (2011). This all highlights the special consideration that must be given to the forward operator for spherical inverse problems.

3.4 Spherical wavelet parameterization

As discussed in Section 2, our inverse problem is parameterized using axisymmetric spherical wavelets (Wiaux et al. 2008; Leistedt et al. 2013; McEwen et al. 2015a) to promote sparsity. We can exploit the multiresolution character of these wavelets for further computational savings. By construction, the wavelets at each scale j each have different bandlimits $k_j = B^{j+1} \leq L$, where B is a wavelet scale parameter (Leistedt et al. 2013). By using a sampling theorem, the transforms at each spherical wavelet scale can be performed up to their own bandlimit k_j, and only the minimum number of samples on the sphere at that bandlimit are needed. This multiresolution transform gives a four to five times speed up of the spherical wavelet transforms (Leistedt et al. 2013), and also dramatically reduces the dimensionality of our parameter space, compared to a full resolution transform where each wavelet scale is sampled at the overall bandlimit L. As an example, for parameters $L = 32$, $B = 2$, $J_0 = 2$, the full resolution algorithm has 10 080 wavelet coefficients, compared to only 4676 for the multiresolution algorithm. While a full assessment of how this affects the convergence speed of the MCMC chain in terms of number of required steps is beyond the scope of this paper, we found real-time speed-ups and significant memory savings in our experiments.

3.5 Uncertainty quantification

By collecting samples from the posterior, we can calculate any measure of uncertainty. For example, a common choice in Bayesian statistics is the credible intervals $[\xi_i^-, \xi_i^+]$ of the model parameters. These intervals contain the values that can be taken by parameter m_i with probability $(1 - \alpha)$, for some chosen small α

$$p(m_i \in [\xi_i^-, \xi_i^+]|\boldsymbol{d}) = 1 - \alpha \tag{28}$$

The lower and upper interval limits are calculated as the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles, respectively, of the posterior. We note that having sampled the spherical wavelet coefficients, inverse spherical wavelet transforms will be necessary to obtain an uncertainty map in real space as opposed to wavelet space. This can be expensive for the same reasons as previously discussed with respect to the forward operator, although to a much lesser extent after burn-in and thinned samples have been discarded. Of course, if the desired summary statistic is linear in the model parameters (e.g. mean) then this can be calculated in wavelet space and only requires a single spherical wavelet transform. Importantly, in this way, we get uncertainties at the pixel level when viewed in real space. As previously discussed, current uncertainty quantification methods in similar contexts only work on superpixels (Price et al. 2020a).

4 SEISMOLOGICAL SURFACE WAVE PHASE VELOCITY MAPS

The deep structures of the Earth's interior are best illuminated from seismic waves. Waves generated from earthquakes travel through the Earth, interacting with the materials along the way. Inverting the measurements of these waves recorded at seismic receivers, typically located at the Earth's surface, reveals the layers and structures in the Earth (Rawlinson, Pozgay & Fishwick 2010). Depending on the type and frequency of the waves, this can be done to image a range of scales from local applications (e.g. imaging oil, gas, geothermal fields) to global problems (e.g. imaging the whole Earth's mantle, down to ∼2800 km depth). The seismic imaging the Earth's interior in three dimensions is known as seismic tomography.

In this section, we introduce the common problem in global seismic tomography of building surface wave phase velocity maps, which we use as an illustrative example. Being a 2D problem that can be described by relatively simple theory, this is a natural example of application that is well suited to illustrate our framework for sampling the posterior for spherical inverse problems. The resolution requirements in global seismic tomography are also much lower than for full-sky mass-mapping, as discussed in the next section, typically only requiring $L \leq 40$. This makes our method directly applicable to this problem.

4.1 Surface wave phase velocity maps

Mapping the phase velocity of surface waves is a common problem in seismic tomography (Trampert & Woodhouse 1995; Ekström, Tromp & Larson 1997; Ekström 2011). Seismic surface waves generated from earthquakes travel along the Earth's surface. These waves are dispersive in that their velocity depends on the wave period, with different wave periods being sensitive to the structures at different depths in the Earth's interior (Dahlen & Tromp 1998). Phase velocity maps show how the velocity of surface waves at a given period varies due to lateral heterogeneities in the Earth's composition and temperature. Hence, creating these maps for waves of different periods is often a first step towards building 3D models of the Earth's mantle (e.g. Durand, Debayle & Ricard 2015).

For a particular wave period, one can measure from a seismogram the average phase velocity along the path that the wave has travelled between a seismic source and receiver. Inverting measurements from many crossing paths, that would ideally cover the Earth uniformly, produces a phase velocity map. In practice, the distribution of paths is determined by the locations of earthquakes, typically along tectonic plate boundaries, and the locations of seismic stations predominantly in the continents of the northern hemisphere, making this an ill-posed inverse problem. On global scales, long-period ($T > 25$ s) phase velocity measurements are typically modelled using linearized ray theory (an infinite frequency approximation analogous to geometrical optics), also known as the great-circle approximation (e.g. Woodhouse & Dziewonski 1984; Parisi & Ferreira 2016) and inverted using least-squares algorithms (e.g. Tarantola & Valette 1982; Trampert & Woodhouse 1995; Ekström 2011; Durand et al. 2015). In this framework, the path travelled by the seismic wave is assumed to correspond to the great circle between the source and the receiver. The observed mean phase velocity anomaly $\langle \delta c / c_0 \rangle$ for a given source-receiver pair is given by the average of the phase velocity field along the minor arc great circle S,

$$\left\langle \frac{\delta c}{c_0} \right\rangle = \frac{1}{\Delta} \int_S \frac{\delta c}{c_0}(\theta, \phi) \mathrm{d}S, \tag{29}$$

where c_0 is the phase velocity value computed for a reference Earth model, $\delta c = c - c_0$, with c being the actual phase velocity, and Δ is the length of path S.

Although there exist more complete theories describing phase velocity anomalies, notably including non-linear and finite frequency effects, recent studies have shown that the great-circle approximation accurately predicts the phase of long-period fundamental mode surface waves for current global tomography models (e.g. Parisi & Ferreira 2016; Godfrey et al. 2019). Hence, in this study, we focus on the great-circle approximation. For simplicity, we also only invert for isotropic lateral variations in phase velocity. We note, however, that our proximal MCMC approach can in principle handle these additional physical parameters (e.g. anisotropy, the directional dependence of wave velocity) with some slight modification.

