Sparse Bayesian mass-mapping with uncertainties: full sky observations on the celestial sphere

M. A. Price[1]⋆, J. D. McEwen[1], L. Pratley[1,2], and T. D. Kitching[1]

[1] *Mullard Space Science Laboratory, University College London, RH5 6NT, UK*
[2] *Dunlap Institute for Astronomy and Astrophysics, University of Toronto, ON M5S 3H4, Canada*

ABSTRACT
To date weak gravitational lensing surveys have typically been restricted to small fields of view, such that the *flat-sky approximation* has been sufficiently satisfied. However, with Stage IV surveys (*e.g. LSST* and *Euclid*) imminent, extending mass-mapping techniques to the sphere is a fundamental necessity. As such, we extend the sparse hierarchical Bayesian mass-mapping formalism presented in previous work to the spherical sky. For the first time, this allows us to construct *maximum a posteriori* spherical weak lensing dark-matter mass-maps, with principled Bayesian uncertainties, without imposing or assuming Gaussianty. We solve the spherical mass-mapping inverse problem in the analysis setting adopting a sparsity promoting Laplace-type wavelet prior, though this theoretical framework supports all log-concave posteriors. Our spherical mass-mapping formalism facilitates principled statistical interpretation of reconstructions. We apply our framework to convergence reconstruction on high resolution N-body simulations with pseudo-Euclid masking, polluted with a variety of realistic noise levels, and show a significant increase in reconstruction fidelity compared to standard approaches. Furthermore we perform the largest joint reconstruction to date of the majority of publicly available shear observational datasets (combining DESY1, KiDS450 and CFHTLens) and find that our formalism recovers a convergence map with significantly enhanced small-scale detail. Within our Bayesian framework we validate, in a statistically rigorous manner, the community's intuition regarding the need to smooth spherical Kaiser-Squires estimates to provide physically meaningful convergence maps. Such approaches cannot reveal the small-scale physical structures that we recover within our framework.

Key words: gravitational lensing: weak – (*Cosmology:*) large-scale structure of Universe – Methods: statistical – Methods: data analysis – techniques: image processing – techniques: compressed sensing

1 INTRODUCTION

Gravitational lensing is an astrophysical phenomenon through which the geometry of distant galaxies becomes distorted by the intervening matter distribution. Mathematically, this lensing effect is a perturbation by the local matter topology of the null geodesics along which photons travel (Grimm & Yoo 2018; Bartelmann & Schneider 2001; Schneider 2005). As such, gravitational lensing is sensitive to all matter (both visible and invisible) and is thus a natural tool with which to probe the nature of dark matter.

Weak gravitational lensing refers to the vast majority of lensing events for which images are not multiply sourced or 'strongly lensed'. Equivalently the weak lensing regime can be defined as the regime in which the lensing perturbations remain (to a good approximation) linear. At first order the effect of weak lensing on distant galaxy images manifests itself as two quantities: the spin-0 magnification referred to as the *convergence* field $_0\kappa$, and a spin-2 perturbation to the ellipticity (third-flattening) referred to as the *shearing* or *shear* field $_2\gamma$.

Due to the *'mass-sheet degeneracy'* there is no way to construct *a priori* estimates of the intrinsic brightness, hence the convergence field is an unobservable quantity – theoretically one could infer the convergence field directly from the galaxy sizes, but the intrinsic dispersion is too high (Alsing et al. 2015). However, as the distribution of instrinsic ellipticities has zero mean and sufficiently tight dispersion,

⋆ E-mail: m.price.17@ucl.ac.uk

averaging sufficient observations within a given pixel can provide an accurate estimator for the shear signal. As such, measurements of the shear field $_2\gamma$ are typically taken and inverted to form estimates of $_0\kappa$ – coined *dark matter mass maps* by Clowe et al. (2006).

A large proportion (Taylor et al. 2018) of cosmological information can be extracted directly from the shear field (Van Waerbeke et al. 2013; Fluri et al. 2019; Giblin et al. 2018), however recently cosmologists have become increasingly interested in extracting information from higher order statistics, such as peak & void statistics and Minkowski functionals, which are typically calculated directly from the convergence field (*e.g.* Munshi & Coles 2017; Coles & Chiang 2000; Peel et al. 2018; Fluri et al. 2018) – motivating research into optimal mass-mapping techniques. Typically, these higher order statistics aim to probe the non-Gaussian information content of the convergence field.

Mapping from shear to convergence (mass-mapping) requires solving an (often seriously) ill-posed inverse problem – mass-mapping takes the form of a typical noisy deconvolution problem with a spin-2 kernel (Wallis et al. 2017a), which is classically ill-posed. The most naive mass-mapping technique for small fields of view is planar Kaiser-Squires (KS; Kaiser & Squires 1993) which is direct inversion of the forward model in Fourier space. This estimator does not take into account noise or boundary effects, and so is typically post-processed *via* convolution with a large Gaussian smoothing kernel, thus heavily degrading the quality of high-resolution non-Gaussian information. Moreover, decomposition of spin-fields on bounded manifolds is known to be degenerate (Bunn et al. 2003) and so for non-trivial masking the KS estimator is ill-defined and can be shown to perform poorly (see Section 5).

Many, perhaps more sophisticated, approaches to mass-mapping on the plane have been developed (*e.g.* VanderPlas et al. 2011; Lanusse et al. 2016; Jeffrey et al. 2018; Jee et al. 2016) though all either lack a principled statistical framework or rely heavily on assumptions or impositions of Gaussianity. In previous work we present a sparse hierarchical Bayesian formalism for planar mass-mapping (Price et al. 2018; Price et al. 2019b,a) that provides fully principled statistical uncertainties without the need to assume Gaussianity and without the computational overhead of MCMC methods (*e.g.* Schneider et al. 2015; Corless et al. 2009; Alsing et al. 2016).

One key assumption of these 'planar' mass-mapping techniques is that the area of interest on the sky can be well approximated as a plane. This assumption is colloquially referred to as the the *flat-sky approximation*. For small-field surveys this approximation is typically justified. However for future wide-field Stage IV surveys mass-mapping must be constructed natively on the sphere (Chang et al. 2018) to avoid errors due to projection effects, which can be large (Wallis et al. 2017a; Vallis et al. 2018). Naturally one can naively invert the spherical forward model to form the *spherical Kaiser-Squires* estimator (SKS; Wallis et al. 2017a) which avoids projection effects but is seriously ill-posed, as is the KS method. It should be noted that alternative techniques for spherical reconstruction have also been developed (*e.g.* Pichon et al. 2010).

In this paper we extend the previously developed hierarchical Bayesian-sparse formalism to the sphere which, for

the first time, allows *maximum a posteriori* (MAP) convergence reconstruction with principled Bayesian uncertainties in very high-dimensions natively on the sphere without making any assumptions or impositions of Gaussianity. Throughout this paper we refer to our estimator, formed within this framework, as the DarkMapper estimator (and by extension the DarkMapper codebase). The reconstruction formalism presented in this paper and any unceraity quantification techniques that follow support any choice of likelihood or prior such that the posterior function belongs to the (rather comprehensive) set of log-concave functions. As such one can incorporate various experimental or systematic effects in future, *e.g.* more complex noise models or intrinsic alignment corrections *etc.*

The structure of this paper is as follows. In Section 2 we provide background mathematical details relevant to the scope of this paper, such as the analysis of spin signals on the sphere, and succiently review weak gravitational lensing. Following this Section 3 provides a cursory introduction to Bayesian analysis before presenting and discussing both the general hierarchical Bayesian formalism and our DarkMapper estimator. In this section we explicitly outline the likelihood and priors used throughout this paper but place emphasis on the generality of this formalism. Furthermore, we outline how to fold uncertainty in regularization parameters into the hierarchy *via* the allocation of a suitable (here a conjugate) hyper-prior distribution. In Section 4 we extend previously developed uncertaintiy quantification techniques to the spherical space and discuss how one should approach constructing custom uncertainty quantification techniques which fit within our formalism. In sections 5, using high resolution N-body (Takahashi et al. 2017) simulations, pseudo-Euclid masking (a masking of the galactic plane and the ecliptic) and noise realisations representative of a variety of weak lensing survey eras (including Stage IV) we demonstrate the drastic increase in reconstruction fidelity of DarkMapper over SKS. Penultimately, in Section 6 we apply both the SKS and DarkMapper estimators to a global weak lensing dataset constructed *via* the concatenation of the majority of publicly available observational datasets. To the best of our knowledge this is the first such global spherical dark-matter mass-maps. Furthermore, we perform global Bayesian uncertainty quantification on these reconstructions. Finally, in Section 7 we draw conclusions.

2   BACKGROUND

Here we present a cursory synopsis of the relevant background required to understand weak lensing on the sphere. In no way is this a complete description and so we recommend the reader follow related papers (Wallis et al. 2017b,a; McEwen et al. 2015b).

2.1   Spin-s Spherical Fields

Local rotations by $\chi \in [0, 2\pi)$ about the tangent plane centered on the spherical coordinate $\omega = (\theta, \psi) \in \mathbb{S}^2$ of square integrable spin-s fields for $s \in \mathbb{Z}$ are defined generally by (Goldberg et al. 1967; Newman & Penrose 1966; Wallis et al.

2017a; McEwen et al. 2013a)

$$_s f'(\omega) = e^{-is\chi} {_s}f(\omega), \tag{1}$$

where $\omega = (\theta, \psi)$ are standard spherical coordinates, given by co-latitude $\theta \in [0, \pi)$ and longitude $\psi \in [0, 2\pi)$. The natural set of orthogonal basis functions for spherical fields are the *spherical harmonics* $Y_{\ell m}(\omega)$.

When considering spin-s fields on $\mathbb{S}^2$ the natural set of orthogonal basis functions are the *spin-weighted spherical harmonics*. The spin weighted spherical harmonics are generated by application of the spin raising and lowering operators ($\eth$ and $\bar{\eth}$ respectively) to the spherical eigenfunctions $Y_{\ell m}(\omega)$. The spin-s raising and lowering operators are given respectively, by

$$\eth \equiv -\sin^s \theta \left(\frac{\partial}{\partial \theta} + \frac{i\partial}{\sin\theta \partial\psi} \right) \sin^{-s}\theta, \tag{2}$$

$$\bar{\eth} \equiv -\sin^{-s}\theta \left(\frac{\partial}{\partial\theta} - \frac{i\partial}{\sin\theta\partial\psi} \right) \sin^s\theta. \tag{3}$$

On application to $_s Y_{\ell m}(\omega)$ we find the recursion relations,

$$\eth\, _s Y_{\ell m}(\omega) = \left[(\ell-s)(\ell+s+1) \right]^{1/2} {_{s+1}}Y_{\ell m}(\omega), \tag{4}$$

$$\bar{\eth}\, _s Y_{\ell m}(\omega) = -\left[(\ell+s)(\ell-s+1) \right]^{1/2} {_{s-1}}Y_{\ell m}(\omega). \tag{5}$$

Following these recursions it is clear that any spin-s weighted spherical harmonic can be represented as $s \in \mathbb{N}$ repeated applications of the spin raising (lowering) operator $\eth$ to the standard spin-0 spherical harmonic $Y_{\ell m}$ such that,

$$_s Y_{\ell m}(\omega) = \left[\frac{(\ell-s)!}{(\ell+s)!} \right]^{\frac{1}{2}} \eth^s Y_{\ell m}(\omega), \tag{6}$$

for positive semi-definite spin $0 \leq s \leq \ell$, and for negative semi-definite spin $-\ell \leq s \leq 0$ by,

$$_s Y_{\ell m}(\omega) = (-1)^s \left[\frac{(\ell+s)!}{(\ell-s)!} \right]^{\frac{1}{2}} \bar{\eth}^{-s} Y_{\ell m}(\omega). \tag{7}$$

The spin-s weighted spherical harmonics form a complete set of orthogonal basis functions which leads to the harmonic representation of a spin-s field $_s f(\omega)$ by

$$_s f(\omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} {_s}\hat{f}_{\ell m}\, _s Y_{\ell m}(\omega). \tag{8}$$

We can then trivially invert this decomposition to give the spin-s field $_s f(\omega)$ projected onto the spin basis eigenfunctions (*i.e.* the spin-spherical harmonic coefficients),

$$_s\hat{f}_{\ell m} = \int_{\mathbb{S}^2} d\Omega(\omega)\, _s f(\omega)\, _s Y_{\ell m}^*(\omega), \tag{9}$$

where the integral is over the sphere $\mathbb{S}^2$, and $d\Omega(\omega) = \sin\theta d\theta d\phi$ is the rotation invariant measure on the sphere. Typically the signal is band-limited at $\ell_{\max}$ which implies $_s f_{\ell m} = 0, \forall \ell \geq \ell_{\max}$ allowing the ℓ summations in equation (8) and the upper limit of the integral in equation (9) to be truncated at $\ell_{\max}$ to make the computation tractable.

