

Sparse image reconstruction on the sphere: a general approach with uncertainty quantification

Matthew A. Price, Luke Pratley, Jason D. McEwen

Abstract—Inverse problems defined naturally on the sphere are becoming increasingly of interest. In this article we provide a general framework for evaluation of inverse problems on the sphere, with a strong emphasis on flexibility and scalability. We consider flexibility with respect to the prior selection (regularization), the problem definition — specifically the problem formulation (constrained/unconstrained) and problem setting (analysis/synthesis) — and optimization adopted to solve the problem. We discuss and quantify the trade-offs between problem formulation and setting. Crucially, we consider the Bayesian interpretation of the unconstrained problem which, combined with recent developments in probability density theory, permits rapid, statistically principled uncertainty quantification (UQ) in the spherical setting. Linearity is exploited to significantly increase the computational efficiency of such UQ techniques, which in some cases are shown to permit analytic solutions. We showcase this reconstruction framework and UQ techniques on a variety of spherical inverse problems. The code discussed throughout is provided under a GNU general public license, in both C++ and Python.

Index Terms—harmonic analysis, sampling, spheres, spherical wavelets, uncertainty quantification

I. INTRODUCTION

INCREASINGLY often one wishes to solve inverse problems natively on the sphere (\mathbb{S}^2) rather than on n dimensional Euclidean space (\mathbb{R}^n), *e.g.* in astronomy and astrophysics [1]–[3], biomedical imaging [4], [5], and geophysics [6]. Straightforwardly from Gauss’ *Theorema Egregium* — which states that the curvature of surfaces embedded in \mathbb{R}^3 is immutable, and thus planar projections of curved manifolds (*e.g.* the sphere) inherently incur (significant) distortions — analysis over such domains must necessarily be conducted natively on the sphere. Though many Euclidean techniques may provide inspiration for counter-parts on the sphere, there are still a great many critical differences between these paradigms which must be considered. Typically, inverse problems of interest, particularly on the sphere, are (often severely) ill-posed and/or ill-conditioned, motivating the injection of prior knowledge to stabilize the reconstruction. Such problems can be solved in a variety of ways (*e.g.* sampling methods and machine learning methods) though, for robustness and scalability, in the spherical setting variational methods (*e.g.* optimization) are the most effective.

Due to recent advances in the theory of compressed sensing [7]–[9] sparsity priors (*e.g.* ℓ_1 -regularization) are now routinely adopted, where the solution to an inverse problem can be constrained and found by promoting sparsity in a dictionary, such as wavelets or gradient space (variational norms). Recent developments in proximal convex optimization algorithms

facilitate the practical application of non-differentiable priors, where they can be distributed and scale to high dimensional parameter spaces [10], [11]. The spherical counterparts for discrete gradient spaces [12], wavelet families [13]–[15], and scale-discretized wavelet families [16]–[20] have been developed, and have found wide applications — see previous papers in this series [12], [21] for a more comprehensive overview on this topic. Somewhat restricted investigations of some aspects have already been conducted, *e.g.* considering sparsity in spherical harmonic space [22], sparsity in various redundant dictionaries [12], [21], [23].

Variational inference techniques to solve inverse problems may be constructed in either the analysis or synthesis setting where signal coefficients or coefficients of a sparse representation are recovered respectively [24]. For Euclidean settings the analysis problem typically provides greater reconstruction fidelity; a characteristic often attributed to the lower cardinality of the analysis solution space [24]–[26], however comparisons between the analysis and synthesis settings on the sphere are not so clear, due to the approximate effective cardinality of different spaces on the sphere [21]. There also exists a more fundamental binary-classification of optimization problems: constrained and unconstrained, corresponding to regularization *via* hard and soft constraints respectively [27]. Hard constraints (constrained formulation) do not depend on variables such as Lagrangian multipliers, the optimal selection of which is an open problem, and instead constrain the solution to a certain sub-space. Soft constraints (unconstrained formulation) can be considered as Bayesian inference problems [28] and thus support a principled statistical interpretation [29]–[31].

Traditionally, although variational approaches may support a probabilistic interpretation they typically recover point estimates and do not quantify uncertainties. Fully probabilistic approaches (*e.g.* Markov chain Monte Carlo sampling methods) exist but are computationally expensive in the high dimensional setting of the sphere, motivating the development of hybrid techniques. Recent developments in the field of probability density theory [30] address precisely this consideration, facilitating flexible generation of scalable, fully principled Bayesian uncertainty quantification (UQ) techniques for variational approaches. Many such techniques have been developed [31]–[35], with applications in a variety of domains. In this article, we leverage these UQ techniques to recover Bayesian local credible intervals, in effect pixel-level error bars, and other forms of hypothesis tests on discrete spherical spaces. Interestingly we show how these uncertainties for a variety of common objectives can be computed rapidly (by exploiting linearity) and in some cases analytically. Such computational

savings are a key component for the future of scalable UQ for spherical inverse problems. Looking forward one might note that these UQ techniques for variational imaging rely only on log-concavity of the posterior (convexity of the objective), as such a great many combinations of likelihood (data-fidelity) and prior (regularization functionals) are permissible.

In the spirit of open access software and scientific reproducibility the spherical reconstruction software (S2INV) developed during this project is made publicly available.¹ S2INV is an object oriented C++ software package (with python extensions) which acts as a spherical extension to the SOPT [11], [36], [37] software package for flexible, efficient sparse optimization. We use fast exact spherical harmonic [38] and spherical wavelet transforms [17] to rapidly solve linear and ill-conditioned spherical inverse problems.

The remainder of this article is structured as follows. In Section II we provide the mathematical context which underpins analysis of spin signals on the sphere. In Section III we present variational regularization approaches to solve spherical inverse problems and consider the unconstrained and constrained formulations, in both the analysis and synthesis settings. Furthermore, we discuss the generalization of planar regularization functionals to their spherical counterparts, and briefly highlight highly optimized, scalable spherical reconstruction open-source software available as a bi-product of this work. In Section IV we develop principled Bayesian uncertainty quantification techniques which can be leveraged for spherical inverse problems, and present acceleration methods exploiting function linearity and/or objective analytic solutions. A diverse selection of numerical experiments are presented in Section V before providing concluding remarks in Section VI.

II. SPIN-SIGNALS ON THE SPHERE AND ROTATION GROUP

One often wishes to consider the frequency space representation of signals; whether this be embedded within regularization methods, necessary to fully capture a desired forward model, or simply adopted to exploit computational symmetries (*e.g.* fast convolution algorithms). In the Euclidean setting, the frequency information of a signal is efficiently expressed through projection onto Fourier space, the *Fourier transform*. For spherical settings frequency information is expressed through projection onto the space of *spin spherical harmonics*. In this section we review mathematical background fundamental to the analysis of signals defined on the sphere.

A. Spin spherical harmonic transforms

The space of square integrable spin- s functions ${}_s f \in L^2[\mathbb{S}^2]$, for $s \in \mathbb{Z}$, with inner product $\langle \cdot | \cdot \rangle_{\mathbb{S}^2}$, are defined by their response under local rotations of $\chi \in [0, 2\pi)$ about the tangent plane centered on the spherical co-ordinate $\omega = (\theta, \psi) \in \mathbb{S}^2$, given by ${}_s f'(\omega) = e^{-is\chi} {}_s f(\omega)$ where ${}_s f'$ is the rotated function [39], [40]. Such functions are most naturally represented by the spin-weighted spherical harmonics ${}_s Y_{\ell m} \in L^2[\mathbb{S}^2]$ which are a set of complete and orthogonal basis functions for degree $\ell \in \mathbb{Z}^+$ and integer $m \in \mathbb{Z}$, $|m| \leq \ell$, $|s| \leq \ell$. We adopt

the Condon-Shortley phase convention [41], which results in conjugate symmetry ${}_s Y_{\ell m}^*(\omega) = (-1)^{s+m} {}_{-s} Y_{\ell -m}(\omega)$, where $(\cdot)^*$ denotes complex conjugation.