4.2 Pixel space path integration

The common way to compute path integrals on the sphere (equation 29) is to rotate the coordinate system such that the path lies along the equator, which is easily expressed in terms of Wigner-D matrices and the spherical harmonic coefficients of the spherical signal (Dahlen & Tromp 1998). For a data set with N_{paths} paths, the measurement operator can be represented by a dense matrix $\mathbf{\Phi}_h \in \mathbb{C}^{N_{\text{paths}} \times L^2}$ acting on the spherical harmonic coefficients $\hat{x} \in \mathbb{C}^{L^2}$. For seismic data sets typically consisting of $\mathcal{O}(10^5)$ paths, this dense matrix multiplication can be quite slow. Our approach is to instead measure the path integral directly on the pixelized sphere using a sparse matrix $\mathbf{\Phi}_p \in \mathbb{R}^{N_{\text{paths}} \times N_{\text{pixels}}}$, where each element of $\mathbf{\Phi}_p$ is a weight representing the normalized distance each path travels in a pixel. This effectively approximates the integral as a weighted Riemann sum over the pixelized function $x \in \mathbb{R}^{N_{\text{pixels}}}$. The adjoint operator is trivially the transpose of $\mathbf{\Phi}_p$. The first step for our numerical path integration is to find discrete geographical points along the great circle minor arc between a source and a receiver using spherical trigonometry. For the discretization, a sampling rate of about 200 points per radian (3.5 points per degree) was generally sufficient for this work. Each of the geographical points along the path is then mapped to its nearest MW sampling point. This mapping assigns a weight to each MW sampling point, which is given by

$$w_{tp} = \frac{n_{tp}}{s\Delta},$$

where n_{tp} is the number of geographical points on the path that are closest to MW sampling point indexed in the θ and ϕ directions by t and p, respectively, s is the path sampling rate and Δ is the full path length. This can easily be done for each path of the data set in parallel to build the full measurement operator.

Fig. 1 compares the forward modelling time for both the harmonic and pixel space path integrations, $\mathbf{\Phi}_h$ and $\mathbf{\Phi}_p$, respectively, for bandlimits $L \in \{20, 28, 32, 64\}$ and a realistic set of ray paths . Also shown are the times for the pixel space integration with an additional spherical wavelet transform $\mathbf{\Phi}_p\mathbf{\Psi}$, as required when sampling wavelet coefficients (equation 5) instead of sampling the image directly. Clearly integration in pixel space is much faster than in harmonic space, even with the computational overhead of the spherical wavelet transforms. Pixel space integration also scales better to higher bandlimits. This is due to the extreme sparsity (less than 2 per cent nonzero elements) of $\mathbf{\Phi}_p$, whereas $\mathbf{\Phi}_h$ is generally dense. Table 1 shows the mean percentage difference and the relative squared error R2E $= \|d_{\text{harm}} - d_{\text{pix}}\|_2^2 / \|d_{\text{harm}}\|_2^2$ between predictions made in harmonic space, d_{harm}, and in pixel space, d_{pix}, when performed on the ground truth map we use in our synthetic
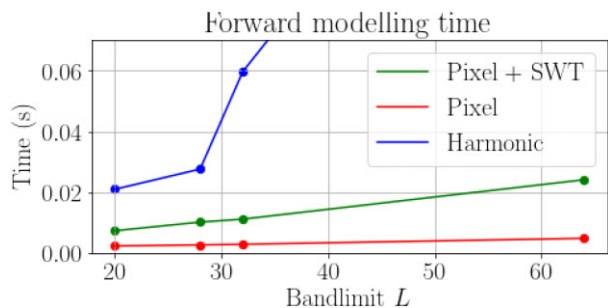
Figure 1. Time taken to perform the forward modelling in pixel space (red), in pixel space with a spherical wavelet transform (green), and in harmonic space (blue). The point for the harmonic approach at bandlimit $L = 64$ is well beyond the vertical scale at around 0.2 s. Timings performed on a 2020 MacBook Pro with an Apple M1 processor.

Table 1. Accuracy of pixel space path integration.

Bandlimit L	Mean difference (%)	R2E
20	−0.01	2.14×10^{-4}
28	−0.02	1.52×10^{-4}
32	−0.02	1.34×10^{-4}
64	−0.01	5.64×10^{-5}

experiment (see Fig. 2 top). As can be expected, as the bandlimit increases the error in the pixel space integration decreases. Crucially, even at relatively low bandlimits the pixel space integration is sufficiently accurate.

4.3 Results and discussion

In this section, we present the results of a synthetic test and real data inversions. We use MYULA to sample the axisymmetric wavelet coefficients $\boldsymbol{\alpha}$ of the spherical image $\boldsymbol{x}$, using the sparse measurement operator described in the previous section.
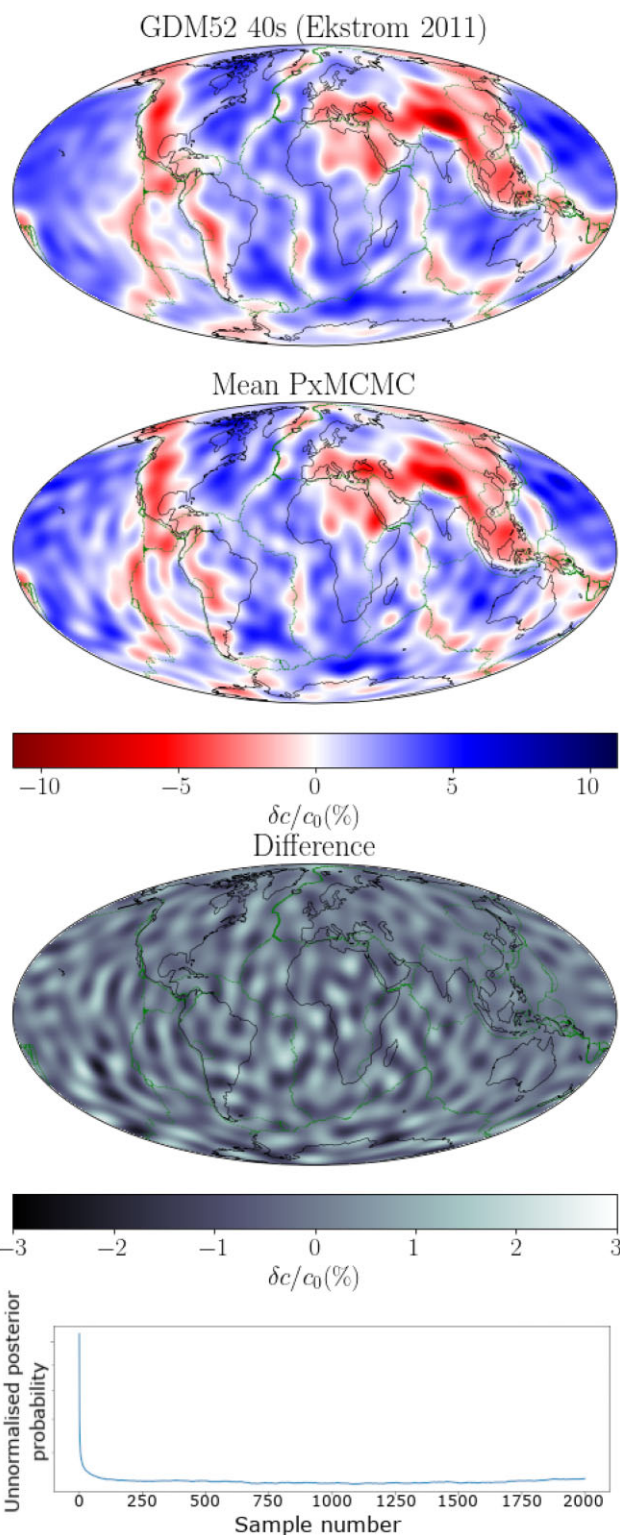
4.3.1 Synthetic experiment