2.2 Weak Lensing on the Sphere

This section provides a basic introduction to weak lensing mass-mapping in the spherical setting. For a more detailed introduction, we refer the reader to popular reviews (*e.g.* Bartelmann & Schneider 2001; Schneider 2005).

Gravitational lensing is an astrophysical effect which describes the deflection of distant photons as they propagate to us here and now by the intervening local matter distribution. As lensing is sensitive to the local matter distribution (both visible and dark), it provides a natural cosmological probe of dark matter.

Specifically, the weak lensing (WL) regime refers to photons which have angular position on the source plane β (relative to the line-of-sight from observer through the lensing mass) smaller than one Einstein radius θ_E to the intervening lensing mass. Mathematically this restricts us to singular solutions of the lens equation,

$$\beta = \theta - \theta_E^2 \frac{\theta}{|\theta|^2}, \quad \text{where} \quad \theta_E = \sqrt{\frac{4GM}{c^2}\frac{f_K(r-r')}{f_K(r)f_K(r')}}, \tag{10}$$

for angular diameter distance f_K, defined in the usual sense, which is dependent on the curvature of the Universe K. The Universe has been observed to be essentially flat (Planck Collaboration et al. 2018) and so to a good approximate $K \approx 0 \Rightarrow f_K(r) \approx r$, where r is the comoving distance.

Galaxies are naturally sparsely distributed across the sky and so the overwhelming majority of observations fall within the weak lensing regime (Bartelmann & Schneider 2001). Now consider a lensing potential ϕ which is the weighted integral along the line of sight of the local Newtonian potential Φ,

$$\phi(r,\omega) = \frac{2}{c^2} \int_0^r dr' \frac{f_K(r-r')}{f_K(r)f_K(r')} \Phi(r',\omega). \tag{11}$$

Poisson's equation must then be satisfied by the local Newtonian potential,

$$\nabla^2 \Phi(r,\omega) = \frac{3\Omega_M H_0^2}{2a(r)} \delta(r,\omega), \tag{12}$$

where $\delta(r,\omega)$ is the fractional over-density, H_0 is the Hubble constant, $a(r)$ is the scale-parameter and Ω_M is the matter density parameter. At first order two physical lensing quantities can be constructed, these being the gravitational shear $_2\gamma$ and the convergence $_0\kappa$ (Bartelmann & Schneider 2001; Schneider 2005), where the subscripts reflect the spin of each field.

These quantities are related to the underlying scalar integrated potential $_0\phi$ by the relations (Castro et al. 2005; Wallis et al. 2017a),

$$_0\kappa(r,\omega) = \frac{1}{4}(\eth\bar{\eth} + \bar{\eth}\eth)\, _0\phi(r,\omega), \tag{13}$$

$$_2\gamma(r,\omega) = \frac{1}{2}\eth\eth\, _0\phi(r,\omega). \tag{14}$$

If we now project these values into their harmonic representations by equation (9) we find the harmonic space relations,

$$_0\hat{\kappa}_{\ell m} = -\frac{1}{2}\ell(\ell+1)\, _0\hat{\phi}_{\ell m}, \tag{15}$$

$$_2\hat{\gamma}_{\ell m} = \frac{1}{2}\sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\, _0\hat{\phi}_{\ell m}. \tag{16}$$

We can then trivially draw a relationship between $_2\hat{\gamma}_{\ell m}$ and $_0\hat{\kappa}_{\ell m}$ in harmonic space,

$$_2\hat{\gamma}_{\ell m} = \mathcal{W}_\ell\, _0\hat{\kappa}_{\ell m}, \tag{17}$$

which is the *spherical forward model*. We have defined a mapping kernel (as in *e.g.* Wallis et al. 2017a) in harmonic space such that,

$$\mathcal{W}_\ell = \frac{-1}{\ell(\ell+1)} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}. \tag{18}$$

This mapping is analogous to the planar forward model (Price et al. 2018) but now defined on $\mathbb{S}^2$. This mapping can trivially be inverted to define the so-called 'Spherical Kaiser-Squires' (SKS, Wallis et al. 2017a) convergence estimator,

$$_0\hat{\kappa}_{\ell m}^{\mathrm{SKS}} = \mathcal{W}_\ell^{-1} \, _2\hat{\gamma}_{\ell m}^{\mathrm{obs}}, \tag{19}$$

where superscript 'obs' refers to the observations (or measurements) of a given shear field $_2\gamma$. A real-space representation of this mapping exists (Wallis et al. 2017a).

It is of interest to notice certain similarities between the SKS estimator $_0\kappa_{\mathrm{SKS}}$ and the *maximum likelihood estimator* (MLE) denoted $_0\kappa_{\mathrm{MLE}}$, which is defined by maximization of the likelihood (*i.e.* an implicit assumption of a flat prior on κ). Suppose the noise properties are assumed to be Gaussian (as is common), then the likelihood is given by

$$P(\gamma|\kappa) \propto |\Sigma|^{-1/2} \exp^{-\frac{1}{2}\chi^2}, \tag{20}$$

for $\chi^2 \equiv (\gamma - \mathbf{\Phi}\kappa)^T \Sigma^{-1}(\gamma - \mathbf{\Phi}\kappa)$ where $\mathbf{\Phi}$ is simply the forward model and Σ is the noise covariance. The solution which minimizes the likelihood is thus given by

$$_0\kappa_{\mathrm{MLE}} = (\mathbf{\Phi}^T\Sigma^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\Sigma^{-1}\gamma. \tag{21}$$

Therefore for the idealised SKS estimator $\mathbf{\Phi}^{-1}\gamma$ to be equivalent to the MLE estimator we require $\mathbf{\Phi}^{-1}$ and $(\mathbf{\Phi}^T\Sigma^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\Sigma^{-1}$ to be equivalent operators. Provided Φ is invertible the Σ terms above trivially cancel resulting in the remaining terms $(\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T$, which reduce to the idealised SKS estimator – note that in the idealised setting Φ is straightforwardly invertible given both the spherical harmonic transform and equation 18 are invertible (ignoring the monopole $\ell = 0$). Thus in this setting the idealised SKS and MLE estimator are equivalent.

However, in practical applications the forward model (*e.g.* with PSF corrections, complex masking, *etc.*) is unlikely to be invertible and hence the SKS and MLE estimators differ in practice. Furthermore, due to limited observation quality (discussed in section 5.2.1) the noise covariance is typically large in magnitude. In such settings the inverse problem is strongly ill-posed and thus significant regularization (introduced through the prior term) is required to stabilize the inversion. As such, a flat prior (MLE) results in unregularized solutions which are highly unlikely to perform well (noise present in γ is very likely to propogate directly into the κ estimate). This noise propogation is well-known, hence the SKS estimator used in practice always includes a smoothing post-processing step (convolution with an arbitrary smoothing kernel) in an attempt to mitigate this noise. Consequently, the SKS estimator used in practice does not support a principled statistical interpretation.

3   SPHERICAL BAYESIAN MASS-MAPPING

Hierarchical Bayesian frameworks facilitate a natural, mathematically principled approach to uncertainty quantification. For an elegant and approachable introduction to Bayesian methods see Trotta (2017). This section introduces Bayesian inference and proceeds to demonstrate how one may cast the spherical mass-mapping inversion as a hierarchical Bayesian inference problem. For notational ease, we drop spin subscripts on κ and γ henceforth.

3.1   Bayesian Inference

First consider the *posterior distribution* given by Bayes' Theorem,

$$p(\kappa|\gamma; \mathcal{M}) = \frac{p(\gamma|\kappa; \mathcal{M})p(\kappa; \mathcal{M})}{\int_{\mathbb{C}^N} p(\gamma|\kappa; \mathcal{M})p(\kappa; \mathcal{M})d\kappa}, \tag{22}$$

where the *likelihood function* $p(\gamma|\kappa; \mathcal{M})$ represents the probability of observing a shear field γ given a convergence field κ and some well defined model $\mathcal{M}$ (which includes both the mapping $\mathbf{\Phi} : \kappa \mapsto \gamma$ and some assumptions of the noise model). The second term in the numerator, $p(\kappa; \mathcal{M})$ is referred to as the *prior* which encodes some *a priori* knowledge as to the nature of κ. Finally, the integral denominator is the *Bayesian evidence* (or *marginal likelihood*) which can be used for model comparison, though we do not consider this within the scope of the current paper.

One approach to estimate the convergence field is given by maximizing the posterior odds conditional on the measurements γ and model $\mathcal{M}$. Such a solution is referred to as the *maximum a posteriori* (MAP) solution, κ^{map}. This can done by either maximization of the posterior or – due to the monotonicity of the logarithm function – minimization of the log-posterior,

$$\operatorname*{argmax}_\kappa \left\{ p(\kappa|\gamma; \mathcal{M}) \right\} \equiv \operatorname*{argmin}_\kappa \left\{ -\log(\, p(\kappa|\gamma; \mathcal{M})\,) \right\}. \tag{23}$$

This is a particularly helpful realization as the latter problem is more straightforward to compute and, for log-concave posteriors, allows one to to pose the problem as a convex optimization problem for which one may draw on the field of convex optimization.

3.2   Spherical Sparse Mass-Mapping

In this paper we consider the ill-posed linear inverse problem of recovering the complex discretized spherical convergence $\kappa \in \mathbb{C}^{N_{\mathbb{S}^2}}$ on the complex $\mathbb{S}^2$-sphere from a typically incomplete ($M < N$) set of M complex discretized shear measurements $\gamma \in \mathbb{C}^M$. Throughout we adopt the McEwen-Wiaux (MW) pixelization scheme, which provides theoretically exact spin spherical harmonic transforms (SSHT) due to exact quadrature (McEwen & Wiaux 2011).

We begin by defining the *measurement operator* (operator which encodes the forward model) which maps from a fiducial convergence field to the observed shear measurements

$$\mathbf{\Phi} \in \mathbb{C}^{M \times N_{\mathbb{S}^2}} : \kappa \in \mathbb{C}^{N_{\mathbb{S}^2}} \mapsto \gamma \in \mathbb{C}^M. \tag{24}$$

In the spherical setting, by noting the spherical lensing forward model given by equation (18) this measurement oper-

ator naturally takes the form,

$$\mathbf{\Phi} = \mathbf{M}\, {}_2\tilde{\mathbf{Y}}\, \mathbf{W}\, {}_0\mathbf{Y}, \tag{25}$$

where ${}_s\mathbf{Y}$ and ${}_s\tilde{\mathbf{Y}}$ represent the forward and inverse spin-s spherical harmonic transforms respectively, $\mathbf{M}$ is a masking operator, and $\mathbf{W}$ is harmonic space multiplication by the kernel $\mathcal{W}_\ell$ defined in equation (18). The adjoint-measurement operator can then be shown to be,

$$\mathbf{\Phi}^\dagger = {}_0\mathbf{Y}^\dagger\, \mathbf{W}\, {}_2\tilde{\mathbf{Y}}^\dagger \mathbf{M}^\dagger, \tag{26}$$

where it should be noted that from symmetry $\mathbf{W}$ is trivially self-adjoint. Additionally, it is important to note that adjoint (†) spin-s spherical harmonic transforms are not equivalent to the corresponding inverse spherical harmonic transforms – an important caveat often overlooked throughout the field.

3.2.1 Likelihood Function

Suppose now that measurements γ are acquired under some additive Gaussian noise $n_i \sim \mathcal{N}(0, \sigma_i^2) \in \mathbb{C}^M$ where σ_i is the noise standard deviation of a given pixel which is primarily dependent on the number of observations within said pixel, which is in turn dependent on the pixel size and number density of galaxy observations. Then the data acquisition model is simply given by

$$\gamma = \mathbf{\Phi}\kappa + n. \tag{27}$$

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

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

$$= p(\gamma|\kappa) \propto \exp\left(\frac{-\|\bar{\mathbf{\Phi}}\kappa - \bar{\gamma}\|_2^2}{2}\right), \tag{28}$$

where $\|\cdot\|_2$ is the ℓ_2-norm and $\bar{\mathbf{\Phi}} = \Sigma^{-\frac{1}{2}}\mathbf{\Phi}$ is a composition of the measurement operator and an inverse covariance weighting as defined in Section 3.2. Effectively this covariance weighting leads to measurements $\bar{\gamma} = \Sigma^{-\frac{1}{2}}\gamma$ which whiten the typically non-uniform noise variance in the observational data (shear field).

This likelihood is therefore structured to correctly account for the covariance of observational data. In this case the covariance matrix is taken to be diagonal but not necessarily proportional to the identity matrix – therefore accounting for varied numbers of observations per pixel. There are several points which should be noted. In the above we have explicitly ignored the complicating factor of intrinsic galaxy alignments which in practice would lead to non-diagonal covariance. This extension can easily be supported, given a sound understanding of the effects of intrinsic alignments on the data covariance (which in practice may be challenging).