A spin- s function ${}_s f \in L^2[\mathbb{S}^2]$ may be decomposed into the spin spherical harmonic basis by

$${}_s f_{\ell m} \equiv \langle {}_s f_{\ell m} | {}_s Y_{\ell m} \rangle_{\mathbb{S}^2} = \int_{\mathbb{S}^2} d\Omega(\omega) {}_s f(\omega) {}_s Y_{\ell m}^*(\omega), \quad (1)$$

where $d\Omega(\omega) = \sin(\theta)d\theta d\psi$ is the standard rotation invariant measure (Haar measure) on the sphere. Equivalently, by the orthogonality and completeness of ${}_s Y_{\ell m}(\omega)$, one can exactly synthesize the signal space representation by

$${}_s f(\omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} {}_s f_{\ell m} {}_s Y_{\ell m}(\omega), \quad (2)$$

where the sum over ℓ is often truncated at L , where it is assumed that ${}_s f_{\ell m} = 0, \forall \ell \geq L$. In this sense, signals are considered to be bandlimited at L . For notation brevity we adopt the shorthand operator notation \mathbf{Y} and \mathbf{Y}^{-1} to denote the forward and inverse spherical harmonic transforms.

This transformation allows one to probe the frequency content of spin signals defined on the sphere, which facilitates, *e.g.* efficient convolutions over spherical manifolds, in much the same way one can compute convolutions over \mathbb{R}^2 through the Fourier convolution theorem. In many cases signals have 0 spin, and so these relations collapse to the simpler form most readers are likely familiar with. Nevertheless, a variety of interesting physical settings exist where signals exhibit non-zero spin, *e.g.* weak gravitational lensing, the cosmic microwave background, or quantum mechanical systems.

B. Scale-discretized directional spherical wavelets

Leveraging the above spin- s spherical harmonic basis, and the associated convolutional properties, one can construct wavelet dictionaries naturally on the sphere. To do so one must first define a general rotation \mathcal{R}_ρ , for Euler angles $\rho = (\alpha, \beta, \gamma) \in \text{SO}(3)$ with $\alpha \in [0, 2\pi)$, $\beta \in [0, \pi)$, and $\gamma \in [0, 2\pi)$, with action $(\mathcal{R}_\rho {}_s f)(\omega) \equiv e^{-is\theta} {}_s f(\mathcal{R}_\rho^{-1}\omega)$. The directional scale-discretized wavelet coefficients of any square integrable spin- s function ${}_s f \in L^2[\mathbb{S}^2]$ are given for scale j by the directional convolution

$$\begin{aligned} W^s \Psi^{(j)}(\rho) &\equiv ({}_s f \otimes {}_s \Psi^{(j)})(\rho) \equiv \langle {}_s f, \mathcal{R}_\rho {}_s \Psi^{(j)} \rangle_{\mathbb{S}^2} \\ &= \int_{\mathbb{S}^2} d\Omega(\omega) {}_s f(\omega) (\mathcal{R}_\rho {}_s \Psi^{(j)})^*(\omega), \end{aligned} \quad (3)$$

where \otimes represents the directional spherical convolution and ${}_s \Psi^{(j)} \in L^2[\mathbb{S}^2]$ is the wavelet kernel at scale $j \in \mathbb{Z}^+$, which determines the compact support of a given wavelet scale [17].

Typically wavelet coefficients have negligible energy concentration over the low-frequency domain in harmonic space, hence a scaling function ${}_s \Upsilon \in L^2[\mathbb{S}^2]$ is introduced [4], [19] with coefficients $W^s \Upsilon \in L^2[\mathbb{S}^2]$ defined by the axisymmetric convolution \odot with a signal ${}_s f \in L^2[\mathbb{S}^2]$ such that

$$\begin{aligned} W^s \Upsilon(\omega) &\equiv ({}_s f \odot {}_s \Upsilon)(\omega) \equiv \langle {}_s f, \mathcal{R}_\omega {}_s \Upsilon \rangle_{\mathbb{S}^2} \\ &= \int_{\mathbb{S}^2} d\Omega(\omega') {}_s f(\omega') (\mathcal{R}_\omega {}_s \Upsilon)^*(\omega'), \end{aligned} \quad (4)$$

¹<https://github.com/astro-informatics/s2inv>

where $\mathcal{R}_\omega = \mathcal{R}_{(\psi, \theta, 0)}$ is a simplification of \mathcal{R}_ρ . One can straightforwardly show that the pixel-space representation of signals may be exactly synthesized if, and only if, the wavelet admissibility condition holds (see *e.g.* [19]). There exist many functions which are admissible, *e.g.* spherical needlets [42], ridgelets [4] and curvelets [43], however in this work we choose to adopt the directional scale-discretized wavelet harmonic space kernel [16], [20]. For notational brevity we define operators Ψ and Ψ^{-1} for the synthesis and analysis wavelet transforms respectively, with corresponding adjoints operators Ψ^\dagger and $(\Psi^{-1})^\dagger$ (for further details see, *e.g.*, [21]).

III. GENERALIZED SPHERICAL IMAGING

Imaging inverse problems are found in countless areas of both science and industry; consequently a great wealth of effort has been spent developing signal processing, Bayesian inference and, more recently, machine learning techniques for solving such problems. However, these techniques have overwhelmingly been restricted to Euclidean settings, in large part due to their prevalence and relative simplicity.

As such, planar imaging benefits greatly from the flexibility such a dictionary of techniques affords, whereas techniques developed for non-Euclidean manifolds (*e.g.* the sphere) are comparatively rare. One might reasonably consider applying planar techniques to spherical settings, *e.g.* through the analysis of planar projections, however these fundamentally fall short [2] as a result of Gauss' *Theorema Egregium* — a core concept of differential geometry, which dictates that one may not flatten a ball without incurring significant distortions. Nonetheless, one can certainly consider the development of analogous techniques defined natively on the sphere. Previously, the spherical total variation TV-norm was constructed [12], and the analysis and synthesis settings were compared in a spherical setting [21]. In this section we extend the discussion to include the constrained and unconstrained formulations, supported by a variety of proximal optimization algorithms, and a variety of regularization functionals.

On the sphere the setup of such imaging problems is as follows: consider the case in which one acquires complex measurements $\mathbf{y} \in \mathbb{C}^M$, which may or may not be natively on the sphere, but can be related to an estimated or true spherical signal $\mathbf{x} \in \mathbb{C}^{N_{\mathbb{S}^2}}$ through the linear mapping

$$\Phi \in \mathbb{C}^{M \times N_{\mathbb{S}^2}} : \mathbf{x} \in \mathbb{C}^{N_{\mathbb{S}^2}} \mapsto \mathbf{y} \in \mathbb{C}^M, \quad (5)$$

commonly referred to as the measurement operator, which simulates measurement acquisition. Suppose observations are contaminated with stochastic noise $\mathbf{n} \in \mathbb{C}^M$ such that $\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}$ which is classically ill-posed. Furthermore, when one considers that spherical observations are often incomplete, *i.e.* $|M| \ll |N_{\mathbb{S}^2}|$, such problem instances quickly become (seriously) ill-conditioned. A diverse set of techniques exist to solve such inverse problems. This article is primarily concerned with variational approaches, for which we develop uncertainty quantification techniques in Section IV. Variational approaches consider the inverse problem as a minimization problem over a chosen objective function, which is typically the combination of a data fidelity term and a regularization

term — selected to stabilize reconstruction with *a priori* assumptions as to the nature of the problem instance. Given an objective function over which to minimize, one must make a variety of decisions regarding optimization formulation.