As a synthetic example, we use the global phase velocity model GDM52 (Ekström 2011) at a wave period of $T = 40$ s as a ground truth $\boldsymbol{x}$. Surface waves of this period are mainly sensitive to Earth structure at depths of around 100 km (Dahlen & Tromp 1998; Durand et al. 2015), so the image (Fig. 2) shows well-known tectonic features such as slow anomalies along spreading ridges (e.g. Ekström 2011). The model is bandlimited to $L = 28$, corresponding to 3724 wavelet coefficients using the multiresolution wavelet transform. We create a synthetic data set $\boldsymbol{d}$

$$\boldsymbol{d} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}, \tag{30}$$

where $\boldsymbol{\Phi}$ is our pixel-space forward operator and $\boldsymbol{n} \sim \mathcal{N}(0, \sigma)$, where σ is the standard deviation of the predictions $\boldsymbol{\Phi}\boldsymbol{x}$ to simulate noise in observed data. $\boldsymbol{\Phi}$ is constructed using the same paths as those used to originally build GDM52 (Ekström et al. 1997; Ekström 2011), thereby ensuring a realistic and non-uniform spatial distribution of the data. In this case, we have 1 79 657 paths. We use the signal-to-noise ratio

$$\mathrm{SNR}(\boldsymbol{x}_0) = 20 \log_{10}\left(\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2}\right) \tag{31}$$

Figure 2. Synthetic GDM52 recovery experiment. Top panel: ground truth from Ekström (2011). Middle panel: mean solution from proximal MCMC. Bottom panel: difference between the truth and our solution. All the maps show perturbations in phase velocity ($\delta c/c_0$) with respect to the reference model PREM (Dziewonski & Anderson 1981). Green lines show the tectonic plate boundaries (Bird 2003). The bottom panel shows the unnormalized posterior probability throughout the MCMC chain, indicating that the sampling has converged.

95% Credible Interval Range

Ray Density (170,657 paths)
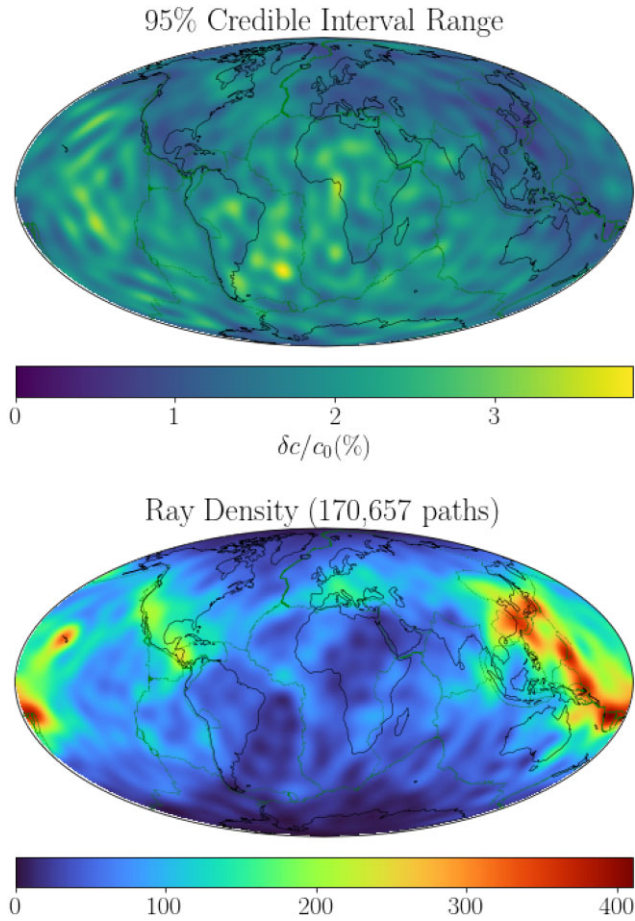
Figure 3. Synthetic GDM52 recovery experiment. Top panel: 95 per cent credible interval range (equation 28) calculated on our MCMC samples in image space ($\boldsymbol{x}$) in units of percentage deviation from the reference value, as in Fig. 2. Bottom panel: row sum of path matrix $\boldsymbol{\Phi}$ (unitless) representing the density of rays in the data set.

and relative squared error

$$\text{R2E}(\boldsymbol{x}_0) = \frac{\|\boldsymbol{d} - \boldsymbol{\Phi} \boldsymbol{x}_0\|_2^2}{\|\boldsymbol{d}\|_2^2} \tag{32}$$

to assess the reconstruction accuracy and predictive accuracy, respectively, of our chosen point solution $\boldsymbol{x}_0$, which in this case we choose to be the post-burn mean mean of our MCMC samples. We perform 10^6 MCMC steps, saving every 500th sample. The first half of the saved samples are discarded as a burn-in when calculating our mean solution and uncertainty. This takes the sampling well beyond the point of convergence, which we take to be the point where there is no longer a significant change in posterior probability (see Fig. 2). This inversion takes about 20 hours on a 2.5-GHz Intel Xeon Platinum 8180M processor. The tuning parameters are set to $\mu = 500$ and $\delta = 10^{-6}$. Fig. 2 shows the ground truth, the post-burn mean of our proximal MCMC samples and the difference between the two. Our solution has an excellent data fit (R2E = 9.96×10^{-3}) and model recovery (SNR = 8.81 dB). Differences between the ground truth and our mean solution are small on average (0.5 per cent) with some small-scale blobs of larger differences. The majority of these blobs occur in the southern hemisphere where data coverage is poorer (see Fig. 3).

We show the map of 95 per cent credible interval ranges as well as a map showing the density of ray paths of our data set in Fig. 3. Here, the uncertainty is calculated on the image space representation of our solution (i.e. $\boldsymbol{x}$). As can be expected, on the whole we find lower uncertainties where we have a higher density of data in areas such as eastern Asia and the western US. We also see much smaller scale regions of higher uncertainty. Looking at the uncertainties in wavelet space (i.e. $\boldsymbol{\alpha}$, which is sampled by the proximal MCMC) in Fig. 4, it is clear that the smaller scale wavelet coefficients have higher uncertainty. Thus, the patterns of differences between our chosen point solution and the truth (Fig. 2) are captured by the uncertainty of our sampled parameters.