Additionally here we, for simplicity, assume each pixel contains a sufficient number of galaxy observations that a *central limit theorem* (CLT) argument for pixel noise can

be justified. Largely this assumption is acceptable, however as the resolution increases (pixel size decreases) the noise becomes increasingly non-Gaussian.

Finally, the forward model considered here (Section 3.2) begins from κ and ends at masked, gridded γ measurements, however there are several steps which must take place before one acquires such measurements. One may therefore wish to extend this model to incorporate such complicating factors as pixelisation effects, reduced shear (see section 3.3.1), point squared function (PSF) errors *etc.*

It should then be explicitly noted that this mass-mapping formalism requires only that the posterior belong to the (rather comprehensive) set of log-concave functions, and as such one can directly interchange the noise model or introduce complicating factors where desired provided the posterior remains log-concave.

3.2.2 Prior Function

As this inverse problem is ill-posed (often seriously), maximum likelihood estimators (MLE) are sub-optimal and must be regularized by some prior assumption as to the nature of the convergence field. In this work we select a sparsity promoting, Laplace-type prior in the form of the ℓ_1-norm $\|.\|_1$ – though as discussed in section 3.2.1 this formalism supports any log-concave priors of which there are many to choose from (*e.g.* most exponential family priors).

Laplace-type priors are often adopted when one wishes to promote sparsity in a given dictionary or basis. Wavelets $\mathbf{\Psi}$ are localised in both the frequency and spatial domains and thus constitute a naturally sparsifying dictionary for most physical signals. There are several wavelet constructions on the sphere that may be considered (see *e.g.* Schröder & Sweldens 1995; Barreiro et al. 2000; Starck et al. 2006; McEwen & Scaife 2008; McEwen et al. 2011; Wiaux et al. 2008; McEwen et al. 2018; Narcowich et al. 2006; Baldi et al. 2009; Marinucci et al. 2007; McEwen & Price 2019; Chan et al. 2017b) with varying localisation and uncorrelation properties. In this paper we adopt the scale-discretised wavelets (Wiaux et al. 2008; Leistedt et al. 2013a; McEwen et al. 2013b, 2015a) scheme as not only does it satisfy qausi-exponential loctalisation and asymptotic uncorrelation properties (McEwen et al. 2018) but also supports directionality which may often be of interest for the weak lensing setting.

We specifically adopt a Laplace-type wavelet log-prior $\|\mathbf{\Psi}(\cdot)\|_1$. Note that as $\|\cdot\|_1$ is a discretization of the continuous ℓ_1-norm it must be reweighted by wavelet pixel size, which in practice is as simple as multiplying a given wavelet coefficient by a factor proportional to $\sin(\theta)$ where θ is the angular deviation of the given pixel from the pole. Throughout this paper any reference to the ℓ_1-norm applied to a spherical space refers explicitly to this spherically reweighted norm.

With our choice of ℓ_1-norm regularization the prior can be written compactly as

$$p(\kappa) \propto \exp\left(-\mu\|\tilde{\mathbf{\Psi}}^\dagger\kappa\|_1\right), \tag{29}$$

where $\tilde{\mathbf{\Psi}}^\dagger$ is the analysis forward-adjoint spherical wavelet transforms (see equation A8 in the appendix) with coefficients $\tilde{\Psi}_i^\dagger$, and $\mu \in \mathbb{R}^+$ is the regularization parameter. It

is assumed here that the spherical wavelet dictionary $\tilde{\boldsymbol{\Psi}}$ is a naturally sparsifying dictionary for the convergence field defined on the sphere. In practice one may select whichever dictionary one's prior knowledge of the convergence indicates is likely to be highly sparsifying.

Conceptually, a sparsity-promoting prior can be though of as a mathematical manifestation of *Occam's Razor* – the philosophical notion that the simplest answer is usually the best answer. Mathematically, this is equivalent to down-weighting solutions with large numbers of non-zero coefficients, which may match the noisy data perfectly, in favour of a less perfect match but with significantly fewer non-zero coefficients.

Alternatively, one may view sparsity priors (in this context) as a relative assumption of the sparsity of the true signal and noise signal when projected into a sparsifying dictionary. This is to say that the assumption is that the noise signal will be less sparse in $\tilde{\boldsymbol{\Psi}}$ than the true signal. Typically noise signals are relatively uniformly distributed in wavelet space, whereas most physical signals are sparsely distributed and therefore this relative interpretation of the sparsity prior makes reasonable sense (Mallat 1999).

Note that the only constraint on the posterior is that it must be log-concave (such that the log-posterior is convex). Hence one can select any log-concave prior within this framework, *e.g.* one could select an ℓ_2-norm prior which with minor adjustments produces Wiener filtering (see Horowitz et al. 2018, for alternate iterative Wiener filtering approaches), or a flat prior which produces the *maximum likelihood estimate* (MLE).

3.3 Implementation

The minimization of the log-posterior in equation (23) is (in the analysis setting) therefore precisely the same as solving,

$$\kappa^{\mathrm{map}} = \underset{\kappa}{\mathrm{argmin}} \underbrace{\left\{ \mu \|\boldsymbol{\Psi}^{\dagger}\kappa\|_1 + \frac{\|\bar{\boldsymbol{\Phi}}\kappa - \bar{\gamma}\|_2^2}{2} \right\}}_{\text{Objective function}}. \tag{30}$$

The bracketed term on the RHS is referred to as the *objective function*. We solve this convex optimization problem using the S2INV (Price et al. 2020) code which is largely built around the SOPT C++ object oriented framework[1] (Pratley et al. 2018; Onose et al. 2016; Carrillo et al. 2012, 2013), utilizing an adapted proximal forward-backward splitting algorithm (Combettes & Pesquet 2009), although a variety of alternate algorithms are provided within S2INV. Wavelet transforms on the sphere are computed using S2LET[2] (McEwen & Wiaux 2011; Chan et al. 2017a; Leistedt et al. 2013a; McEwen et al. 2015b,a; McEwen et al. 2018), which in turn makes use of SSHT[3] (McEwen & Wiaux 2011; McEwen et al. 2013a) to compute spherical harmonic transforms and SO3[4] (McEwen et al. 2015a) to compute Wigner transforms.
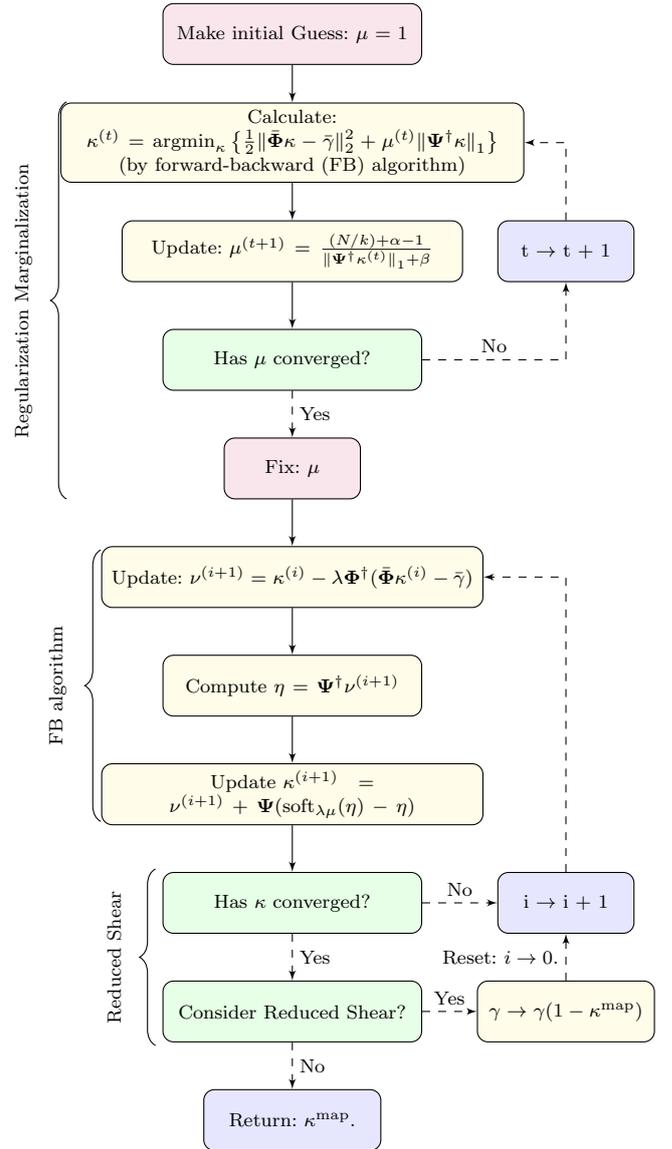
Figure 1. Schematic of proximal forward-backward splitting algorithm used (Combettes & Pesquet 2009). Note that the first iterative block represents the Majorize-Minimization (MM) algorithm marginalization over the regularization parameter (which here is treated as a nuisance parameter), the second itertaive block represents the primary proximal forward backward iterations, and the final (optional) block represents the reduced shear outer iterations. Note that the soft$_{\lambda,\mu}(\eta)$ operation is the soft thresholding operation, which is the proximal projection of the ℓ_1-norm (see *e.g.* Pereyra 2017; Cai et al. 2018a,b, for details).

To deal with the non-differentiable ℓ_1-norm prior, gradient operators ∇ are in some sense replaced by proximal operators when applied to the non-differentiable term (Moreau 1962). The iteration steps are provided in the schematic of Figure 1, for full details of the derivation of the proximal forward-backward algorithm iterations look to Combettes & Pesquet (2009). These primary optimizations are terminated once the objective function is updated by less than a set threshold (in our experiments 10^{-6}) between iterations.

3.3.1 Reduced shear

Figure 1 displays a schematic representation of the steps taken in computing κ^{map}. A degeneracy between the convergence field κ and shear field γ exists, and as such γ is not a true observable. Instead the *reduced shear g* is the true observable where,

$$g(\omega) = \frac{\gamma(\omega)}{1 - \kappa(\omega)}. \tag{31}$$

When working sufficiently within the weak lensing regime $\kappa \ll 1$ and $\gamma \approx g \ll 1$. Although typically the reduced shear need not be accounted for, for completeness we correct for the reduced shear (Mediavilla et al. 2016; Wallis et al. 2017a; Price et al. 2018). We add correcting iterations outside our primary iterations to maintain the linearity of the overall reconstruction. Our reduced shear correction iterations are displayed schematically in the final loop of Figure 1.

Reduced shear iterations are deemed to have converged once the convergence update $\max_j |\kappa_j^{(i)} - \kappa_j^{(i+1)}| < 10^{-10}$ where j runs over all pixels (as in Wallis et al. 2017a).

3.3.2 Bayesian Regularization Parameter

For recovered statistics to be truly principled, the regularization parameter must necessarily be folded into the hierarchy or correctly marginalized over. One way to do this was recently developed (Pereyra et al. 2015) and shown to work well in the planar weak lensing setting (Price et al. 2018).

This Bayesian hierarchical inference approach assumes a gamma distribution hyper-prior

$$p(\mu) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\beta\mu} \mathbb{I}_{\mathbb{R}^+}(\mu), \tag{32}$$

with weakly dependent hyper-parameters α and β which without loss of generality can be fixed at $\alpha = \beta = 1$. We then iterate (Pereyra et al. 2015) to effectively marginalize over μ which is treated as a nuisance parameter in the main body of our hierarchy. These iterations are,

$$\kappa^{(t)} = \underset{\kappa}{\mathrm{argmin}} \left\{ \frac{1}{2} \|\bar{\bar{\boldsymbol{\Phi}}}\kappa - \bar{\gamma}\|_2^2 + \mu^{(t)} \|\boldsymbol{\Psi}^\dagger \kappa\|_1 \right\}, \tag{33}$$

$$\mu^{(t+1)} = \frac{(N/k) + \alpha - 1}{\|\boldsymbol{\Psi}^\dagger \kappa^{(t)}\|_1 + \beta} \tag{34}$$

where the log-prior $\|\boldsymbol{\Psi}^\dagger \kappa\|_1$ is k-homogeneous. Note that a sufficient statistic (log-prior) is k-homogeneous if $\exists\, k \in \mathbb{R}^+$ such that,

$$f(\eta x) = \eta^k f(x), \ \forall x \in \mathbb{R}^M, \ \forall \eta > 0. \tag{35}$$

Further note that all norms, composite norms, and compositions of norms with linear operators are 1-homogeneous, *i.e.* $k = 1$. See Pereyra et al. (2015) for further details. These regularization marginalization iterations are terminated when the update to the regularization parameter is less than 1% *i.e.* $|\mu^{(i+1)} - \mu^{(i)}|/\mu^{(i)} < 0.01$.