A. Constrained and unconstrained optimization

Suppose one selects data-fidelity term $f(\mathbf{x})$ and regularization functional $g(\mathbf{x})$, then the unconstrained optimization problem has the Lagrangian formulation [27]

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \Omega} \{f(\mathbf{x}) + \lambda g(\mathbf{x})\}, \quad (6)$$

where $\Omega = \{\mathbb{C}^{N_{\mathbb{S}^2}}, \mathbb{R}^{N_{\mathbb{S}^2}}, \mathbb{R}_+^{N_{\mathbb{S}^2}}\}$, and the regularization parameter $\lambda \in \mathbb{R}^+$ is a Lagrangian multiplier that balances the relative contributions of the two functions to the objective. In effect λ allows for a smooth re-weighting (soft constraint) of the solution space instead of the strict boundary (hard constraint) imposed in the constrained problem. When one formulates such optimizations in the unconstrained setting, the solution which minimizes the objective is in fact the *maximum a posteriori* (MAP) estimator $\mathbf{x}^* = \mathbf{x}^{\text{MAP}}$.

Interestingly, it is well known that the unconstrained problem has direct links to Bayesian inference and supports a principled statistical interpretation. However, until recently such Bayesian interpretations have been restricted to point estimators and/or severely restricted objective functional forms. One can leverage recent advances in probability concentration theory [30] to develop unconstrained optimization techniques which support principled uncertainty quantification, as discussed in Section IV. Therefore, when considering spherical imaging problems, where Bayesian sampling methods are impractical, in scientific domains, where uncertainty quantification is a desirable feature, unconstrained optimization exhibits significant advantages. However, this advantage comes at the additional complexity of optimal selection of the regularization parameter λ . Popular methods for selection of λ have adopted criteria such as: the Akaike information criterion (AIC) [44], Bayesian information criterion (BIC) [45], or Stein's unbiased risk estimator (SURE) [46], [47] and others [48]. Optimal regularization parameter selection is still very much an open problem (for various reasons including bias vs variance considerations). In this work we adopt a recently developed hierarchical Bayesian inference approach [49] which treats the regularization parameter as a nuisance variable [28] over which a majorization-minimization algorithm marginalizes. Effectively this method produces automatic, somewhat robust λ selection with a straightforward, natural Bayesian interpretation, facilitating principled uncertainty quantification.

Suppose instead that one is unwilling to accept a trade-off in either the data-fidelity or regularization functional, *i.e.* one requires that the data-fidelity is strictly below a given threshold, or that solutions belong to a restrictive sub-space of the regularization support or measurement operator. For such inverse problems, the problem instance is formulated as a constrained optimization problem, in which one function is minimized subject to the constraint that the other function belongs to some constrained set [27]. Here we consider the common form in which the regularization functional is

minimized subject to the constraint that the solution belongs within a level-set of the data-fidelity term, *i.e.*

$$\mathbf{x}^* = \underset{\mathbf{x} \in \Omega}{\operatorname{argmin}} \{g(\mathbf{x})\} \quad \text{s.t.} \quad f(\mathbf{x}) \leq \delta, \quad (7)$$

where δ is a specified threshold (defining an iso-contour or level-set) of the data-fidelity term, typically determined by the noise variance. This optimization restricts solutions to the subspace $\mathbf{x} \in \mathcal{B}_f^\delta$ where \mathcal{B}_f^δ is the f -ball centered at $\mathbf{z} \in \Omega$ with radius δ , *i.e.* $\mathcal{B}_f^\delta(\mathbf{z}) := \{\mathbf{x} : f(\mathbf{x}) \leq \delta\}$.

This formulation of the constrained problem requires calibration of δ which can be computed from the estimated noise variance, and has a well defined interpretation. The calibration of additional Lagrangian multipliers (regularization parameters) is not required, hence the constrained setting is typically more straightforward to adopt. For many problem instances the constrained setting provides greater reconstruction fidelity, though this is likely to be problem dependent. In this sense the soft constraint adopted by the unconstrained setting (when selected appropriately) allows for bias to be traded for variance (and *vice versa*) and thus in particularly ill-posed problem instances, where the prior weighting is large (*i.e.* high bias situations), may produce estimates that are more accurate. Furthermore, the constrained problem does not have an associated or well defined posterior distribution over the latent space, prohibiting principled uncertainty quantification.

B. Analysis and synthesis settings

Often one adopts regularization functions which are computed on projections of the image space, *e.g.* wavelet space, harmonic space, gradient space *etc.*. In such settings one can formulate optimization problems that consider the inverse problem latent space to be the image space or the projected space, giving rise to the analysis and synthesis formulations respectively. In this way one recovers solutions in pixel-space \mathbf{x}^* (analysis) or projected space $\boldsymbol{\alpha}^*$, which are then inverted to form pixel-space estimates $\mathbf{x}^* = \boldsymbol{\Psi}\boldsymbol{\alpha}^*$ (synthesis) [24]–[26].

This is most easily illustrated by considering a simple example. Consider the wavelet Lasso regression problem in the analysis form, *i.e.* an ℓ_1 wavelet regularization functional $g_A(\mathbf{x}) = \|\boldsymbol{\Psi}^{-1}\mathbf{x}\|_1$ and an ℓ_2 data-fidelity term $f_A(\mathbf{x}) = \|\boldsymbol{\Phi}\mathbf{x} - \mathbf{y}\|_2^2$. Clearly in the analysis formulation the optimization problems are precisely those given in Section III-A. However, in the synthesis settings the regularization functional takes the form $g_S(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_1$, while the data-fidelity term is given by $f_S(\boldsymbol{\alpha}) = \|\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} - \mathbf{y}\|_2^2$. With these definitions the synthesis optimization problem reads in much the same way as those presented in Section III-A and, in fact, for situations in which the measurement operator is orthogonal, *i.e.* $\boldsymbol{\Psi}^{-1} = \boldsymbol{\Psi}^\dagger$, these formulations are equivalent. However, they have very different geometric properties when this is not the case [24]–[26]. Notice that we adopt overcomplete spherical wavelet transforms where $\boldsymbol{\Psi}^{-1} \neq \boldsymbol{\Psi}^\dagger$, and sampled spherical harmonic transforms which are not orthogonal, *i.e.* $\mathbf{Y}^{-1} \neq \mathbf{Y}^\dagger$ [12] — a notable difference to the discrete Fourier transform in Euclidean settings. Therefore on the sphere the analysis and synthesis settings are not equivalent, and often produce noticeably different results.