4.3.2 Real data inversions

To demonstrate our method on real data, we invert the same data that was used to build GDM52 (Ekström et al. 1997; Ekström 2011) at wave periods $T = 25, 40, 75$ s, which have a strong sensitivity to about 100, 160 and 300 km depths, respectively (Dahlen & Tromp 1998; Durand et al. 2015). The results of these inversions are shown in Fig. 5. The data sets at these wave periods consist of about 103 633, 179 657, and 286 302 paths, respectively. Again we use a bandlimit of $L = 28$. For these inversions, we perform 750 000 chain steps, saving every 500th sample and discarding the first 500 saved samples as burn-in. This takes between 8 and 33 h, depending on the number of ray paths, on a 2.5-GHz Intel Xeon Platinum 8180M processor. The tuning parameter δ is chosen on a case-by-case basis. Our mean solutions show all the expected velocity anomalies, being very similar to the GDM52 phase velocity maps (Fig. 5, left-hand panel). For example, the $T = 25$ s map (Fig. 5, top panel) shows a clear distinction between the slow continents and the fast oceans. On the other hand, the $T = 40$ s map (Fig. 5, middle panel) depicts a good correlation between slow anomalies and plate boundaries, while the $T = 75$ s map (Fig. 5, bottom panel) shows deeper mantle signals, such as high velocities associated with cratons. A notable difference with GDM52 (Fig. 5, left-hand panel) is a north–south streak of fast velocities off the coast of the western US seen at all wave periods. This is a well-known artefact resulting from not modelling azimuthal anisotropy (Forsyth 1975; Ekström 2011). This streak corresponds to a region of high uncertainty in our solutions (Fig. 5, right-hand panel). Again, the uncertainty maps correlate with ray density as expected.

We emphasize that the main purpose of this study is not to build improved phase velocity maps, which can be constructed quickly using, e.g. least-squares approaches (Tarantola & Valette 1982; Trampert & Woodhouse 1995), but rather to illustrate and validate our framework for sampling the posterior of spherical inverse problems with a useful, well-known first application. Hence, we do not consider more sophisticated theoretical formulations, such as, e.g. full ray theory (Ferreira & Woodhouse 2007), finite frequency theory (Zhou et al. 2005), including anisotropic effects (Ekström 2011), etc. Future work will focus on the application of the method to more-sophisticated problems, such as, for example, depth inversions using non-linear theory.

5 LOW-RESOLUTION COSMOLOGICAL MASS-MAPPING

One of the predictions of Einstein's theory of general relativity was that the gravitational influence of massive objects will cause light to bend around them (Einstein 1905). As a result, distant objects often look distorted when observed by astronomers. This is the phenomenon of gravitational lensing. Light from distant stars and galaxies travels through the universe, bending around all the mass along the way to Earth. Inverting measurements of the
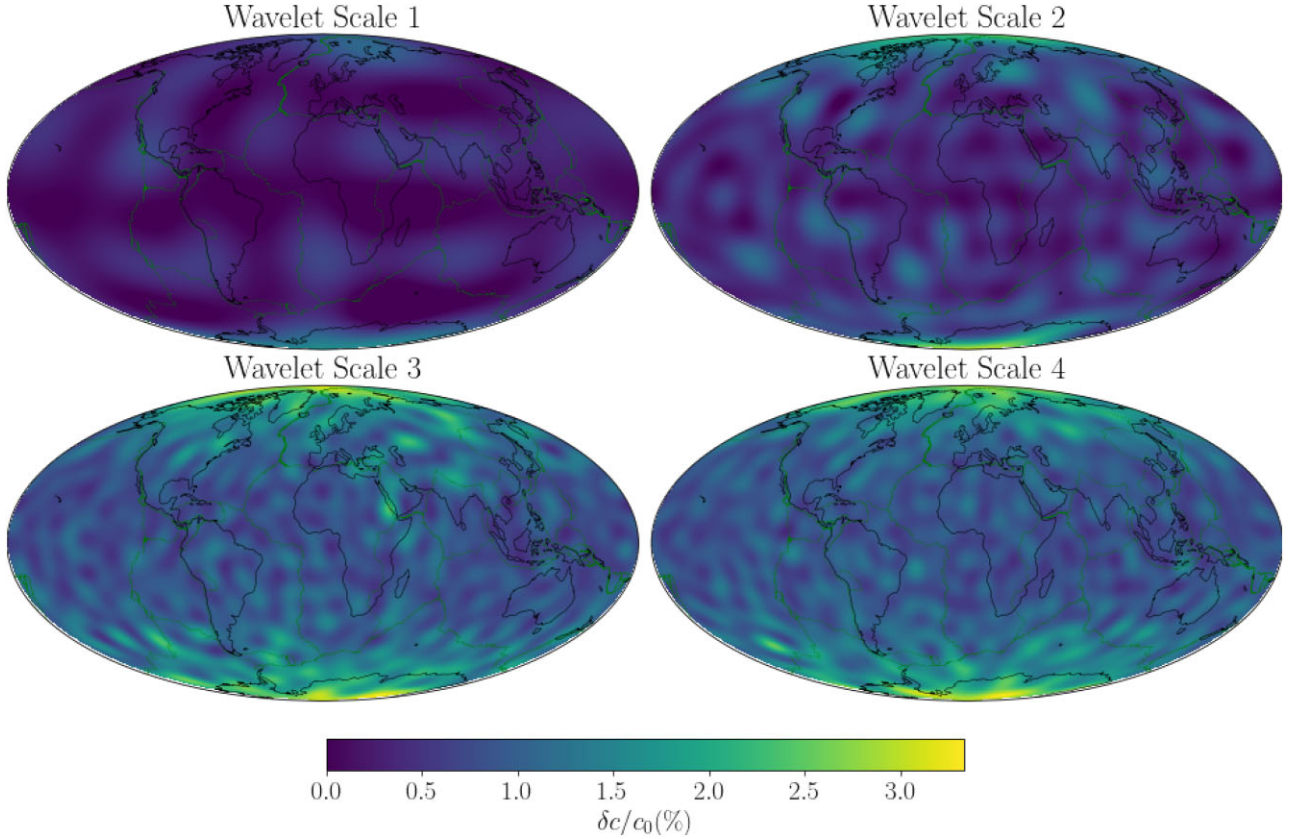
Figure 4. Synthetic GDM52 recovery experiment. 95 per cent credible interval ranges of spherical wavelet coefficients at different scales. The range of the colourbar is the same for all maps in units of percentage deviation from the reference value, as in Figs 2 and 3.

distortions in the images of the light source reveals the hidden masses that the light passed by (e.g. Kaiser & Squires 1993), including dark matter (Heavens 2009), producing so-called cosmological mass maps.