3.3.3 Computational efficiency

As discussed in 3.3 all iterations consist of a forward step which includes application of the measurement operator before computing the data fidelity term, followed by the backward step which includes application of the spherical wavelet transform.

The measurement operator is dominated by the spin spherical harmonic transforms which scale as $\mathcal{O}(L^3)$. Similarly the computational efficiency of the wavelet transform is dominanted by underlying harmonic transforms, however with directionality N (*i.e.* wavelet on the rotation group) the transform scales as $\mathcal{O}(N \times L^3)$. The overall forward-backward algorithm scales additively as $\mathcal{O}(K \times (N+1) \times L^3) \sim \mathcal{O}(K \times N \times L^3)$ where K is the total number of iterations required for convergence.

The SKS operator also requires the application of spin spherical harmonic transforms and therefore scales as $\mathcal{O}(L^3)$. However the SKS method requires only a single application of the transform and thus the ratio of computational efficiency between the two algorithms effectively scales as $\mathcal{O}(K \times N)$ – which is to say the difference in computational efficiency is primarily determined by the choice of wavelet complexity and the magnitudes of the associated convergence criteria.

In practice, including the marginalization preliminary iterations and subsequent annealing iterations to optimize convergence, we find $\mathcal{O}(10^2)$ iterations are sufficient for convergence. We consider axisymmetric wavelets ($N = 1$) in this article, thus the DarkMapper algorithm is $\mathcal{O}(10^2)$ times slower than SKS but with greatly superior reconstruction performance and the ability to quantify uncertainties in a statistically principled manner.

It is interesting to note that MCMC methods typically require a very large number of samples, with each individual sample requiring at least one spin spherical harmonic transform. Therefore the increase in computational efficiency of this approximate Bayesian inference over sampling methods is roughly given by $\mathcal{O}(n_{\mathrm{samples}}/10^2)$ where n_{samples} is the total number of samples required for convergence of a given MCMC sampling method. As MCMC methods often require at least $\mathcal{O}(10^6)$ this increase in computation speed is (many) orders of magnitude. In the spherical setting an $\mathcal{O}(10^4)$ increase in computation speed results in computations which would take $\mathcal{O}(\text{decades})$ taking $\mathcal{O}(\text{days})$.

4 BAYESIAN UNCERTAINTY QUANTIFICATION

Though MAP estimates provide high fidelity estimates of the convergence field uncertainties on these estimates are a necessity if one aims to make statistically principled inferences. Generally, for scientific inference one should prioritize principled uncertainties over image aesthetics.

Bayesian inference approaches as presented in Section 3 provide principled statistical frameworks through which quantification of uncertainty on recovered statistics comes naturally from the posterior. Typically the posterior cannot be evaluated analytically, and so *Markov Chain Monte Carlo* (MCMC) sampling methods must be used. In moderate to low dimensional settings for computationally cheap functions, MCMC chains can feasibly be computed. However, in high dimensional spherical settings, MCMC techniques quickly become challenging to compute.

Bespoke MCMC techniques have been developed for the weak lensing setting (*e.g.* Schneider et al. 2015; Corless et al.

2009; Alsing et al. 2016) which can improve computational efficiency, yet these methods will find it challenging to accommodate future 'Big Data' from high-resolution wide-field surveys. Furthermore, these sometimes come with additional restrictions (*e.g.* some are restricted to Gaussian priors). This provides strong motivation for the development of fast, approximate Bayesian inference approachs Pereyra (2017); Cai et al. (2018b); Price et al. (2018); Price et al. (2019b,a), the uncertainty quantification of which we extend to the complex $\mathbb{S}^2$-sphere and present in this section.

4.1   Highest Posterior Density Region

At $100(1-\alpha)\%$ confidence a sub-set $C_\alpha \in \mathbb{C}^{N_{\mathbb{S}^2}}$ of the posterior space is considered a credible region of the posterior *iif* the integral equation,

$$p(\kappa \in C_\alpha | \gamma) = \int_{\kappa \in \mathbb{C}^{N_{\mathbb{S}^2}}} p(\kappa|\gamma) \mathbb{I}_{C_\alpha} d\kappa = 1 - \alpha, \qquad (36)$$

is satisfied, where we have used the set indicator function $\mathbb{I}_{C_\alpha}$, defined to be,

$$\mathbb{I}_{C_\alpha} = \begin{cases} 1 & \text{if,} \quad \kappa \in C_\alpha \\ 0 & \text{if,} \quad \kappa \notin C_\alpha. \end{cases} \qquad (37)$$

Theoretically there are infinitely many regions which satisfy the integral in equation (36). However, the decision-theoretic optimal region – in the sense of minimum volume – is the *Highest Posterior Density* (HPD) credible-region, which is given by (Robert 2001)

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

where the combination $f(\kappa) + g(\kappa)$ is our objective function derived in Section 3.3, and ϵ_α is an isocontour (*i.e.* level-set) of the log-posterior.

However, in high dimensional ($N \gg 1$) settings ϵ_α (and therefore C_α) becomes particularly difficult to compute, thus motivating the development of alternate approaches that are fast and approximate. Recent advances in probability concentration theory have led to the derivation (Pereyra 2017) of a conservative approximate credible-region C'_α for log-concave distributions. This approximate region is defined as,

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

such that,

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

is the approximate level-set threshold with constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$. Recall that N is the dimension of $\kappa \in \mathbb{C}^{N_{\mathbb{S}^2}}$ which for equiangular spherical sampling (MW; McEwen & Wiaux 2011) is given by,

$$N_{\mathrm{MW}} \equiv \ell_{\max}(2\ell_{\max} - 1) \approx 2\ell_{\max}^2, \qquad (41)$$

for signals with angular band-limit $\ell_{\max}$. An upper bound on the error introduced through this approximation has been shown to exist (Pereyra 2017) and is given by,

$$0 \le \epsilon'_\alpha - \epsilon_\alpha \le \eta_\alpha \sqrt{N} + N, \qquad (42)$$

where $\eta_\alpha = \sqrt{16 \log(3/\alpha)} + \sqrt{1/\alpha}$. This error scales at most linearly with N and in high dimensional settings can be

somewhat large, though in practice we find this error upper-bound to be extremely conservative (Price et al. 2019b).

Note that the error is positive semi-definite which corroborates the assertion that C'_α is a conservative approximation. Mathematically, this is to say that the true HPD credible region C_α is sub-set of the approximate HPD credible region C'_α *i.e.* $C_\alpha \subseteq C'_\alpha$. This ensures that if some convergence field $\kappa \notin C'_\alpha$ then necessarily $\kappa \notin C_\alpha$.

Further note that although we adopt the approximate level-set threshold derived in Pereyra (2017) in this work, research into these types of bounds is a relatively new area of study. Thus, if and when new (more constraining) bounds are derived they can trivially be substituted here.

4.2   General Application

Having introduced the concept of an approximate HPD-credible region C'_α of the posterior, the question then immediately arises as to how one can utilize this information in practice. In MCMC sampling type approaches, one may simply use the recovered samples to quantify uncertainty at well defined confidence on specific properties of the recovered posterior. In our setting, we have recovered only the MAP solution in a form which supports trivial computation of the approximate level-set threshold, and thus C'_α.

With such limited posterior knowledge one may only ask whether a surrogate solution (an adjusted convergence map) does or does not belong to C'_α. In effect this is to say that in our formalism all questions of the posterior must be cast as Bayesian hypothesis tests of varying complexity (Price et al. 2018; Price et al. 2019a). Some examples are provided in the following subsections.

4.2.1   Bayesian hypothesis testing

A Bayesian hypothesis test on the posterior (see Figure 2) is simply: the MAP convergence is recovered, a feature of that map is removed[5] to form a surrogate map κ^{sur}, if $\kappa^{\mathrm{sur}} \notin C'_\alpha$, then $\kappa^{\mathrm{sur}} \notin C_\alpha$, and thus the hypothesis that feature of interest is insignificant is rejected at some well defined confidence, implying that the feature cannot be deemed insignificant at said confidence (for more details look to Cai et al. 2018b; Price et al. 2018).

One can invisage constructing substantially more complicated uncertainty quantification techniques *via* iterative application of Bayesian hypothesis testing or by more complicated individual Bayesian hypothesis tests.

4.2.2   Local credible intervals

The next most straightforward uncertainty quantification technique is given by the notion of local credible intervals (Cai et al. 2018b; Price et al. 2019b) which are in effect pixel-level Bayesian uncertainty (error) bars on recovered maps. Conceptually these are formed by splitting the recovered MAP estimate into superpixels (groups of adjacent pixels), then within each superpixel (keeping all other pixels fixed at their MAP values) iteratively increasing (decreasing) the

---

[5] In practice one may simply adjust κ^{sur} to suit a specific question of the posterior.

recovered pixel intensity and thus constructing surrogate solutions, checking whether these surrogate solutions belong to C'_α. Once the maximum (minimum) super-pixel intensity is located (typically *via* bisection) the difference (maximum - minimum) is taken to be the range of values which cannot be rejected for a recovered super-pixels intensity – hence the notion of these representing pixel level Bayesian uncertainty (error) bars.

4.2.3 Uncertainty quantification of global features

For science, in particularly for cosmology, it is often perhaps more informative to leverage the concept of Bayesian hypothesis testing to consider global structure, and therefore consider global (or aggregate) statistics of a recovered field. To do so one must simply define a logically consistant algorithm which constructs surrogate convergence solutions that are representative of the global question they wish to ask of the recovered convergence field, after which the process follows in much the same way as demonstrated for forming local credible intervals.

It should, however, be noted that one must be careful how one poses these global questions, as the questions of interest are often inherently non-convex and must be solved *via* decision theory methods. A good example of how one can apply hypothesis testing to global structure can be found in Price et al. (2019a) where the Bayesian uncertainty in the aggregate peak statitic is recovered.

Here we have discussed only a few possible uncertainty quantification techniques which are supported by this formalism, though in practice following the methodology outlined above one can form uncertainty quantification techniques around a far more comprehensive set of global features (or equally statistics) provided a few important caveats are understood: the process of Bayesian hypothesis tests suggested to quantify a specified uncertainty are well defined and clearly explained, the limitations of any method are fully acknowledged, and the results are interpretted correctly so as to mitigate unjustified statistical statements. We present a specific example on current cosmic shear data in Section 6.

4.2.4 The curse of dimensionality

Finally it is academic to note that the concept of changing only a small number of pixels of a given map whilst fixing the remaining at their MAP values is explicitly recovering conditional probabilities which are by definition the largest possible uncertainties. Though this is precisely what one requires of such approximations it highlights an inherent drawback of such approaches. As the approximate level-set threshold scales with the total dimension of the inference in high dimensional cases, the uncertainty of any individual local structure within an image becomes large.

Conceptually this makes sense as the higher the dimensionality of the problem, the more statistical fluctuations occur and thus the higher the chance that a statistical fluctuation produced the feature of interest. As such, for anything higher than moderate dimensional settings local uncertainties become very large and one should prioritize global or aggregate statistics.
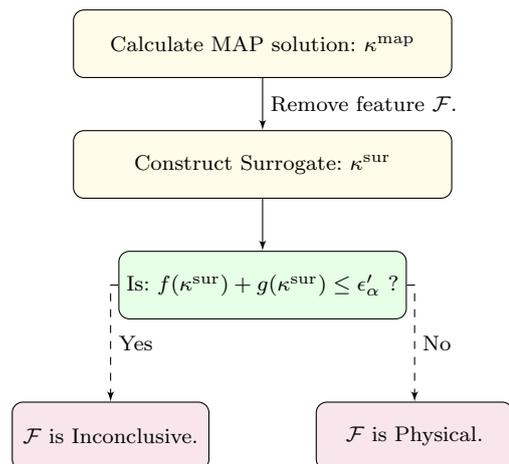
Figure 2. Schematic of hypothesis testing (Price et al. 2018). The feature $\mathcal{F}$ is entirely general and can be constructed by any well defined operator on the MAP solution κ^{map}.

5 SIMULATIONS

In this section we apply the spherical Kaiser-Squires (SKS) estimator , both with and without post-processing smoothing, and the spherical sparse hierarchical Bayesian (DarkMapper) estimator developed in this article to a range of realistic N-body simulations which are masked throughout by a pseudo-Euclid mask so as to best match upcoming Stage IV surveys.