In practice the analysis setting has consistently been demonstrated to exhibit greater reconstruction fidelity, a feature attributed to the lower cardinality of the analysis solution space [24]–[26]. However, in previous work it was concluded that this characteristic may not be generalized to the spherical setting [21]. In Section V we revisit this analysis and find that the variation in relative performance, both in terms of reconstruction fidelity and computational efficiency, of each setting is dependent on the problem instance under consideration. Therefore, flexibility with respect to reconstruction formulation supports development of scalable spherical imaging algorithms tailored for specific applications. In this work we discover that implicit bandlimiting is often a determining factor when one considers inverse problems on the sphere, which impacts the effective cardinality of the spaces considered. In this sense it is beneficial to either (i) adopt the synthesis setting in which signals are implicitly bandlimited during reconstruction or (ii) explicitly bandlimit the analysis setting, which in some settings can be computationally inefficient on the sphere. An example of such computational savings in the synthesis setting is demonstrated in Section V-C, where an iterative Wiener filtering approach is adopted. In this case the analysis/synthesis formulations require 6, 2 spherical harmonic transforms respectively.

C. Regularization functionals on the sphere

Having discussed the variety of ways one may formulate and construct an optimization on the sphere, we should now consider spherical counterparts to common regularization functionals and how one can develop these for the spherical setting. Such regularization functionals include *e.g.* sparsity promoting $\|\cdot\|_1$ regularizers, typically in a sparsifying dictionary $\boldsymbol{\Psi}$, which are often motivated by the theory of compressed sensing [7]–[9]; Gaussian $\|\cdot\|_2^2$ regularizers, which are often iterative implementations of harmonic Wiener filters [50], [51]; and spherical total-variation (TV) priors [12], which are effective for edge detection and segmentation tasks.

Most imaging problems exist in the discrete settings, and so depend on approximations to the underlying continuous ℓ_p -norms. In spherical settings one often adopts equiangular sampling [38], which does not uniformly sample the continuous norms. Typically this results in disproportionate weight being applied to pixels located at the poles, due to progressing increased sampling density away from the equator. To account for this spherical (directional wavelet) counterparts to the traditional norms are defined by

$$\mathbb{S}^2\|\mathbf{x}\|_p = \|\mathbf{w} \circ \mathbf{x}\|_p \Rightarrow \mathbb{S}^2\|\boldsymbol{\alpha}\|_p = \left(\sum_j \sum_n \mathbb{S}^2\|\alpha_{j,n}\|_p^p \right)^{\frac{1}{p}}, \quad (8)$$

respectively, where $\mathbf{w} \in \mathbb{S}^2$ is the corresponding map of reciprocal pixel areas on the sphere, \circ is the Hadamard product, and $j, n \in \mathbb{Z}^+$ are wavelet scale and direction respectively. This reformulation provides a closer approximation to the underlying continuous ℓ_p -norm on the sphere.

With these corrected norms one can straightforwardly consider, *e.g.*, sparsity in spherical wavelet space $\mathbb{S}^2\|\boldsymbol{\Psi}^\dagger\mathbf{x}\|_1$. Such a generalization permits multi-resolution algorithms [17] resulting in wavelet scale projections of varying resolution,

which provide a significant increase in computational efficiency, a fundamental bottleneck of variational methods on the sphere. In theory one could leverage the exact quadrature weights inherent to the underlying spherical sampling theorems [38], however in this work we find simple weights $w_{j,n}$ which capture the pixel area to be sufficient.

D. Efficient flexible imaging on the sphere

Variational approaches efficiently locate optimal solutions via iterative algorithms, which typically leverage 1st-order (gradient) information to navigate towards extremal values. Furthermore, for convex objectives, such algorithms permit strong guarantees of both convergence and the rate of convergence. Imaging problems often adopt non-differentiable regularization functionals (e.g. ℓ_1 -norms) for which proximal operators may be used to navigate the objective function, thus motivating proximal convex optimization algorithms.

Convex optimization algorithms require successive iterations to converge; as such, any operators evaluated must be efficient and precise, so as to facilitate accurate, scalable methods. These considerations are more pronounced when considering optimization over spherical manifolds, wherein underlying operators (e.g. spin- s spherical harmonic transforms) scale poorly with dimension ($\propto \mathcal{O}(L^3)$ in the best case scenario). Additionally, a large subset of optimization algorithms require adjoint \dagger operators, which are often incorrectly approximated by their inverse operators, introducing unpredictable errors and breaking convergence guarantees. Furthermore, on the sphere one must also consider the weighting scheme presented in Section III-C, which can be incorporated into proximal optimization algorithms through; a direct operator that performs the weighting, or by weighting norms. To avoid additional complications (e.g. under certain norms weighting operators do not represent tight frames, necessitating additional sub-iterations) we simply weight the norms directly.

During this research we developed a highly optimized object oriented (OOP) C++ software framework (S2INV) which permits all the aforementioned flexibility. The equiangular sampling theorem on the sphere of [38] is adopted through the SSHT² package, which permits fast and efficient spin- s spherical harmonic transforms, whilst permitting machine precision computation. Additionally, we adopt optimized scale-discretized directional wavelets on the sphere [16], [18], [19] through the S2LET³ package [17], [21], which are optimally sampled and support machine precision synthesis. We leverage a recently developed, highly optimized C++ OOP sparse optimization framework SOPT⁴ [11], [36], [37], which facilitates a variety of proximal convex optimization algorithms, e.g. forward-backward [52], [53], primal dual [27], [54], [55], and the alternating direction method of multipliers [56], with appropriate modifications for the spherical setting. In this way S2INV provides a scalable, flexible, open-source software package, which is fully customizable and supports a wide

variety of novel, fully principled, uncertainty quantification techniques on the sphere, which we discuss in Section IV.

IV. SPHERICAL BAYESIAN UNCERTAINTY QUANTIFICATION

The unconstrained reconstruction problem has a straightforward Bayesian interpretation which is as follows. The posterior distribution of a spherical image $\mathbf{x} \in \mathbb{C}^{N_{s^2}}$ defined over, e.g., the celestial sphere or the globe, given observations $\mathbf{y} \in \mathbb{C}^M$ is given by Bayes' theorem,

$$P(\mathbf{x}|\mathbf{y}; \mathcal{M}) \equiv \frac{P(\mathbf{y}|\mathbf{x}; \mathcal{M})P(\mathbf{x}; \mathcal{M})}{\int_{\mathbb{C}^{N_{s^2}}} P(\mathbf{y}|\mathbf{x}; \mathcal{M})P(\mathbf{x}; \mathcal{M})d\mathbf{x}}, \quad (9)$$

where the likelihood encodes data fidelity, the prior encodes *a priori* information of the image, and \mathcal{M} represents some model, which includes the mapping $\Phi \in \mathbb{C}^{M \times N_{s^2}} : \mathbf{x} \mapsto \mathbf{y}$, and some understanding of the noise inherent to \mathbf{y} [28]. Note that the marginal likelihood (Bayesian evidence) is a constant scaling of the posterior and can be used for model comparison, which we do not consider further in this article.

Typically sampling methods, e.g. Markov chain Monte Carlo, are adopted to sample from the posterior distribution from which one can determine a point estimation of the solution to the inverse problem and the distribution of uncertainty about such a solution. Although these methods recover asymptotically exact estimates of the posterior distribution, they typically require large numbers of samples to converge. Each sample requires at least a single evaluation of the posterior which in spherical settings is computationally demanding — for moderate resolutions $L > 10^3$ sampling methods rapidly become computationally intractable.