In this section, we demonstrate our framework on the retrieval of cosmological mass-maps from simulated data. This is a 2D problem described by a simple linear forward model. We treat this as a simple demonstration and perform the inversion at a much lower bandlimit than would normally be required for cosmological mass-mapping. Due to the poor scaling of the spherical wavelet transforms with L, performing the millions of transforms required at $L \sim \mathcal{O}(10^3)$ (Price et al. 2020b; Wallis et al. 2022) would simply be too slow. However, with future computational improvements to speed up the transforms (e.g. by exploiting GPUs), our framework will be well-suited for this particular problem.

5.1 Mass-mapping on the celestial sphere

Images of galaxies are typically distorted, as the light they emit gets lensed by the mass between us and the source galaxies. Gravitational lensing occurs regardless of the nature of the intervening mass, and as such lensing is an excellent probe for dark matter (Heavens 2009). Mass-mapping maps the total density perturbation along a line of sight between a source galaxy and the observer based on measurements of the distortion of galaxy images. Up until recently, lensing surveys only covered relatively small sky-fractions, so planar approximations were made. As the area of coverage has increased with newer surveys, planar approximations are no longer valid (Wallis et al. 2022), resulting in mass-maps now being constructed

on the sphere (e.g. Price et al. 2020b; Jeffrey et al. 2021; Wallis et al. 2022).

Gravitational lensing studies deal with two main fields: the convergence field $_0\kappa(\theta, \phi)$, which causes magnification of the galaxy image; and the shear field $_2\gamma(\theta, \phi)$, which causes rotation and stretching of the galaxy image (Dodelson 2017). Mass maps reveal the convergence field $_0\kappa(\theta, \phi)$, which can be shown to be the integrated mass density along the line of sight (Bartelmann & Schneider 2001). This is linearly related in spherical harmonic space to the spin-2 shear field $_2\gamma(\theta, \phi)$, measured from observations of galaxy shapes, by a linear kernel $\mathcal{K}_\ell$ given by (Kaiser & Squires 1993; Wallis et al. 2022)

$$\mathcal{K}_\ell = \frac{-1}{\ell(\ell+1)}\sqrt{\frac{(\ell+2)!}{(\ell-2)!}}. \tag{33}$$

Thus, our forward model, in this case, is given by $\Phi = m_2 S^{-1} K_0 S$, where K encodes the linear kernel above and m is a masking matrix to account for areas on the sky without reliable data (e.g. the galactic plane and ecliptic). $_2S$ and $_0S$ denote the spin-2 and spin-0 spherical harmonic transforms needed for to account for the spin symmetries of the shear and convergence fields, respectively. From hereon in we drop the spin subscripts for clarity. We note that in reality there exists a degeneracy between γ and κ, and as a result the true observable is not the shear but the reduced shear $g = \gamma/(1-\kappa)$. In the weak-lensing regime, this non-linear effect is very small, so we ignore it here. Accounting for non-linearities in our MCMC is possible, provided the relevant gradients of the non-linear forward model can be computed efficiently.
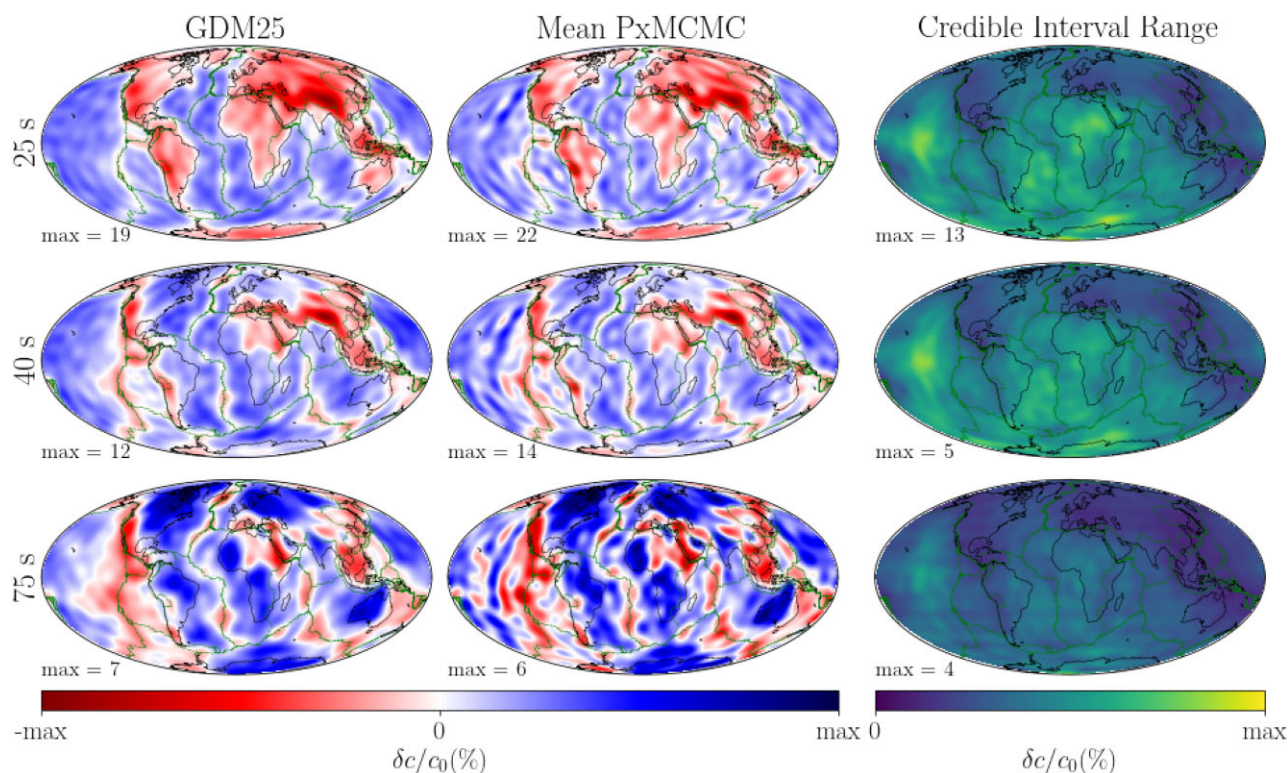
Figure 5. Real data inversions. Phase velocity maps from GDM52 (Ekström 2011) (left-hand panel), mean proximal MCMC solutions from this study (middle panel) and credible interval ranges (right-hand panel) for wave periods 25, 40, and 75 s (top to bottom panels). All the maps show perturbations in phase velocity ($\delta c / c_0$) with respect to the reference model PREM (Dziewonski & Anderson 1981).

The choice of using sparsity promoting priors in this case is motivated by the need to recover non-Gaussian structures, particularly at high ℓ. These are created by non-linear structure growth of the density field throughout the evolution of the Universe. Promoting sparsity has been used previously (e.g. Leonard, Lanusse & Starck 2014; Lanusse et al. 2016; Price et al. 2020b; Starck et al. 2021) to incorporate non-Gaussian structures, as have Wiener filters on a Gaussian prior (e.g. Jeffrey et al. 2018),log-normal priors (e.g. Böhm et al. 2017; Fiedorowicz et al. 2022) and physically informed power spectrum priors (e.g. Porqueres et al. 2021).