5.1 Data-set

Throughout this article we perform reconstructions and uncertainty quantification on simulated convergence maps generated from the high resolution Takahashi N-body simulation datasets (Takahashi et al. 2017)[6]. These mock convergence maps are generated *via* multiple-lens plane raytracing, and are provided for a range of comoving distances. Specifically, simulated convergence maps are presented at every 150Mpc/h for redshift $z_s \in [0.05, 5.3]$. The cosmological parameters selected for this suite of simulations are $\Omega_m = 1, \Omega_\Lambda = 0.279, \Omega_{\mathrm{cdm}} = 0.233, \Omega_b = 0.046, h = 0.7, \sigma_8 = 0.82$ and $ns = 0.97$ which are consistent with the WMAP 9 year result (Hinshaw et al. 2013).

We select redshift slice 16 which corresponds to the slice with redshift $z_s \sim 1$. To mitigate the Poisson noise present in such N-body snapshots we convolve the Takahashi convergence with a very small smoothing kernel sufficient only to remove the noise whilst adjusting the signal as little as possible. Finally we apply a pseudo-Euclid masking (a straightforward masking of the galactic plane and the ecliptic) so as to best mimic the setting of upcoming Stage IV surveys.

5.2 Methodology

As in previous work (Price et al. 2018; Price et al. 2019b,a) we begin by applying the measurement operator $\mathbf{\Phi}$ (see

---

[6] These datasets can be found at http://cosmo.phys.hirosaki-u.ac.jp/takahasi/allsky_raytracing/.

$\ell_{\max} = 2048$		SKS		SKS (optimal smoothing)		DarkMapper		Difference
Setting	n_{gal}	SNR (dB)	$P_{\mathrm{correlation}}$	SNR (dB)	$P_{\mathrm{correlation}}$	SNR (dB)	$P_{\mathrm{correlation}}$	Δ SNR (dB)
Stage III	5	-9.792	0.403	0.962	0.759	5.494	0.904	+15.286 (+4.532)
	10	-6.794	0.532	1.108	0.806	7.299	0.935	+14.093 (+6.191)
Stage IV	30	-2.091	0.732	1.254	0.854	9.767	0.964	+11.858 (+8.513)
Idealized	100	2.956	0.887	n/a	n/a	12.132	0.980	+9.176 (n/a)

Table 1. Numerical results from reconstructions of Takahashi simulations as discussed in Sections 5.1 and 5.2. In each case the DarkMapper estimator drastically outperforms both the SKS estimator and the optimally smoothed SKS estimator (which cannot in practice be achieved due to *ad hoc* smoothing kernel selection) in both recovered signal to noise ratio (SNR) and the Pearson correlation coefficient $P_{\mathrm{correlation}}$. Highlighted are the results most representative of the imminent Stage IV surveys, such as Euclid and LSST. As Stage IV surveys forecast large sky fractions to avoid projection effects (Wallis et al. 2017a; Vallis et al. 2018) mass-mapping must be performed natively on the sphere. Thus this spherical mass-mapping formalism is, at least currently, the optimal choice for Stage IV weak lensing mass-mapping. Note that no post-processing by smoothing increased the recovered SNR for the idealized $n_{\mathrm{gal}} = 100$ setting for the SKS estimator and so was recorded as n/a.

equation 25) to the fiducial ground truth, full-sky Takahashi convergence map κ to create artificial masked clean shear measurements $\gamma \in \mathbb{C}^M$.

A noise standard deviation σ_i is computed (see Section 5.2.1) for each pixel i individually and used to construct known diagonal covariance Σ.[7] Hence we create noisy simulated shear observations $\gamma_n = \gamma + n$ and a simulated data covariance Σ which would in practice be provided by the observation team – this covariance is defined by the number of galaxy observations within a given pixel of the sky.

We then apply the standard SKS estimator and the DarkMapper estimator presented in this paper to these noisy artificial measurements γ_n to create estimates of the fiducial convergence map κ. For DarkMapper we simply adopt diadic axisymmetric spherical wavelets ($N = 1$ and $\lambda = 2$ for simplicity), with scale-discretised harmonic tiling (McEwen et al. 2018) (adopting minimum wavelet scale $j_0 = 0$ and maximum wavelet scale $j_{\max} = 10$ resulting in a total of 11 wavelet scales). Additional complexity may produce better results at the cost of computational efficiency. Furthermore scale-discretised wavelets are only one possible choice of spherical wavelets (see Section 3.2.2). Other wavelets on the sphere could be adopted and are interchangable within this reconstruction formalism, provided they support exact synthesis of a signal from its wavelet coefficients.

We adopt the signal to noise ratio (SNR) as a metric to compare how closely each convergence estimator matches the true convergence map. This recovered SNR in decibels (dB) is defined to be,

$$\text{Recovered SNR} = 20 \times \log_{10}\left(\frac{\|\kappa\|_2}{\|\kappa - \kappa^{\mathrm{map}}\|_2}\right), \quad (43)$$

from which it is clear that the larger the recovered SNR the more accurate[8] the convergence estimator. Additionally we record the Pearson correlation coefficient between recovered convergence estimators $\kappa^{\mathrm{map}} \in \mathbb{C}^{N_{\mathbb{S}^2}}$ and the fiducial convergence $\kappa \in \mathbb{C}^{N_{\mathbb{S}^2}}$ as a measure of topological fidelty of the

estimator. The Pearson correlation coefficient is defined to be

$$r = \frac{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa^{\mathrm{map}}(i) - \bar{\kappa}^{\mathrm{map}}\}\{\kappa(i) - \bar{\kappa}\}}{\sqrt{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa^{\mathrm{map}}(i) - \bar{\kappa}^{\mathrm{map}}\}^2}\sqrt{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa(i) - \bar{\kappa}\}^2}}, \quad (44)$$

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

In practice the SKS estimator (as with its predecessor the KS estimator) is post-processed *via* axisymmetric convolution with an often quite large Gaussian smoothing kernel. The absolute scale of this kernel is typically chosen 'by eye' (which is to say arbitrarily), but in order to maximise the performance of the SKS estimator we iteratively compute the smoothing scale which maximises the recovered SNR, yielding the best possible reconstruction that can be provided by the SKS estimator (*i.e.* with optimal smoothing). We then use this optimal SKS estimator for comparison. Note that this may only be performed in simulation settings where the fiducial convergence is known. Further note that such *ad hoc* parameters do not exist within the DarkMapper formalism, for which a principled statistical problem is posed and solved by automated optimization algorithms.

5.2.1 Noise

For weak-lensing surveys the noise level of a given pixel is dependent on: the number density of galaxy observations n_{gal} (typically given per arcmin^2), the size of said pixel, and the variance of the intrinsic ellipticity distribution σ_e^2.

Knowing the area A of a given pixel the noise standard deviation σ_i is simply given by,

$$\sigma_i = \sqrt{\frac{\sigma_e^2}{A \times (180/\pi)^2 \times 3600 \times n_{\mathrm{gal}}}}, \quad (45)$$

where $3600(180/\pi)^2$ converts steradians to arcmin^2 – this relation is simply a reduction in the noise standard deviation by the root of the number of data-points. Thus, larger pixels which (assuming a roughly uniform spatial distribution of galaxy observations) contain more observations have smaller noise variance. In practice the value of σ_i (and therefore the

[7] Note we here do note consider off diagonal terms which may arise due to intrinsic galaxy alignments though in future this can be incorporated straightforwardly.

[8] Accuracy here is in regard to the pixel-level deviation not structural correlation, for which specific estimators may be designed.

covariance Σ) can be determined using the true number of galaxies in a given pixel rather than n_{gal}.

The typical intrinsic ellipticity standard deviation is $\sigma_e \sim 0.37$. Upcoming Stage IV surveys (*e.g.* Euclid (Laureijs et al. 2011) and LSST) are projected to achieve a number density of $n_{\text{gal}} \sim 30$ per arcmin^2 – a soft limit due to blending complications. For academic discussion we also consider the case of a potential future space-based survey which may push the number density as high as $n_{\text{gal}} \sim 100$ per arcmin^2, in addition to lower number densities $n_{\text{gal}} \in [5, 10]$ per arcmin^2 which are representative of past Stage III surveys.

5.3 Reconstruction results

For an angular bandlimit $\ell_{\text{max}} = 2048$, a pseudo-Euclid mask and input $n_{\text{gal}} \in [5, 10, 30, 100]$ we compute the spherical Kaiser-Squires (SKS) estimator, an idealized (optimally smoothed) SKS estimator, and the DarkMapper estimator. The results can be found in Figure 3 and numerically in Table 1. In all cases the DarkMapper estimator provides the highest reconstruction fidelity both in terms of recovered SNR and Pearson correlation coefficient. Note that for $\ell_{\text{max}} = 2048$ and the number density of galaxy observations selected the mean number of galaxies per pixel is $\mathcal{O}(10-10^3)$.

It is important to note that the optimal smoothing kernel for the SKS estimator cannot be known and thus in practice is often selected 'by eye' which is to say selected *ad hoc*. Therefore the smoothed SKS results here constitute an upper bound. The DarkMapper framework is fully principled and requires no *ad hoc* parameter selection and is therefore likely to perform in much the same way when applied to observational data.

For Stage III survey settings with $n_{\text{gal}} = 5, 10$ the increase in SNR (Δ SNR) of the DarkMapper estimator over the SKS (optimally smoothed SKS) estimator was +15.286 (+4.532) dB and +14.093 (+6.191) dB respectively. Recall that dB is measured on a logarithmic scale (see equation 43) and so this increase is quite dramatic. Furthermore the Pearson correlation coefficient increased from 0.403 (0.759) to 0.904 and 0.532 (0.860) to 0.935 for $n_{\text{gal}} = 5, 10$ respectively.

For the Stage IV Euclid-type setting with $n_{\text{gal}} = 30$ Δ, SNR was found to be +11.858 (+8.513) dB, along with which the Pearson correlation coefficient rose from 0.723 (0.854) to 0.964. As this setting is highly representative of the observations which will be made in Stage IV surveys this strongly suggests that algorithms such as DarkMapper should be adopted for weak lensing mass-mapping.

6 APPLICATION TO PUBLIC DATA

Finally we apply both the SKS and DarkMapper estimators to a collated map of the majority of the public wide field weak lensing observational datasets in order to reconstruct a single global dark-matter mass-map computed natively on the sphere. Furthermore we demonstrate straightforward global uncertainty quantification on our reconstruction.

Specifically we perform convergence reconstructions on the DESY1 (Abbott et al. 2018; Morganson et al. 2018; Flaugher et al. 2015), CFHTLens (Erben et al. 2012), and

the KiDS450 (Hildebrandt et al. 2017; Fenech Conti et al. 2017) weak lensing shear datasets. See specific acknowledgements and related papers for further details. Note that throughout we have not chosen to perform reduced shear iterations, assuming that the observed shear is approximately the reduced shear $\gamma \sim g$ (in a more detailed analysis one could perform such further iterations)

6.1 Joint Spherical Mass-Map

All aforementioned weak lensing shear observational datasets were collated into a single joint global dataset. For each data-set we select only galaxies with non-zero catalog weight $w(i) > 0$ and perform a correction for the multiplicative bias by $w(i)$ and additive by $c_{1,2}(i)$ biases per observation. Specifically this correction for ellipticities $e_{1,2}(i)$ is given by

$$\Re\{\gamma(i)\} = \frac{\sum_j w(j)\{e_1(j) - c_1(j)\}}{\sum_j w(j)\{1 + \text{mcorr}(j)\}},$$
$$\Im\{\gamma(i)\} = \frac{\sum_j w(j)\{e_2(j) - c_2(j)\}}{\sum_j w(j)\{1 + \text{mcorr}(j)\}}, \quad (46)$$

where j are observations such that observation j belongs to pixel i, mcorr is the catalog magnification correction and $\Re, \Im$ denote the real and imaginary components of the shear field $\boldsymbol{\gamma}$ respectively.

This joint global dataset was then projected onto an equiangularly sampled (MW) spherical shear map γ_{obs} with an angular bandlimit of $\ell_{\text{max}} = 2048$. During this projection the number of galaxies projected into each pixel was recorded to create a complimentary map of observations per pixel, from which the data covariance Σ_{obs} is directly determined (as discussed in Sections 3.2.1 and 5.2.1)

To this spherical shear map γ_{obs} (with corresponding data covariance Σ_{obs}) we apply DarkMapper outlined in Section 3 with the same parameter choices outlined in Section 5 (see Price et al. 2018, for a planar equivalent). Additionally, we provide the SKS (Wallis et al. 2017a) reconstruction which we present in both its fundamental form (without post-processing Gaussian smoothing) and in its typical form (with post-processing Gaussian smoothing with full width at half maximum FWHM = Θ = 25 arcmins). All data products aforementioned within this section are publicly available and may be found at `https://doi.org/10.5281/zenodo.3980652`.

The results of all reconstruction algorithms can be seen globally in Figure 4 and with enhanced regions in Figure 5, where all subplots share the same colourscale (Green 2011). It is immediately apparent that the SKS estimator, in the absence of smoothing, is overwhelming dominated by noise (hence the motivation for post-processing).