Instead consider a variational approach that maximizes the posterior odds, referred to as the *maximum a posteriori* (MAP) solution defined by

$$\begin{aligned} \mathbf{x}^{\text{MAP}} &\equiv \underset{\mathbf{x}}{\operatorname{argmax}} \{P(\mathbf{x}|\mathbf{y}; \mathcal{M})\}, \\ &\propto \underset{\mathbf{x}}{\operatorname{argmin}} \{ -\log(P(\mathbf{y}|\mathbf{x}; \mathcal{M})P(\mathbf{x}; \mathcal{M})) \}, \\ &\propto \underset{\mathbf{x}}{\operatorname{argmin}} \{h(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})\}, \end{aligned} \quad (10)$$

where the second line follows by the monotonicity of the logarithm function. For convex objective functions $h(\mathbf{x})$ this takes the form of a convex optimization problem [27], and therefore Equation 6 explicitly returns the MAP solution, as asserted in Section III-A. Hence, leveraging state-of-the-art convex optimization techniques one can efficiently locate the solution which maximizes the posterior odds. However this is still a point estimate which, though useful, does not naively support uncertainty quantification. Recently, approximate contours of the latent space have been derived facilitating variational regularization methods with principled uncertainty quantification. We discuss these approximate methods and develop uncertainty quantification techniques on the sphere, which we accelerate by exploiting function linearity.

A. Highest posterior density credible regions

A credible region $C_\alpha \subset \mathbb{C}^{N_{s^2}}$ of the posterior latent space at credible confidence $100(1 - \alpha)\%$, for $\alpha \in [0, 1]$, satisfies

²<https://astro-informatics.github.io/ssht/>

³<https://astro-informatics.github.io/s2let/>

⁴<http://astro-informatics.github.io/sopt/>

the integral equation [28]

$$P(\mathbf{x} \in C_\alpha | \mathbf{y}; \mathcal{M}) = \int_{\mathbf{x} \in \mathbb{C}^{N_{s^2}}} P(\mathbf{x} | \mathbf{y}; \mathcal{M}) \mathbb{I}_{C_\alpha} d\mathbf{x} = 1 - \alpha, \quad (11)$$

where \mathbb{I}_{C_α} is the standard set indicator function. The optimal credible region in the sense of minimal volume [28] is the highest posterior density (HPD) credible region defined by $C_\alpha := \{\mathbf{x} : h(\mathbf{x}) \leq \epsilon_\alpha\}$, where $\epsilon_\alpha \in \mathbb{R}^+$ is an isocontour of the log-posterior. Determination of the HPD region requires computation of the integral in Equation 11, which is computationally infeasible in even moderate dimensional spherical settings, due to dimensionality and functional complexity considerations. Convex objectives $h(\mathbf{x})$ support the conservative approximate HPD credible region C'_α defined by [30]

$$C_\alpha \subseteq C'_\alpha \subset \mathbb{C}^{N_{s^2}} := \{\mathbf{x} : h(\mathbf{x}) \leq \epsilon'_\alpha\},$$

where $\epsilon'_\alpha = h(\mathbf{x}^{\text{MAP}}) + \sqrt{16N \log(3/\alpha)} + N$, (12)

which allows one to approximate C_α with only knowledge of the MAP solution \mathbf{x}^{MAP} and the dimension $N = \mathbb{C}^{N_{s^2}}$. An upper bound on the approximation error exists [30]. Therefore for convex objectives, given \mathbf{x}^{MAP} , one may draw statistically principled conclusions. This credible set approximation has been leveraged to develop fast Bayesian uncertainty quantification techniques in a variety of settings [30]–[35].

B. Bayesian hypothesis testing on the sphere

The most straightforward uncertainty quantification technique one may generate by leveraging the approximation of Equation 12 is that of hypothesis testing [1], [31], [33]. The concept of hypothesis testing is to adjust a feature of the recovered estimator \mathbf{x}^{MAP} generating a surrogate solution \mathbf{x}^{sur} , of which we ask $\mathbf{x}^{\text{sur}} \in C'_\alpha$? If $\mathbf{x}^{\text{sur}} \notin C'_\alpha \Rightarrow \mathbf{x}^{\text{sur}} \notin C_\alpha$, which follows from the conservative nature of the approximation in Equation 12, the feature of interest is considered to be statistically significant (necessary to the reconstruction) at $100(1 - \alpha)\%$ confidence. Conversely $\mathbf{x}^{\text{sur}} \in C'_\alpha$ indicates that the surrogate solution remains within the approximate credible set and we conclude that the feature is indeterminate.

C. Local credible intervals on the sphere

Suppose one recovers an optimal solution \mathbf{x}^{MAP} through unconstrained convex optimization (see Section III-A) and wishes to quantify the uncertainty associated with a given pixel or super pixel (collection of pixels). With knowledge of the approximate level set threshold ϵ'_α , and therefore the approximate HPD credible set C'_α at well defined confidence $(100 - \alpha)\%$, one simply needs to iteratively compute the extremal values a given region of interest may take, such that the resulting solution falls outside of the approximate HPD credible set, *i.e.* $\mathbf{x}^{\text{sur}} \notin C'_\alpha$. One must then define which types of regions (super-pixels) on the sphere one is interested in.

Formally, select independent partitions of the latent space $\Omega = \cup_i \Omega_i$ for which we define super-pixel indexing functions $\zeta_{(\cdot)}$ such that $\mathbf{x}_i \in \Omega_i \Rightarrow \zeta_{\Omega_i} = 1$ and $\mathbf{x}_i \notin \Omega_i \Rightarrow \zeta_{\Omega_i} = 0$. For a given Ω_i locate upper (lower) bounds $\xi_{\Omega_i}^+$, $\xi_{\Omega_i}^-$ respectively,

which saturate the HPD credible region C'_α [31], [34]. This is achieved by the following optimizations,

$$\xi_{\Omega_i}^\pm = \max_{\xi \in \mathbb{R}^+} \{ \pm \xi | f(\mathbf{x}_{i,\xi}) + g(\mathbf{x}_{i,\xi}) \leq \epsilon'_\alpha \}, \quad (13)$$

where $\mathbf{x}_{i,\xi} = \mathbf{x}^{\text{MAP}} \zeta_{\Omega/\Omega_i} + \xi \zeta_{\Omega_i}$ is a surrogate solution where the super pixel region has been replaced by a uniform intensity ξ . The collective set of these bounds $\{|\xi_{\Omega_i}^+ - \xi_{\Omega_i}^-|\}$ is taken to be the local credible interval map [31], which can simply be recovered *via* bisection. Though conditional local uncertainty quantification techniques such as this have demonstrated utility in certain circumstances [31], [33]–[35], in the high dimensional spherical settings they can quickly become dilute [1]. This makes intuitive sense, as small (local) objects (super-pixels) in high dimensional settings become statistically insignificant. As such, in high-dimensional spherical settings global or aggregate (statistical) uncertainty quantification techniques are often more meaningful (see *e.g.* [1]).

1) *Gridding schemes*: One can construct rectangular partitions directly on the latent space (*e.g.* uniform gridding). However, in the spherical setting it is sometimes more meaningful to define a super pixel by a fixed physical area surrounding a defined central pixel. Practically this is computed as follows: define a central pixel on the sphere, rotate this pixel to a pole (where higher angular resolution provides greater fine tuning of super-pixel area), select a given angular deviation from the pole, define this spherical cap as the super pixel. In this way all super pixels are, by definition, of equal area.