A Bayesian sampling method for mass-mapping was recently implemented by Fiedorowicz et al. (2022). They used HMC to efficiently navigate the large parameter space with a log-normal prior. In comparison with our framework, the log-normal prior is differentiable, which allows the gradient-based HMC to be used, with the exploitation of efficient autodifferentiation making this computationally tractable. However, the log-normal prior makes certain assumptions about the cosmological parameters, and thus the mass maps produced by their method cannot be used for inference about cosmology, without additional further work (Fiedorowicz et al. 2022). Our framework makes no such assumptions.

5.2 Results and discussion

As a synthetic example, we attempt to reconstruct a simulated mass map from the Takahashi *N*-body simulation (Takahashiet al. 2017).[3] Slices are provided at a range of redshifts, and we select redshift slice

[3]Data available at http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/

16 corresponding to $z \sim 1$. This slice is bandlimited at $L = 64$, giving our ground truth mass map, to which we apply a basic Galactic plane and ecliptic mask.

Synthetic shear data is generated by applying the measurement operator Φ to the ground truth map. The noise in weak-lensing surveys depends on the number of galaxy counts per unit area and the variance of the intrinsic ellipticity distribution $\sigma_e \sim 0.37$. We choose an overall number density of galaxies per arcmin^2, n_{gal}, and add zero-mean Gaussian noise to the synthetic data, with the variance of the noise given by

$$\sigma_t^2 = \frac{\sigma_e^2}{\sqrt{2n_t}}, \tag{34}$$

where t is the colatitude index, and n_t is the expected number of galaxies in a pixel at colatitude t given the overall number density n_{gal}. This dependence on colatitude comes from the equiangular nature of the MW sampling theorem (McEwen & Wiaux 2011).

When reporting the SNR, R2E summary statistics and our solution maps, we use a slightly larger mask than what was applied to the synthetic shear data. This is to remove leakage artefacts that occur around the edge of the mask due to the sudden lack of data and also wavelets that have support both inside and outside the mask. At the high resolutions typically of interest for mass-mapping, this leakage will be minimal.

Maps of the ground truth, mean of our MCMC samples, the difference between them and our measured uncertainty are shown in Fig. 6. For this inversion, we perform 12.5×10^6 chain steps, saving every 500th sample and using the last 3000 samples for our results, taking about 4 d on a 2.5-GHz Intel Xeon Platinum 8180M processor. The step-size parameter δ is chosen to be 10^{-9}, the largest value that
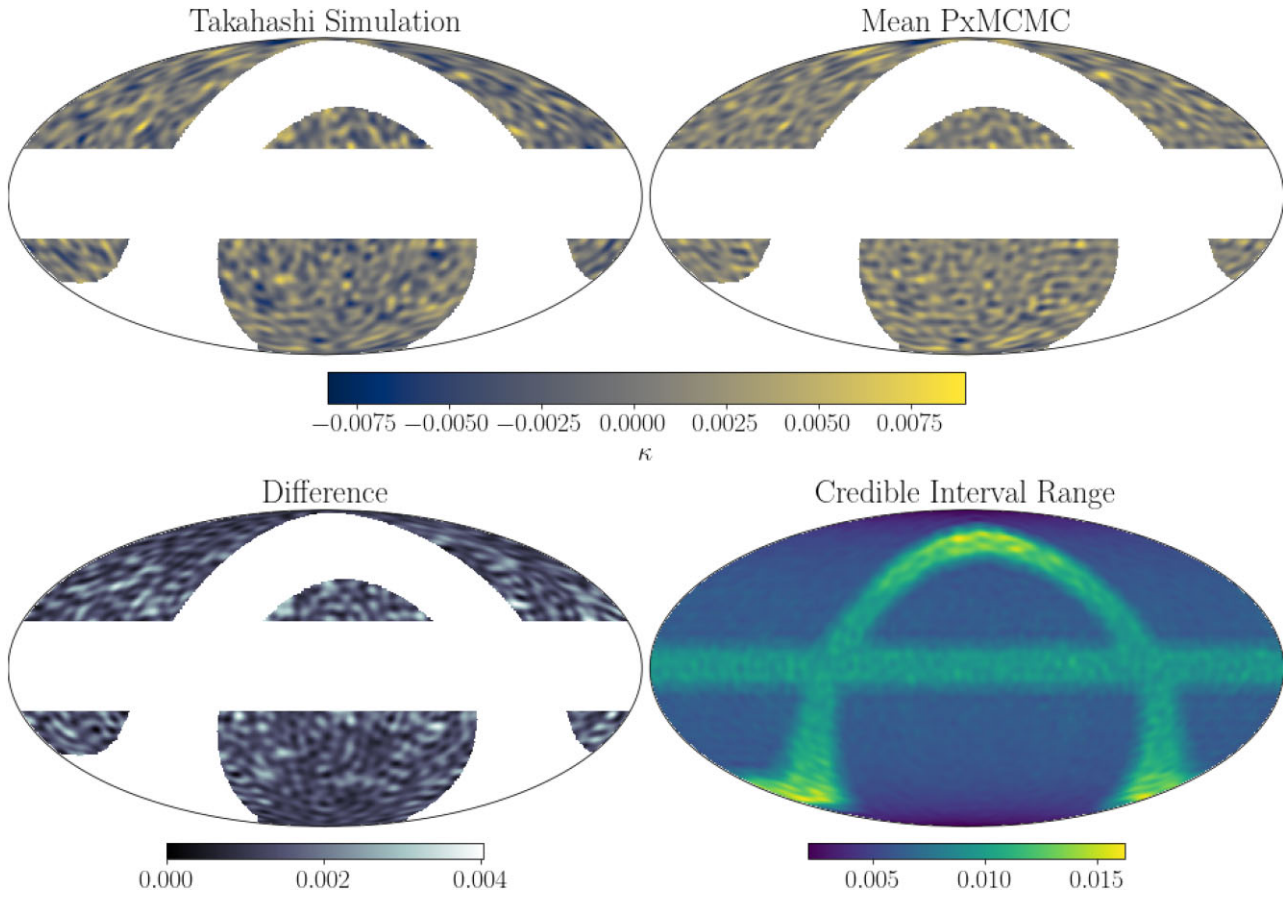
Figure 6. Results of a simple mass-mapping example where the ground truth is known. Top left-hand panel: ground truth simulation from Takahashiet al. (2017). Top right-hand panel: Mean of the proximal MCMC samples. Bottom left-hand panel: difference between the top two maps. Bottom right-hand panel: 95 per cent credible interval range obtained from proximal MCMC. A Galactic plane and ecliptic mask are show in all maps except the uncertainty map, which, as expected, shows high uncertainty in these regions.

allows a stable Euler approximation of the Langevin diffusion. The regularization parameter μ is chosen to be 5×10^5, to constrain the wavelet coefficients to the appropriate order of magnitude. At this resolution ($L = 64$), there are 18 916 wavelet coefficients to be sampled. The mean solution has an SNR of 7.83 dB and an R2E of 0.1. The uncertainty map, measured as the range of the 95 per cent credible intervals (equation 28) show, as expected, high uncertainty within the masked regions where the solution is not constrained by any data.