In contrast to this, the SKS estimator with a Θ = 25 arcmin post-processing Gaussian smoothing is largely in agreement with the DarkMapper estimator, however this smoothed SKS estimator unsurprisingly lacks any significant small-scale structure. Further note that the smoothed SKS estimator does not mirror all high intensity structure (*e.g.* peaks and voids) recovered by the DarkMapper estimator, which indicates more significant deviations between the two estimators. The most egregious of these cases is highlighted in the red boxed KiDS450 patch of Figure 5.
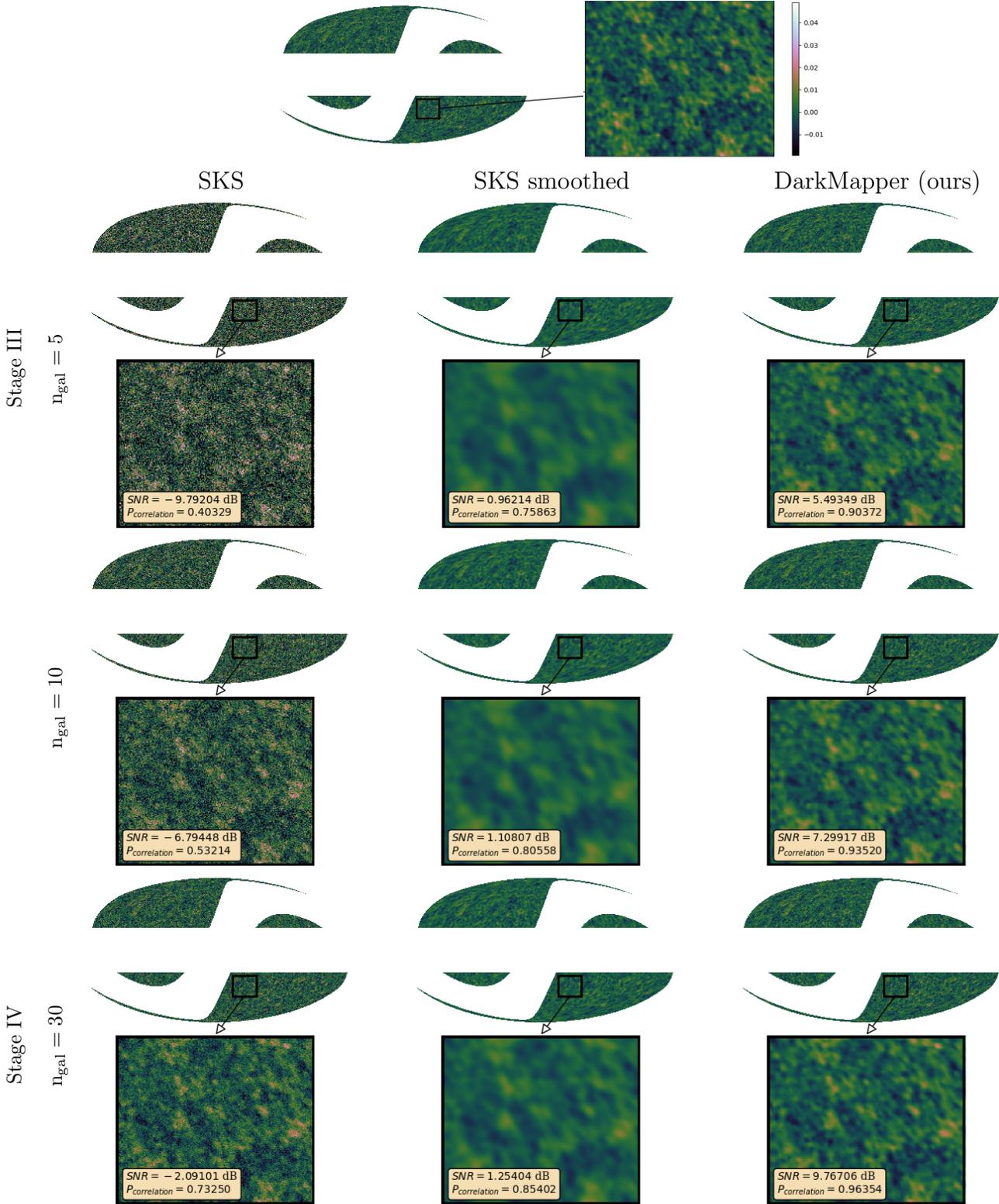
Figure 3. The top row displays the ground truth Takahashi convergence map as described in Section 5.1 with close up of a small region. **Top to bottom:** Increasing number density of galaxies (n_{gal}) and therefore decreasing noise levels. At the top we have $n_{\mathrm{gal}} = 5$ which is representative of current Stage III surveys, at the bottom we have $n_{\mathrm{gal}} = 30$ which has been forecast for upcoming Stage IV surveys (*e.g.* Euclid / LSST). **Left to right:** The spherical Kaiser-Squires (Wallis et al. 2017a) estimator without the *ad hoc* smoothing kernel postprocessing, the optimally smoothed spheircal Kaiser-Squires estimator, and finally the DarkMapper estimator. **Discussion:** Clearly the DarkMapper estimator is visibly superior in all cases, numerically recovering both significantly larger SNR and Pearson correlation coefficients. All reconstructions have are plotted on the same colorscale to aid comparison (Green 2011).

SKS
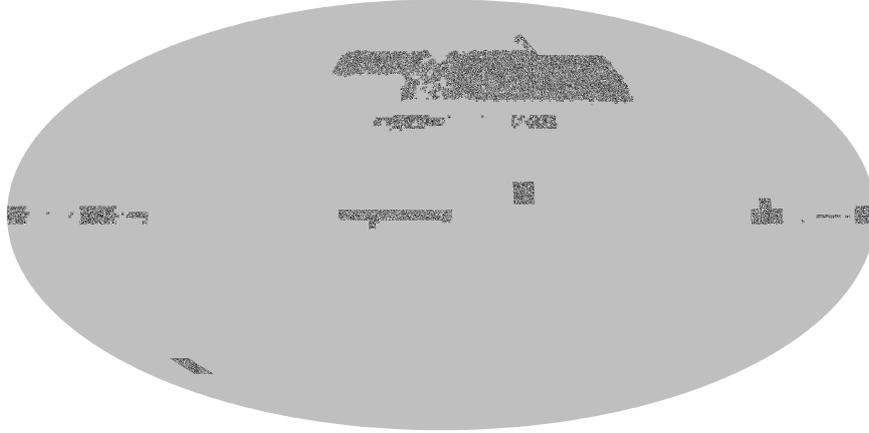
SKS smoothed

DarkMapper (ours)

$$-0.00798 \qquad 0.00000 \qquad 0.00814$$
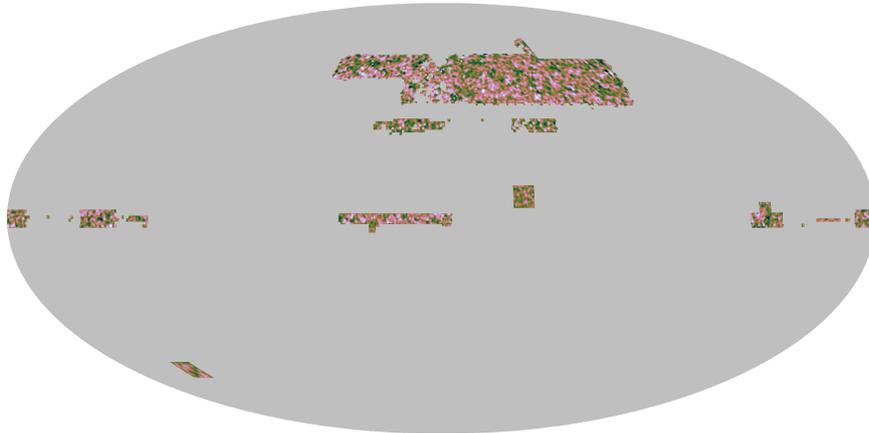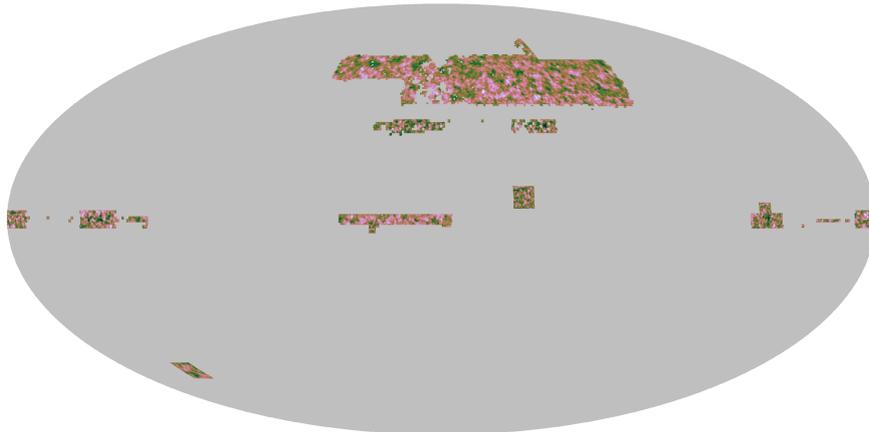
Figure 4. Mollweide projections of global reconstruction of the majority of public weak lensing datasets. **Top to bottom:** Spherical Kaiser-Squires (SKS) estimator without Gaussian smoothing kernel, SKS estimator with FWHM = Θ = 25 arcmin smoothing kernel (as in other studies), DarkMapper (our) estimator. All reconstructions are plotted on the same colorscale to aid comparison (Green 2011). The data-sets can be found online at https://doi.org/10.5281/zenodo.3980652.
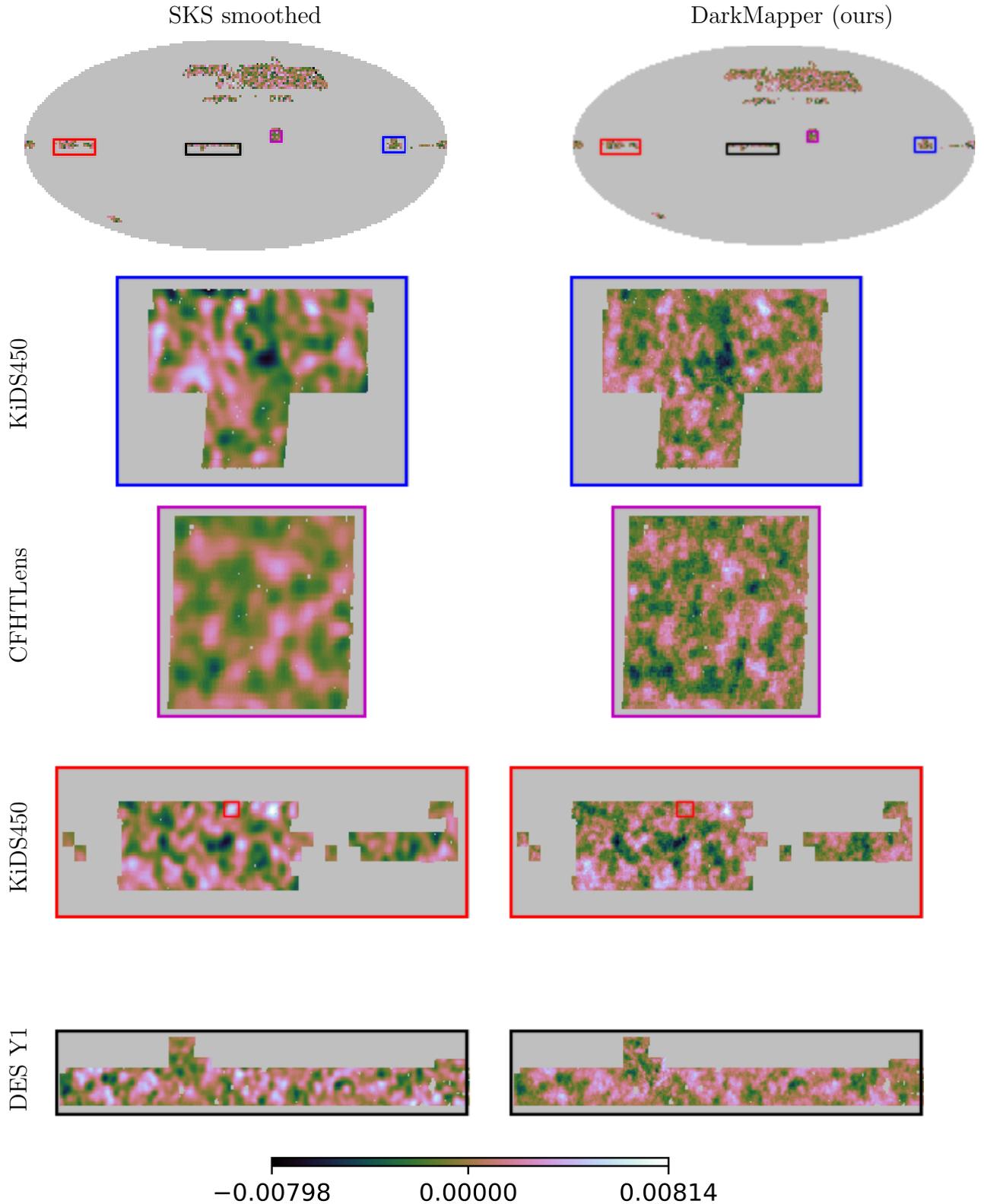
Figure 5. Top to bottom: Global reconstruction of the majority of public weak lensing datasets, magnified view of a variety of patches. **Left to right:** Spherical Kaiser-Squires (SKS) estimator without Gaussian smoothing kernel, SKS estimator with FWHM = Θ = 25 arcmin smoothing kernel (as in other studies), DarkMapper (our) estimator. **Discussion:** Note the overall agreement between both the smoothed SKS estimator and our DarkMapper estimator, however notice the significant increase in small-scale detail captured by the DarkMapper estimator. All reconstructions are plotted on the same colorscale to aid comparison (Green 2011). These data-sets can be found online at https://doi.org/10.5281/zenodo.3980652.