D. Acceleration through linearity

Naive computation of local credible intervals through bisection can often require many evaluations of the objective function, which is particularly costly when one considers functions on the sphere. To avoid this computational bottleneck we exploit the linearity of such operators. Consider the generalized convex objective function for the analysis setting, without loss of generality, which can be written as

$$h(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) = \|\Phi \mathbf{x} - \mathbf{y}\|_{p_1}^{p_2} + \|\Psi^\dagger \mathbf{x}\|_{q_1}^{q_2}. \quad (14)$$

Consider again the partition $\mathbf{x}^{\text{sur}} = \mathbf{x} \zeta_{\Omega/\Omega_i} + \xi \zeta_{\Omega_i}$ upon which the applications of any linear operator \mathcal{L} is given trivially by linearity to be $\mathcal{L} \mathbf{x}^{\text{sur}} = \mathcal{L} \mathbf{x} \zeta_{\Omega/\Omega_i} + \xi \mathcal{L} \zeta_{\Omega_i}$. Explicitly expanding Equation 14 with linearity one finds

$$\begin{aligned} & \|\Phi \mathbf{x} \zeta_{\Omega/\Omega_i} + \xi \Phi \zeta_{\Omega_i} - \mathbf{y}\|_{p_1}^{p_2} + \|\Psi^\dagger \mathbf{x} \zeta_{\Omega/\Omega_i} + \xi \Psi^\dagger \zeta_{\Omega_i}\|_{q_1}^{q_2}, \\ & \Rightarrow \|\mathbf{a} + \xi \mathbf{b}\|_{p_1}^{p_2} + \|\mathbf{c} + \xi \mathbf{d}\|_{q_1}^{q_2}, \end{aligned} \quad (15)$$

for constant (per credible interval) vectors defined to be

$$\begin{aligned} \mathbf{a} &= \Phi \mathbf{x} \zeta_{\Omega/\Omega_i} - \mathbf{y} & \mathbf{b} &= \Phi \zeta_{\Omega_i} \\ \mathbf{c} &= \Psi^\dagger \mathbf{x} \zeta_{\Omega/\Omega_i} & \mathbf{d} &= \Psi^\dagger \zeta_{\Omega_i} \end{aligned} \quad (16)$$

In this way the local credible optimization problems in Equation 13 can be re-written instead as

$$\|\mathbf{a} + \xi \mathbf{b}\|_{p_1}^{p_2} + \|\mathbf{c} + \xi \mathbf{d}\|_{q_1}^{q_2} \leq \epsilon'_\alpha, \quad (17)$$

which is clearly just a 1-dimensional polynomial root finding problem. One could approach this inequality from an iterative perspective, forming an upper bound through the

Minkowski inequality, which is then leveraged as initialization for bisection. For polynomials of order < 5 (see Abel-Ruffini theorem) Equation 17 permits analytic solutions. In practice the computational difference between the analytic solution and solving an inequality bounded bisection problem is marginal, though in high dimensions this speed up is non-negligible.

1) *Gaussian regression*: Suppose one adopts both a Gaussian likelihood and prior (e.g. iterative Wiener filtering approaches), in such a setting we have $p_1 = p_2 = q_1 = q_2 = 2$ which reduces Equation 17 to the binomial inequality $\|\mathbf{a} + \xi\mathbf{b}\|_2^2 + \|\mathbf{c} + \xi\mathbf{d}\|_2^2 \leq \epsilon'_\alpha$, which expands to give

$$[\|\mathbf{b}\|_2^2 + \|\mathbf{d}\|_2^2]\xi^2 + 2[\mathbf{a} \cdot \mathbf{b} + \mathbf{c} \cdot \mathbf{d}]\xi + [\|\mathbf{a}\|_2^2 + \|\mathbf{c}\|_2^2] \leq \epsilon'_\alpha. \quad (18)$$

One could gain some geometric insight by considering the case in which $\mathbf{a} \cdot \mathbf{b} + \mathbf{c} \cdot \mathbf{d} = 0$, as in such a case the resulting credible region about the posterior is symmetric, however we do not consider this further here. Nonetheless, per credible region the interval is trivially recovered.

2) *Lasso regression*: Consider the Lagrangian dual of the Lasso regression (e.g. sparse reconstruction), in which we have $p_1 = p_2 = 2$ and $q_1 = q_2 = 1$, such that the general polynomial Equation 17 reduces to a 2nd-order polynomial $\|\mathbf{a} + \xi\mathbf{b}\|_2^2 + \|\mathbf{c} + \xi\mathbf{d}\|_1 \leq \epsilon'_\alpha$, which, assuming the intersection of the partitions projected into Ψ is negligible, i.e. the dictionary Ψ has sufficient localization properties on the sphere, results in the inequality

$$\|\mathbf{b}\|_2^2 \xi^2 + 2[\mathbf{a} \cdot \mathbf{b} + \|\mathbf{d}\|_1]\xi + [\|\mathbf{a}\|_2^2 + \|\mathbf{c}\|_1] \leq \epsilon'_\alpha, \quad (19)$$

which can be analytically solved. Typically the partitions Ω_i projected into Ψ exhibit overlapping support and so this inequality is not exact. In such cases one can perform bisection, computing only the ℓ_1 term at each iteration.

V. NUMERICAL EXPERIMENTS

In this section we showcase the variational regularization and uncertainty quantification techniques presented in Section III and Section IV on a diverse set of numerical experiments. For each scenario we create mock observations $\mathbf{y} = \Phi\mathbf{x}$ of a ground truth signal \mathbf{x} which are related through the forward model $\Phi : \mathbf{x} \mapsto \mathbf{y}$ from which we formulate an ill-posed inverse problem (e.g. add noise, mask, blur, etc.). We solve this inverse problem for an estimator \mathbf{x}^* (\mathbf{x}^{MAP}) of \mathbf{x} the success of which we quantify by the recovered *signal to noise ratio*, defined by $\text{SNR} = 20 \log_{10}(\|\mathbf{x}\|_2 / \|\mathbf{x} - \mathbf{x}^*\|_2)$.

A. Earth satellite topography

Suppose a satellite performs observations $\mathbf{y} \in \mathbb{R}^M$ of the Earth's topography (geographic elevation) which can be related to the true topography $\mathbf{x} \in \mathbb{R}^{N_{S^2}}$ through a mapping (forward model) $\Phi_{\text{ET}} \in \mathbb{R}^{M \times N_{S^2}} : \mathbf{x} \mapsto \mathbf{y}$. Consider the scenario in which incomplete, i.e. $|M| \ll |N_{S^2}|$, observations \mathbf{y} are contaminated with independent and identically distributed (i.i.d.) Gaussian noise $\mathbf{n} \sim \mathcal{N}(0, \sigma^2) \in \mathbb{R}^M$, and blurred with an axisymmetric smoothing kernel with full width half maximum (FWHM) Θ . In such a case observations are modeled by $\mathbf{y} = \Phi_{\text{ET}}\mathbf{x} + \mathbf{n}$ for measurement operator

$$\Phi_{\text{ET}} = \mathbf{D}\mathbf{Y}^{-1}\Theta\mathbf{Y} \quad \text{and} \quad \Phi_{\text{ET}}^\dagger = \mathbf{Y}^\dagger\Theta(\mathbf{Y}^{-1})^\dagger\mathbf{D}^\dagger, \quad (20)$$