While we obtain encouraging results here with a good reconstruction and physically reasonable pixel-level uncertainties, for this method to be adopted in full-sky mass mapping, computational advances are needed in the implementation of the spherical harmonic and wavelet transforms such that the scaling with L is not as severe. Currently, the complexity scaling of the spherical harmonic transforms is $\mathcal{O}(L^3)$ (McEwen & Wiaux 2011), and dominates over the efficient harmonic space wavelet transform (Leistedt et al. 2013) discussed in Section 2. With at least two spherical harmonic transforms per iteration of MYULA, current implementations will not allow for the desired $L \sim \mathcal{O}(10^3)$ bandlimits. New efforts implementing these transforms on GPUs should go a significant way towards pushing our method to higher resolutions. Further theoretical advances for the representation of spherical signals could also lead to computational savings (e.g. Ocampo, Price & McEwen 2022).

6 CONCLUSIONS

In this paper, we have presented a general framework for posterior sampling of inverse problems on the sphere with sparsity promoting priors that allows for flexible uncertainty quantification and extends naturally to non-linear problems. We demonstrated the potential applicability of this method to both astrophysical and geophysical problems. As with all MCMC methods, the suitability of this method depends on the time taken to take the next chain sample, particularly in relation to the forward modelling step. The computational complexity of transforms on the sphere mean that this framework is generally feasible for problems of low to moderate resolution (roughly $L \leq 64$), such as those commonly considered in global seismic tomography. At higher resolutions, as needed for full-sky mass-mapping, posterior sampling quickly becomes intractable largely due to the poor scaling of spherical harmonic transforms present in the forward operator. In either case, special consideration must be given to the forward operator and whether it should be formulated in harmonic or pixel space, and also if its adjoint is known. A harmonic formulation would avoid repeated spherical harmonic transforms, but these savings could be lost on the measurement operator. Making use of a more efficient proximal algorithm based on a stochastic Runge–Kutta approximation of the Langevin equation (Pereyra et al. 2020) could be a promising way forward for higher resolution spherical inverse problems. This algorithm is more complex but converges to a solution much faster than the Euler approximation algorithm used in

this work, thereby potentially requiring fewer spherical harmonic transforms. Additionally, faster implementations of the spherical harmonic transform leveraging GPUs would immediately increase the potential of our method.

The examples shown in this work are largely illustrative, as simplifications have been made. In the mass-mapping example, the bandlimit is much lower than the angular orders at which the convergence spectrum has the most power due to the computational restrictions imposed by the spherical harmonic transforms. We also have not fully considered the effect of reduced shear. Nonetheless, ongoing work to implement the spherical harmonic and wavelet transforms on accelerators (GPUs) provide a route to scale to higher resolutions, making our framework is a promising addition to other recent methods for obtaining mass-maps with uncertainties (e.g. Price et al. 2020b; Fiedorowicz et al. 2022). For the surface wave tomography example, we have used the great circle approximation and not accounted for anisotropic effects, in which case a least-squares approach is fast and efficient. None the less, our results demonstrate the feasibility of our framework methods on global scale tomographic inverse problems. In this case, the more commonly used harmonic formulation of the forward problem proved to be too slow, and we were able to reformulate it in pixel space such that it was much faster, even with the computational overhead of spherical wavelet transforms. Further, the uncertainties calculated from our posterior samples make physical sense, being correlated with the distribution of data. Bayesian methods in seismic tomography on large-to-global scales have largely been used for independent 1D inversions (e.g. Shapiro & Ritzwoller 2002; Khan, Boschi & Connolly 2011; Ravenna & Lebedev 2017), although new advances in gradient-based or variational methods (e.g. Fichtner, Zunino & Gebraad 2018; Gebraad, Boehm & Fichtner 2020; Zhang & Curtis 2020; Zhao, Curtis & Zhang 2022) are promising for 2D and 3D probabilistic tomography. Our framework is a further contribution to this advance in methodology, with the novelty of being able to use a non-differentiable prior.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

We provide a new Python package, `pxmcmc`[4] that contains implementations of the proximal MCMC algorithms used in this work as well as the measurement operators, wavelet transforms and priors with their proximal mappings. The code is designed to be flexible, with base classes that will allow users to implement their own forward (measurement and transform) operators and priors. The proximal MCMC algorithms implemented are themselves not restricted to spherical problems, as the spherical aspects of the inversions appear in the likelihood and prior proximal calculations. We also provide scripts and data to reproduce the synthetic experiment described in this paper. MCMC chains for the real data inversions are available from the authors upon request. The discretization of great circle paths

is implemented in our Python code `greatcirclepaths`,[5] which is publicly available.