These structural dissimilarities between the smoothed SKS and DarkMapper estimator may reasonably be attributed to large noise fluctuations and boundary effects, both of which are not reasonably accounted for by the SKS estimator. Observation of such significant differences indicates that more principled reconstruction algorithms (such as DarkMapper) are important considerations when attempting to perform future statistical and scientific inference from dark matter mass-maps.

All reconstructions were performed on a 2016 MacBook air and took ~ 30 hours to compute. A further ~ 100 hours were optionally undertaken for annealing iterations to optimise the convergence. Note that this is by no means a benchmark of computational performance.

6.2 Local uncertainty quantification

Given significant structural dissimilarities between the SKS and DarkMapper convergence estimators we performed several hypothesis tests of local structure. Specifically we addressed the missing peaks observed in the smoothed SKS estimator of the lower (red) region of Figure 5 but not in the corresponding DarkMapper estimator. We did so by performing local hypothesis testing of structure as described in Section 4.2.1 (see Price et al. 2018, for more comprehensive details).

In all cases the hypothesis test of local structure could not reject the existance of such structure at reasonable confidence. This is unsurprising given the notably high noise level inherent to Stage III weak lensing surveys (which reduces the magnitude of the objective function thus making the approximate level set threshold ϵ'_α more difficult to reach; see Section 4) and the extremely high dimensionality $\sim \mathcal{O}(10^7)$ of the reconstruction (which directly increases the level set threshold ϵ'_α in equation 40).

6.3 Global uncertainty quantification

For high dimensional cases it is often more informative to consider global features of the reconstruction, as discussed in Section 4. A question one may wish to address is for which smoothing scales Θ does the SKS estimator provide solutions that are not in disagreement with the DarkMapper estimator at some well defined confidence.

To address this question within our global uncertainty quantification we consider the SKS estimator with a variety of Gaussian smoothing kernels, specifically $\Theta = 5i$ for integer $i \in [0, 6]$, which is to say a uniform sampling of different (in practice arbitrary) smoothing choices ranging from no smoothing (the basic SKS estimator) to the typically adopted case of $\Theta \sim 30$ arcmin smoothing. In this way we can directly address the question of which smoothing scales produces solutions $\kappa^{\mathrm{sur}}(\Theta)$ belong to the DarkMapper approximate HPD-credible region C'_α (are consistent with the DarkMapper estimator) and which solutions are unacceptable (*i.e.* those solutions which reject the null hypothesis that the surrogate is within the credible set) at $100(1 - \alpha)\%$ confidence.

The results of this global uncertainty quantification (at 99% confidence) can be found numerically in Table 2. Despite the high noise level present in the joint dataset, the uncertainty quantification technique is sensitive enough to reject the SKS estimator for $\Theta = 0, 5, 10$ arcminutes, which is to say that these smoothing scales are in disagreement with the DarkMapper estimator at 99% confidence and are unlikely to be physically meaningful. This provides statistically rigourous evidence for the community's intuition that SKS estimators require considerable smoothing to be considered meaningful.

This raises an interesting point worth noting: the SKS estimator (by construction) locates solutions within C'_α which exhibit relatively little small-scale structure, whereas the DarkMapper estimator locates solutions within C'_α which retain significantly greater small-scale structure. Therefore, though the two solutions do not disagree at $100(1 - \alpha)\%$ confidence, the DarkMapper estimator places relatively more probability on small-scale structures.

Note that if both estimators provided details of the HPD credible set then a stronger discussion of the relative cardinality of the intersection of both HPD credible sets could be used to quantify the level of statistical agreement. However, in this case the SKS estimator does not support a principled statistical interpretation and so can only justifiably be treated as a point estimate.

7 CONCLUSIONS

In this paper we have extended the previously presented (Price et al. 2018) sparse Bayesian reconstruction formalism to the spherical setting, resulting in a sparse spherical Bayesian mass-mapping algorithm which we refer to as DarkMapper. This algorithm is general and accomodates any log-concave posterior. In this paper we adopt a Laplace-type sparsity promoting wavelet prior with a multivariate Gaussian likelihood.

The DarkMapper mass-mapping algorithm was benchmarked against spherical Kaiser-Squires (Wallis et al. 2017a) in a variety of realistic weak lensing settings (ranging from Stage III to future space based surveys) using the Takahashi (Takahashi et al. 2017) N-body simulations and a pseudo-Euclid masking. In all cases we perform analysis at a typically adopted angular bandlimit of $\ell_{\max} = 2048$. We do not consider intrinsic alignments in this paper, but highlight how they may be included should one wish it.

In all simulations the DarkMapper algorithm dramatically outperforms (in both recovered SNR and recovered Pearson correlation coefficient) the SKS estimator, even when artificially selecting the optimal SKS smoothing kernel (*i.e.* even when biasing our evaluation in favour SKS as strongly as possible).

We extend approximate Bayesian uncertainty quantification methods (Pereyra 2017; Price et al. 2018; Price et al. 2019b,a; Cai et al. 2018b; Repetti et al. 2018) to the spherical setting and explain how one may leverage these methods from local uncertainty quantification to general global (or aggregate) uncertainty quantification.

The DarkMapper estimator was applied to a joint observational shear dataset constructed by collating the majority of publicly available weak lensing data – specifically the DESY1 (Abbott et al. 2018; Morganson et al. 2018; Flaugher et al. 2015), CFHTLens (Erben et al. 2012), and the KiDS450 (Hildebrandt et al. 2017; Fenech Conti et al.

Surrogate κ^{sur}		Analysis ($\epsilon'_{99\%} = 842789$)		Hypothesis test
Estimator	Θ (arcmin)	$\mathrm{Obj}(\kappa^{\mathrm{sur}})$	$\mathrm{Obj}(\kappa^{\mathrm{sur}})/\epsilon'_{99\%}$	$\kappa^{\mathrm{sur}} \in C'_{99\%}$
SKS	0	6151070	7.298	✗
	5	2584510	3.067	✗
	10	891266	1.058	✗
	15	586741	0.696	✓
	20	513223	0.609	✓
	25	488887	0.580	✓
	30	478245	0.567	✓

Table 2. Description: Uncertainty quantification of convergence estimators and smoothing scales, in each case the convergence surrogate solution κ^{sur} is defined by estimator (*i.e.* SKS) and Gaussian smoothing scale FWHM = Θ in arcmins. For each surrogate both the objective function and the ratio of the objective function to the level set threshold at 99% confidence $\epsilon'_{99\%}$ is presented. The right hand column indicates whether a given surrogate κ^{sur} belongs to the credible set (and is therefore not rejected as a possible solution to the reconstruction). Shaded in red are solutions which are rejected by Bayesian hypothesis testings. **Discussion:** Clearly, the SKS estimator without smoothing is unequivically rejected, which is concurrent with the community's intuition that smoothing is required for the SKS estimator to produce physically meaningful solutions. The minimal smoothing scale required for any SKS solution to not be rejected is $\Theta \sim 15$ arcminutes, therefore with a typical smoothing of $\Theta \in [25, 30]$ arcmins the SKS solution belongs to the DarkMapper credible set and cannot be rejected at 99% confidence (*i.e.* the two estimators are not necessarily conflicting). Nevertheless, the DarkMapper estimator contains greater fine-scale structure.

2017) surveys. To the best of our knowledge this is the first joint spherical reconstruction of all public weak lensing shear observations.

For comparison we also computed the SKS estimator of this joint dataset. We find, as with the simulated benchmarking, that the DarkMapper algorithm recovers significantly more fine-scale structure without the need for any assumptions of Gaussianity or *ad hoc* smoothing parameters (*i.e.* the smoothing scale for SKS post-processing). This demonstrates that the algorithm works as expected on observational data.

Finally, uncertainty quantification was carried out to determine for which smoothing scales the SKS point estimates provide solutions that are acceptable solutions to the DarkMapper Bayesian inference problem (*i.e.* within the highest posterior density credible region) – this is to say the smoothing scales at which both convergence estimates are not conflicting at 99% confidence. It was found that all SKS reconstructions with smoothing scales below ~ 15 arcmin were rejected at 99% confidence, indicating that significant smoothing is required for agreement between the SKS and DarkMapper estimators. This reaffirms the community's understanding that SKS estimators must undergo significant smoothing to recover physically meaningful convergence maps.

Moreover, we demonstrate that the DarkMapper estimator locates permissible solutions with significantly greater small-scale structure than those which are located by the SKS estimator. More constraining statistical statements were limited by the inherently high noise level in current observational shear data.

With the advent of Stage IV surveys the pixel noise level is projected to drop dramatically (due to increased galaxy number density), which will inevitably facilitate significantly more constraining statistical statements. As the DarkMapper estimator not only provides dramatically increased reconstruction fidelity over the SKS estimator but also supports a principled Bayesian interpretation, it will be of important use for application to Stage IV datasets.

Note that just as we have extended this sparse hierarchical Bayesian mass-mapping formalism to the sphere ($\mathbb{S}^2$)

one can extend it to the ball ($\mathbb{B}^3$) and thus recover similar results for the case of full 3D mass-mapping. This is a possible avenue for future investigation.

---

[9] https://github.com/astro-informatics/sopt
[10] https://astro-informatics.github.io/ssht/
[11] http://astro-informatics.github.io/s2let/
[12] http://astro-informatics.github.io/so3/

7.2 Kilo Degree Survey Acknowledgements

8 DATA AVAILABILITY

All observational data utilized throughout this paper is publicly available and can be found in the corresponding references. All joint reconstruction data-sets are publicly available and can be found online at `https://doi.org/10.5281/zenodo.3980652`. The DarkMapper reconstruction software will be publicly released in the near future.