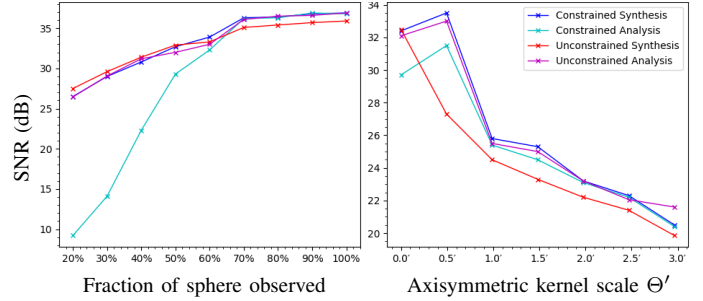


Fig. 1. **Left**: Recovered SNR for a variety of problem setups over a variety of inpainting scenarios versus % of pixels observed. **Right**: Recovered SNR for a variety of problem setups versus variety of axisymmetric blurring kernel scales, for a fixed inpainting of 50% masked pixels. **Discussion**: Generally each setup performs similarly and it would appear that no single setting is optimal in all cases. Notice the drastic underperformance of the analysis constrained formulation in the heavily masked regime for the inpainting problem. We find that this asymmetry is due to the implicit spherical harmonic bandlimiting inherent to the synthesis problem (see text).

where $\mathbf{Y}, \mathbf{Y}^{-1}$ are forward and inverse spherical harmonic transforms correspondingly, $\mathbf{D}, \mathbf{D}^\dagger$ are masking and projection operators correspondingly, Θ is the axisymmetric convolution with the harmonic representation of Θ which is trivially self-adjoint, and \dagger represents the adjoint operation.

As \mathbf{n} is a univariate Gaussian the data-fidelity (log-likelihood) term is simply given by $\frac{1}{2\sigma^2} \|\Phi_{\text{ET}}\mathbf{x} - \mathbf{y}\|_2^2$. Here we adopt a sparsity promoting ℓ_1 -norm wavelet regularization $\Psi\|\Psi^\dagger(\cdot)\|_1$ (Laplace distribution log-prior), and solve both the constrained formulation through the proximal ADMM algorithm [56] and the unconstrained formulation through the proximal forward backward algorithm [27], [52], [53], in both the analysis and synthesis settings. Notice the use of spherical (wavelet) space norms, outlined in Section III-C, which better approximate spherical continuous norms. To quantify the impact of analysis versus synthesis (and constrained versus unconstrained) settings we consider all settings in two paradigms (i) varying levels of inpainting without deconvolution (ii) varying scales of deconvolution with 50% masked pixel inpainting. Generally, each problem setup performs comparably in all settings considered (see Figure 1). Certainly it cannot be said that one reconstruction paradigm is optimal in all settings, which leads us to conclude that it is likely that problem formulation optimality is ambiguous and should be selected on a case by case basis. Interestingly, notice the underperformance of the constrained analysis problem in the heavily masked regime (see the left plot of Figure 1). This was observed in prior analysis [21] and reported as evidence that the synthesis setting may produce more optimal results.

Note that the synthesis setting implicitly bandlimits the observations \mathbf{x} , therefore restricting the solution space cardinality — a factor known to impact reconstruction fidelity [24]. To account for this bias we reran the analysis optimization with an explicitly bandlimited measurement operator, results of which can be seen in Figure 2 (which also demonstrates the uncertainty quantification of Section IV-C). It was found that the analysis setting performs similarly to the synthesis setting for this problem, leading us to conclude that the optimality of optimization formulation is at best ambiguous.

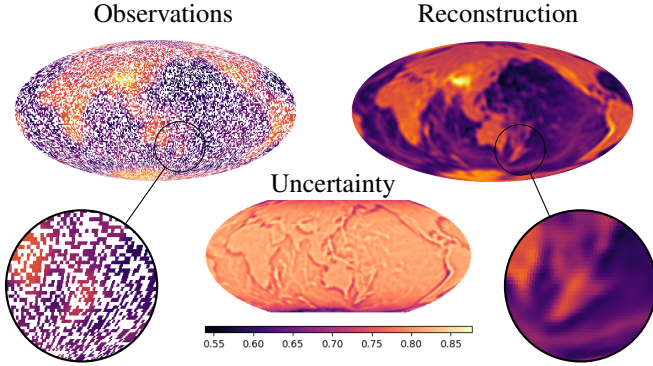


Fig. 2. **Left:** Simulated observations contaminated with 30dB Gaussian *i.i.d.* noise, convolved with a ~ 268 arc-minute Gaussian blurring kernel, and with 60% of pixels masked. **Right:** Unconstrained reconstruction using ℓ_1 -norm wavelet sparsity regularization (log-prior) in the analysis setting, adopting the proximal forward-backward algorithm [27], [52], [53]. **Bottom:** Bayesian local credible intervals at 99% confidence (reconstruction values are $\in [-1, 1]$), which are pixel-level uncertainties (see Section IV).

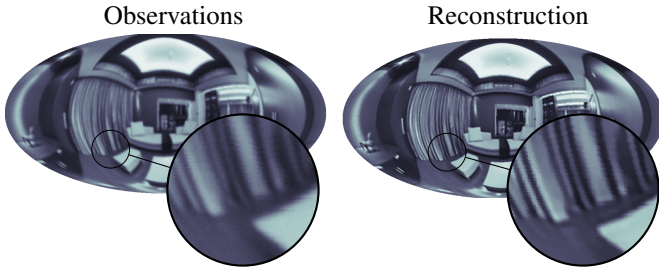


Fig. 3. **Left:** Simulated observations contaminated with 30dB Gaussian noise and convolved with a ~ 78 arc-minute Gaussian blurring kernel (*e.g.* motion blur). **Right:** Spherical TV-norm $\|\mathbf{x}\|_{\text{TV}}$ regularized reconstruction through proximal primal dual algorithm in the analysis setting.

B. 360° camera blur deconvolution

Suppose a 360° camera captures a greyscale spherical image $\mathbf{y} \in \mathbb{R}^{M_{\mathbb{S}^2}}$ which can be related to the true image $\mathbf{x} \in \mathbb{R}^{N_{\mathbb{S}^2}}$ through the forward model $\Phi_{360^\circ} \in \mathbb{R}^{M_{\mathbb{S}^2} \times N_{\mathbb{S}^2}} : \mathbf{x} \mapsto \mathbf{y}$. Consider that the camera captures complete $|M_{\mathbb{S}^2}| = |N_{\mathbb{S}^2}|$ observations but introduces low-level *i.i.d.* Gaussian noise $\mathbf{n} \sim \mathcal{N}(0, \sigma^2) \in \mathbb{R}^{M_{\mathbb{S}^2}}$ and a certain amount of lens blurring characterized by axisymmetric convolution with a Gaussian smoothing kernel with FWHM = Θ . In this case observations are modeled by $\mathbf{y} = \Phi_{360^\circ} \mathbf{x} + \mathbf{n}$ for measurement operator

$$\Phi_{360^\circ} = \mathbf{Y}^{-1} \Theta \mathbf{Y} \quad \text{and} \quad \Phi_{360^\circ}^\dagger = \mathbf{Y}^\dagger \Theta (\mathbf{Y}^{-1})^\dagger, \quad (21)$$

where $\mathbf{Y}, \mathbf{Y}^{-1}$ are forward and inverse spherical harmonic transforms correspondingly (see section II), Θ is the axisymmetric convolution with the harmonic representation of Θ .