REFERENCES

Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bauschke H. H., Combettes P. L., 2017, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York
Bird P., 2003, Geochem. GeoPhys. Geosys., 4, 1027
Bodin T., Sambridge M., Gallagher K., Rawlinson N., 2012, J. Geophys. Res.: Solid Earth, 117, B02301
Böhm V., Hilbert S., Greiner M., Enßlin T. A., 2017, Phys. Rev. D, 96, 123510
Cai X., Pereyra M., McEwen J. D., 2018a, MNRAS, 408, 4154
Cai X., Pereyra M., McEwen J. D., 2018b, MNRAS, 480, 4170
Cai X., McEwen J. D., Pereyra M., 2022, Stat. Comput., 32, 87
Candès E. J., Eldar Y. C., Needell D., Randall P., 2011, Appl. Comput. Harmon. Anal., 31, 59
Carrillo R. E., McEwen J. D., Wiaux Y., 2014, MNRAS, 439, 3591
Chang S.-J., Ferreira A. M. G., Ritsema J., van Heijst H. J., Woodhouse J. H., 2015, J. Geophys. Res.: Solid Earth, 120, 4278
Charléty J., Voronin S., Nolet G., Loris I., Simons F. J., Sigloch K., Daubechies I. C., 2013, J. Geophys. Res.: Solid Earth, 118, 4887
Combettes P. L., Pesquet J.-C., 2011, Proximal Splitting Methods in Signal Processing. Springer, New York, NY, p. 185
Corless V. L., King L. J., Clowe D., 2009, MNRAS, 393, 1235
Dahlen F. A., Tromp J., 1998, Theoretical Global Seismology. Princeton Univ. Press, Princeton
Dodelson S., 2017, Gravitational Lensing. Cambridge Univ. Press, Cambridge
Donoho D. L., 2006, IEEE Transact. IT, 52, 1289
Driscoll J. R., Healy D. M., 1994, Adv. Appl. Math., 15, 202
Duane S., Kennedy A., Pendleton B. J., Roweth D., 1987, Phys. Lett. B, 195, 216
Durand S., Debayle E., Ricard Y., 2015, Geophis. Res. Lett., 42, 3266
Dziewonski A. M., Anderson D. L., 1981, Phys. Earth Planet. Interiors, 25, 297
Einstein A., 1905, Annalen der Physik, 322, 891
Ekström G., 2011, Geophys. J. Int., 187, 1668
Ekström G., Tromp J., Larson E. W. F., 1997, J. Geophys. Res.: Solid Earth, 102, 8137
Ferreira A. M. G., Woodhouse J. H., 2007, Geophys. J. Int., 168, 109
Ferreira A. M. G., Marignier A., Attanayake J., Frietsch M., Berbellini A., 2020, Geophys. J. Int., 221, 1232
Fichtner A., Zunino A., Gebraad L., 2018, Geophys. J. Int., 216, 1344
Fiedorowicz P., Rozo E., Boruah S. S., Chang C., Gatti M., 2022, MNRAS, 512, 73
Forsyth D. W., 1975, Geophys. J. Int., 43, 103
Gebraad L., Boehm C., Fichtner A., 2020, J. Geophys. Res.: Solid Earth, 125, e2019JB018428
Godfrey K. E., Dalton C. A., Ma Z., Hjörleifsdóttir V., Ekström G., 2019, Geophys. J. Int., 217, 1496
Goldberg J. N., Macfarlane A. J., Newman E. T., Rohrlich F., Sudarshan E. C. G., 1967, J. Math. Phys., 8, 2155
Golub G. H., Hansen P. C., O'Leary D. P., 1999, SIAM J. MAA, 21, 185
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, AJ, 622, 759
Heavens A., 2009, Nucl. Phys. B., 194, 76
Jaynes E. T., 2003, Probability Theory: The Logic of Science. Cambridge Univ. Press, Cambridge
Jeffrey N. et al., 2018, MNRAS, 479, 2871
Jeffrey N. et al., 2021, MNRAS, 505, 4626
Kaiser N., Squires G., 1993, AJ, 404, 441
Khan A., Boschi L., Connolly J. A. D., 2011, J. Geophys. Res.: Solid Earth, 116, B01301
Lanusse F., Starck J. L., Leonard A., Pires S., 2016, A&A, 591, A2

---

[4]https://github.com/auggiemarignier/pxmcmc

[5]https://github.com/auggiemarignier/GreatCirclePaths

Leistedt B., McEwen J. D., Vandergheynst P., Wiaux Y., 2013, A&A, 558, A128

Leonard A., Lanusse F., Starck J.-L., 2014, MNRAS, 440, 1281

Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511

Loris I., Nolet G., Daubechies I., Dahlen F. A., 2007, Geophys. J. Int., 170, 359

Malinverno A., 2002, Geophys. J. Int., 151, 675

McEwen J. D., Wiaux Y., 2011, IEEE J. SP, 59, 5876

McEwen J. D., Leistedt B., Büttner M., Peiris H. V., Wiaux Y., 2015a, preprint (arXiv:1509.06749)

McEwen J. D., Büttner M., Leistedt B., Peiris H. V., Wiaux Y., 2015b, IEEE Signal Process. Lett., 22, 2425

McEwen J. D., Durastanti C., Wiaux Y., 2018, Appl. Computational Harmonic Analysis, 44, 59

Moreau J. J., 1962, CRHS_AS, 255, 2897

Mosegaard K., Tarantola A., 1995, J. Geophys. Res.: Solid Earth, 100, 12431

Neal R. M., 2012, Handbook of Markov Chain Monte Carlo. CRC Press, Boca Raton, Florida, p. 113

Newman E. T., Penrose R., 1966, J. Math. Phys., 7, 863

Ocampo J., Price M. A., McEwen J. D., 2022, preprint (arXiv:2209.13603)

Parikh N., Boyd S., 2014, Foundat. Trends Opt., 1, 127

Parisi L., Ferreira A. M. G., 2016, Geophys. J. Int., 205, 146

Pereyra M., 2016, Stat. Comput., 26, 745

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

Pereyra M., Mieles L. V., Zygalakis K. C., 2020, SIAM J. Imag. Sci., 13, 905

Planck Collaboration I, 2016, A&A, 594, A1

Porqueres N., Heavens A., Mortlock D., Lavaux G., 2021, MNRAS, 502, 3035

Pratley L., McEwen J. D., d'Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2017, MNRAS, 473, 1038

Price M. A., Cai X., McEwen J. D., Pereyra M., Kitching T. D., 2020a, MNRAS, 492, 394

Price M. A., McEwen J. D., Pratley L., Kitching T. D., 2020b, MNRAS, 500, 5436

Ravenna M., Lebedev S., 2017, Geophys. J. Int., 213, 278

Rawlinson N., Pozgay S., Fishwick S., 2010, Phys. Earth Planet. Int., 178, 101

Ritsema J., Deuss A., van Heijst H. J., Woodhouse J. H., 2011, Geophys. J. Int., 184, 1223

Roberts G. O., Rosenthal J. S., 1998, JRSS B., 60, 255

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

Shapiro N. M., Ritzwoller M. H., 2002, Geophys. J. Int., 151, 88

Skilling J., 2006, Bayesian Anal., 1, 833

Starck J. L., Themelis K. E., Jeffrey N., Peel A., Lanusse F., 2021, A&A, 649, A99

Takahashi R., Hamana T., Shirasaki M., Namikawa T., Nishimichi T., Osato K., Shiroyama K., 2017, AJ, 850, 24

Tarantola A., Valette B., 1982, Rev. Geophys. Space Phys., 20, 219

Trampert J., Woodhouse J. H., 1995, Geophys. J. Int., 122, 675

Wallis C. G., Wiaux Y., Mcewen J. D., 2017, IEEE J. IP, 26, 5176

Wallis C. G. R., Price M. A., McEwen J. D., Kitching T. D., Leistedt B., Plouviez A., 2022, MNRAS, 509, 4480

Wiaux Y., McEwen J. D., Vandergheynst P., Blanc O., 2008, MNRAS, 388, 770

Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vandergheynst P., 2009, MNRAS, 395, 1733

Woodhouse J. H., Dziewonski A. M., 1984, J. Geophis. Res., 89, 5953

Zhang X., Curtis A., 2020, Geophys. J. Int., 222, 406

Zhao X., Curtis A., Zhang X., 2022, Geophys. J. Int., 228, 213

Zhou Y., Dahlen F. A., Nolet G., Laske G., 2005, Geophys. J. Int., 163, 1087

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