REFERENCES

Abbott T. M. C., et al., 2018, ApJS, 239, 18
Alsing J., Kirk D., Heavens A., Jaffe A. H., 2015, MNRAS, 452, 1202
Alsing J., Heavens A., Jaffe A. H., Kiessling A., Wandelt B., Hoffmann T., 2016, MNRAS, 455, 4452
Baldi P., Kerkyacharian G., Marinucci D., Picard D., et al., 2009, The Annals of Statistics, 37, 1150
Barreiro R. B., Hobson M. P., Lasenby A. N., Banday A. J., Górski K. M., Hinshaw G., 2000, MNRAS, 318, 475
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Bunn E. F., Zaldarriaga M., Tegmark M., de Oliveira-Costa A., 2003, Phys. Rev. D, 67, 023501
Cai X., Pereyra M., McEwen J. D., 2018a, MNRAS, 480, 4154
Cai X., Pereyra M., McEwen J. D., 2018b, MNRAS, 480, 4170
Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223
Carrillo R. E., McEwen J. D., Van De Ville D., Thiran J.-P., Wiaux Y., 2013, IEEE Signal Processing Letters, 20, 591
Castro P. G., Heavens A. F., Kitching T. D., 2005, Phys. Rev. D, 72, 023516
Chan J. Y. H., Leistedt B., Kitching T. D., McEwen J. D., 2017a, IEEE Transactions on Signal Processing, 65, 5
Chan J. Y. H., Leistedt B., Kitching T. D., McEwen J. D., 2017b, IEEE Transactions on Signal Processing, 65, 5
Chang C., et al., 2018, MNRAS, 475, 3165
Clowe D., Brada M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJ, 648, L109
Coles P., Chiang L.-Y., 2000, Nature, 406, 376
Combettes P. L., Pesquet J.-C., 2009, preprint, (arXiv:0912.3522)
Corless V. L., King L. J., Clowe D., 2009, MNRAS, 393, 1235
Erben T., et al., 2012, Monthly Notices of the Royal Astronomical Society, 433
Erben T., et al., 2013, MNRAS, 433, 2545
Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, MNRAS, 467, 1627
Flaugher B., et al., 2015, AJ, 150, 150
Fluri J., Kacprzak T., Sgier R., Refregier A., Amara A., 2018, J. Cosmology Astropart. Phys., 2018, 051
Fluri J., Kacprzak T., Lucchi A., Refregier A., Amara A., Hofmann T., Schneider A., 2019, Phys. Rev. D, 100, 063514
Giblin B., et al., 2018, Monthly Notices of the Royal Astronomical Society, 480, 5529
Goldberg J. N., Macfarlane A. J., Newman E. T., Rohrlich F., Sudarshan E. C. G., 1967, Journal of Mathematical Physics, 8, 2155
Green D. A., 2011, Bulletin of the Astronomical Society of India, 39, 289
Grimm N., Yoo J., 2018, J. Cosmology Astropart. Phys., 2018, 067
Hildebrandt H., et al., 2017, MNRAS, 465, 1454
Hinshaw G., et al., 2013, ApJS, 208, 19
Horowitz B., Seljak U., Aslanyan G., 2018, arXiv preprint arXiv:1810.00503
Jee M. J., Dawson W. A., Stroe A., Wittman D., van Weeren R. J., Brüggen M., Bradač M., Röttgering H., 2016, ApJ, 817, 179
Jeffrey N., et al., 2018, MNRAS, 479, 2871
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kuijken K., et al., 2015, MNRAS, 454, 3500
Lanusse F., Starck J.-L., Leonard A., Pires S., 2016, A&A, 591, A2
Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
Leistedt B., McEwen J. D., Vandergheynst P., Wiaux Y., 2013a, A&A, 558, A128
Leistedt B., McEwen J. D., Vandergheynst P., Wiaux Y., 2013b,

A&A, 558, A128

Mallat S., 1999, A Wavelet Tour of Signal Processing

Marinucci D., et al., 2007, Monthly Notices of the Royal Astronomical Society, 383, 539

McEwen J. D., Price M. A., 2015, arXiv e-prints, p. arXiv:1510.01595

McEwen J. D., Price M. A., 2019, in 27th European Signal Processing Conference (EUSIPCO). (arXiv:1510.01595)

McEwen J. D., Scaife A. M. M., 2008, MNRAS, 389, 1163

McEwen J. D., Wiaux Y., 2011, IEEE Transactions on Signal Processing, 59, 5876

McEwen J. D., Wiaux Y., Eyers D. M., 2011, A&A, 531, A98

McEwen J. D., Puy G., Thiran J.-P., Vandergheynst P., Van De Ville D., Wiaux Y., 2013a, IEEE Transactions on Image Processing, 22, 2275

McEwen J. D., Vandergheynst P., Wiaux Y., 2013b, in Proc. SPIE. p. 88580I (arXiv:1308.5706), doi:10.1117/12.2022889

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

McEwen J. D., Buttner M., Leistedt B., Peiris H. V., Wiaux Y., 2015b, IEEE Signal Processing Letters, 22, 2425

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

McFarland J. P., Verdoes-Kleijn G., Sikkema G., Helmich E. M., Boxhoorn D. R., Valentijn E. A., 2013, Experimental Astronomy, 35, 45

Mediavilla E., Muñoz J. A., Garzón F., Mahoney T. J., 2016, Astrophysical Applications of Gravitational Lensing

Miller L., et al., 2013, MNRAS, 429, 2858

Moreau J. J., 1962, C.R. Acad. Sci. Paris Ser. A Math., 255, 2897

Morganson E., et al., 2018, PASP, 130, 074501

Munshi D., Coles P., 2017, J. Cosmology Astropart. Phys., 2, 010

Narcowich F. J., Petrushev P., Ward J. D., 2006, SIAM Journal on Mathematical Analysis, 38, 574

Newman E. T., Penrose R., 1966, Journal of Mathematical Physics, 7, 863

Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016, MNRAS, 462, 4314

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

Pereyra M., 2017, SIAM Journal on Imaging Sciences, 10, 285

Pereyra M., Bioucas-Dias J., Figueiredo M., 2015, Maximum-a-posteriori estimation with unknown regularisation parameters. pp 230–234, doi:10.1109/EUSIPCO.2015.7362379

Pichon C., Thiébaut E., Prunet S., Benabed K., Colombi S., Sousbie T., Teyssier R., 2010, MNRAS, 401, 705

Planck Collaboration et al., 2018, arXiv preprint arXiv:1807.06209

Pratley L., McEwen J. D., d'Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2018, MNRAS, 473, 1038

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

Price M. A., McEwen J. D., Cai X., Kitching T. D., 2019a, Monthly Notices of the Royal Astronomical Society, 489, 3236

Price M. A., Cai X., McEwen J. D., Pereyra M., Kitching T. D., 2019b, Monthly Notices of the Royal Astronomical Society, 492, 394

Price M. A., Pratley L., McEwen J. D., 2020, in prep

Repetti A., Pereyra M., Wiaux Y., 2018, preprint, (arXiv:1803.00889)

Robert C.-P., 2001, The Bayesian Choice, doi:https://doi.org/10.1007/0-387-71599-1.

Schneider P., 2005, ArXiv Astrophysics e-prints,

Schneider M. D., Hogg D. W., Marshall P. J., Dawson W. A., Meyers J., Bard D. J., Lang D., 2015, ApJ, 807, 87

Schröder P., Sweldens W., 1995, in Computer Graphics Proceedings (SIGGRAPH '95). pp 161–172

Starck J.-L., Moudden Y., Abrial P., Nguyen M., 2006, Astronomy & Astrophysics, 446, 1191

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

Taylor P. L., Kitching T. D., McEwen J. D., 2018, Phys. Rev. D, 98, 043532

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

Vallis Z. M., Wallis C. G. R., Kitching T. D., 2018, Astronomy and Computing, 24, 84

Van Waerbeke L., et al., 2013, MNRAS, 433, 3373

VanderPlas J. T., Connolly A. J., Jain B., Jarvis M., 2011, ApJ, 727, 118

Wallis C. G. R., McEwen J. D., Kitching T. D., Leistedt B., Plouviez A., 2017a, preprint, (arXiv:1703.09233)

Wallis C. G. R., Wiaux Y., McEwen J. D., 2017b, IEEE Transactions on Image Processing, 26, 5176

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

de Jong J. T. A., et al., 2015, A&A, 582, A62

APPENDIX A: WAVELETS ON THE SPHERE

This section will provide extremely brief (and quite technical) overview of scale-discretised spherical wavelets; for extensive details see the related articles (Wiaux et al. 2008; Wallis et al. 2017b; McEwen et al. 2015b, 2013a; Leistedt et al. 2013b; McEwen et al. 2015a; McEwen & Price 2015; Chan et al. 2017a; McEwen et al. 2018).

The scale-discretised wavelet transform on the sphere is given by the directional convolution of each wavelet Ψ^j of scale $j \in [J_0, J_{\max}]$ with a field $_0\kappa$, such that the wavelet coefficients $w^j(\rho) \in \ell^2_{\max}(SO(3))$ are given by,

$$w^j(\rho) = \int_{\mathbb{S}^2} d\rho(\rho') \, _0\kappa(\rho')(\mathcal{R}_\rho \Psi^j)^*(\rho'). \tag{A1}$$

Additionally, the low frequency component of $_0\kappa$ is encapsulated by the axisymmetric convolution with $\Upsilon \in \ell^2_{\max}(\mathbb{S}^2)$ such that,

$$s(\omega) = \int_{\mathbb{S}^2} d\Omega(\omega') \, _0\kappa(\omega')(\mathcal{R}_\omega \Upsilon)^*(\omega'). \tag{A2}$$

It is possible to project these directional and axisymmetric convolution operators into harmonic space (Wallis et al. 2017b),

$$(w^j)^\ell_{mn} \equiv \langle w^j, \mathcal{D}^{\ell*}_{mn} \rangle = \frac{8\pi^2}{2\ell+1} \, _0\kappa_{lm} \, \Psi^{j*}_{\ell n}, \tag{A3}$$

$$s_{\ell m} \equiv \langle s, Y_{\ell m} \rangle = \sqrt{\frac{4\pi}{2\ell+1}} \, _0\kappa_{lm} \, \Upsilon^*_{\ell 0}. \tag{A4}$$

The coefficients of a general spherical signal $_s f$ can be synthesised exactly by

$$_s f(\omega) = \int_{\mathbb{S}^2} d\Omega s(\omega')(\mathcal{R}_{\omega'} s)(\omega') + \sum_{j=J_0}^{J} \int_{SO(3)} d\rho w^j(\rho)(\mathcal{R}_{\rho'} w^j)(\rho'), \tag{A5}$$

from which we can define the discretized form of $_0\kappa$ by

$$
_0\kappa(\omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left\{ \sqrt{\frac{4\pi}{2\ell+1}} s_{\ell m} \Upsilon_{\ell 0} Y_{\ell m}(\omega) \right.
$$
$$
\left. + \frac{2\ell+1}{8\pi^2} \sum_{j=J_0}^{J_{\max}} \sum_{n=-\ell}^{\ell} (w^j)^{\ell}_{mn} \Psi^j_{\ell n} Y_{\ell m}(\omega) \right\},
$$
$$(A6)$$

where the harmonic coefficients $_0\kappa_{\ell m}$ are explicitly given by

$$
_0\kappa_{\ell m} = \sqrt{\frac{2\ell+1}{4\pi}} s_{\ell m} \Upsilon_{\ell 0} + \frac{2\ell+1}{8\pi^2} \sum_{j=J_0}^{J_{\max}} \sum_{n=-\ell}^{\ell} (w^j)^{\ell}_{mn} \Psi^j_{\ell n}
$$
$$(A7)$$

Combining equations (A3) and (A4) the harmonic representation of the spherical wavelet transform at scale j can be written as a combination of the following linear operators (Wallis et al. 2017b):

$$
\hat{w}^j = \mathbf{N}^j \mathbf{W}^j \, _0\hat{\kappa}_{\ell m},
$$

are the Wigner coefficients of the directional wavelet coefficients. Here $\mathbf{W}^j$ is harmonic space multiplication by the wavelet kernel $\Psi^{j*}_{\ell n}$, and $\mathbf{N}^j$ corresponds to wavelet normalization.

$$
\hat{s} = \mathbf{S} \, _0\hat{\kappa}_{\ell m},
$$

are the spherical harmonic coefficients of the scaling wavelet coefficients, where $\mathbf{S}$ is harmonic space multiplication by scaling kernel $\Upsilon^*_{\ell 0}$.

We further specialize by grouping the harmonic representation of the coefficients into $\hat{\alpha} = \mathbf{NW} \, _0\hat{\kappa}_{\ell m}$, where the wavelet scale normalization terms are collectively $\mathbf{N} = \mathrm{diag}(\mathbb{I}, \mathbf{N}^{J_0}, \ldots, \mathbf{N}^{J_{\max}})$, and the Wigner/harmonic space wavelet convolutions are collectively $\mathbf{W} = \mathrm{diag}(\mathbf{S}, \mathbf{W}^{J_0}, \ldots, \mathbf{W}^{J_{\max}})$. Finally we introduce a diagonal array of operators $\mathbf{H}$ corresponding to inverse spherical harmonic and Wigner transforms, $\mathbf{H} = \mathrm{diag}(\mathbf{Y}, \mathcal{D}^{J_0}, \ldots, \mathcal{D}^{J_{\max}})$ In the scope of this paper we restrict ourselves to the analysis setting, thus these are the only operators required though a complete set are provided in Wallis et al. (2017b). Concatenating these operators we can finally form analysis forward and adjoint wavelet transforms,

$$
\tilde{\Psi} = \mathbf{HNW\tilde{Y}} \qquad \text{and} \qquad \tilde{\Psi}^{\dagger} = \mathbf{\tilde{Y}^{\dagger} W^{\dagger} N H^{\dagger}}. \qquad (A8)
$$

Note that † is the adjoint (or conjugate transpose) operator, and as $\mathbf{N}$ is simply wavelet scaling it is trivially self-adjoint $\mathbf{N} = \mathbf{N}^{\dagger}$. Crucially as we work in the discrete setting the inverse wavelet transforms are not equivalent to the adjoint wavelet transforms[13], and so we adopt the optimized algorithms to compute fast Wigner and Wigner adjoint functions presented in the appendix of McEwen et al. (2015b).

This paper has been typeset from a TeX/LaTeX file prepared by the author.

---

[13] An approximation often made throughout the literature.