As in the previous example the data-fidelity is given by the \mathbb{S}^2 $\|\Phi_{360^\circ} \mathbf{x} - \mathbf{y}\|_2^2$. Depending on the degree to which \mathbf{x} is piece-wise constant the TV-norm $\mathbb{S}^2 \|\mathbf{x}\|_{\text{TV}} = \mathbb{S}^2 \|\nabla \mathbf{x}\|_2$ (promoting gradient sparsity) constitutes a good choice of regularizer. For image deconvolution an analysis wavelet sparsity promoting regularizer $\Psi \|\Psi^{-1} \mathbf{x}\|_1$ is often also considered. Here we consider both regularization functionals $g(\mathbf{x}) = \{\mathbb{S}^2 \|\mathbf{x}\|_{\text{TV}}, \Psi \|\Psi^{-1} \mathbf{x}\|_1\}$ in the constrained analysis setting:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{C}^{N_{\mathbb{S}^2}}}{\operatorname{argmin}} \{g(\mathbf{x})\} \text{ s.t. } \frac{1}{2\sigma^2} \mathbb{S}^2 \|\Phi_{360^\circ} \mathbf{x} - \mathbf{y}\|_2^2 \leq \delta, \quad (22)$$

where $\delta \in \mathbb{R}^+$ is the radius of the ℓ_2 -ball $\mathcal{B}_{\ell_2}^\delta$ which balances

sparsity against data-fidelity, and is defined straightforwardly from the known (in general unknown) noise variance. We perform an example reconstruction with both priors in the constrained formulation of the analysis problem using the proximal primal dual algorithm [27], [54], [55]. Both priors produce similar results, with wavelet sparsity regularization recovering SNR=17.60 dB and TV-norm marginally superior at SNR = 17.65 dB — which can be seen in Figure 3.

C. CMB temperature anisotropies

Suppose one captures masked (and therefore incomplete) measurements of the cosmic microwave background (CMB; [3]) $\mathbf{y} \in \mathbb{C}^{M_{\mathbb{S}^2}}$ that can be related through a mapping operator $\Phi_{\text{CMB}} \in \mathbb{C}^{M_{\mathbb{S}^2} \times N_{\mathbb{S}^2}}$ to the full-sky CMB signal $\mathbf{x} \in \mathbb{C}^{N_{\mathbb{S}^2}}$ which can be decomposed into harmonic coefficients $\hat{x}_{\ell m} = \langle \mathbf{x}, Y_{\ell m} \rangle$ which for Gaussian fields such as CMB [3] are uncorrelated and isotropic $\mathbb{E}[\hat{x}_{\ell m}^* \hat{x}_{\ell' m'}] = \delta_{\ell \ell'} \delta_{m m'} C_\ell$, where C_ℓ is the angular power spectrum.

These considerations motivate the choice of a multivariate Gaussian prior $P(\hat{\mathbf{x}}|\mathcal{M}) = \exp(-\hat{\mathbf{x}}^\dagger \mathbf{C}^{-1} \hat{\mathbf{x}}/2)$ for vectorized harmonic coefficients $\hat{\mathbf{x}}$ and covariance \mathbf{C} given by diagonal elements C_ℓ . Consider the case in which $|M_{\mathbb{S}^2}| \ll |N_{\mathbb{S}^2}|$ with *i.i.d.* Gaussian noise $\mathbf{n} \sim \mathcal{N}(0, \sigma^2) \in \mathbb{C}^{M_{\mathbb{S}^2}}$ then the whitened harmonic coefficients $\hat{\mathbf{x}}' = \mathbf{C}^{-\frac{1}{2}} \hat{\mathbf{x}}$ are modeled as $\mathbf{y} = \Phi_{\text{CMB}} \hat{\mathbf{x}}' + \mathbf{n}$ for measurement operator

$$\Phi_{\text{CMB}} = \mathbf{D} \mathbf{Y} \mathbf{C}^{\frac{1}{2}} \quad \text{and} \quad \Phi_{\text{CMB}}^\dagger = \mathbf{C}^{\frac{1}{2}} \mathbf{Y}^\dagger \mathbf{D}^\dagger, \quad (23)$$

for spherical harmonic transform \mathbf{Y} and masking and projection operators $\mathbf{D}, \mathbf{D}^\dagger$ respectively. For diagonal noise covariance $\Sigma = \sigma \mathbb{I}$ the univariate Gaussian likelihood is given by $P(\mathbf{y}|\hat{\mathbf{x}}'; \mathcal{M}) = \frac{1}{2\sigma^2 \mathbb{S}^2} \|\Phi_{\text{CMB}} \hat{\mathbf{x}}' - \mathbf{y}\|_2^2$, and so the synthesis unconstrained optimization is given by

$$\hat{\mathbf{x}}^{\text{MAP}'} = \underset{\hat{\mathbf{x}}'}{\operatorname{argmin}} \left\{ \|\hat{\mathbf{x}}'\|_2^2 + \frac{1}{2\sigma^2 \mathbb{S}^2} \|\Phi_{\text{CMB}} \hat{\mathbf{x}}' - \mathbf{y}\|_2^2 \right\}, \quad (24)$$

where the pixel space signal is recovered by $\mathbf{x}^{\text{MAP}} = \mathbf{Y}^{-1} \mathbf{C}^{\frac{1}{2}} \hat{\mathbf{x}}^{\text{MAP}'}$. This is the convex optimization formulation of what is commonly known as Wiener filtering, which is often adopted for highly Gaussian fields, *e.g.* the CMB. As expected for Wiener filtering problems of this type (*e.g.* Figure 6 in [51]), we recover maps which exhibit inpainting of low- ℓ modes (large scale structure) into the masked regions. The results of this experiment can be seen in Figure 4.

D. Weak gravitational lensing

The following example considers spherical imaging of dark matter. A more extensive analysis that applies the method presented to observational data, and not just simulations, is performed in [1] which leverages many of the methods developed in this work. At first order, gravitational lensing manifests itself into the spin-0 convergence ${}_0\phi(r, \omega) \in \mathbb{C}^{N_{\mathbb{S}^2}}$ (the integrated matter field along the line of sight) and the spin-2 shear ${}_2\mathbf{y}(r, \omega) \in \mathbb{C}^{N_{\mathbb{S}^2}}$ (the ellipticity of observed images) which can be related to the lensing potential ${}_0\phi(r, \omega) \in \mathbb{C}^{N_{\mathbb{S}^2}}$ by ${}_0\mathbf{x}(r, \omega) = \frac{1}{4}(\bar{\partial}_+ \bar{\partial}_- + \bar{\partial}_- \bar{\partial}_+) {}_0\phi$ and ${}_2\mathbf{y}(r, \omega) = \frac{1}{2}\bar{\partial}_+ \bar{\partial}_+ {}_0\phi$, where $\bar{\partial}_\pm$ are spin-raising/lowering operators [39], [40]. One can then relate ${}_0\mathbf{x}$ and ${}_2\mathbf{y}$ to one another in harmonic space by

${}_2\hat{\mathbf{y}}_{\ell m} = W_\ell {}_0\hat{\mathbf{x}}_{\ell m}$, for harmonic space kernel defined in [1], [2]. As ${}_0\mathbf{x}$ is not directly observable, typically observations of ${}_2\mathbf{y}$ are collected and used to reconstruct ${}_0\mathbf{x}$. Suppose one recovers observations $\mathbf{y} \in \mathbb{C}^M$ which can be related to the $\mathbf{x} \in \mathbb{C}^{N_{S^2}}$ via the forward model $\Phi_{\text{WL}} \in \mathbb{C}^{M \times N_{S^2}} : \mathbf{x} \mapsto \mathbf{y}$. Consider the scenario in which \mathbf{y} are contaminated with *i.i.d.* Gaussian noise $\mathbf{n} \sim \mathcal{N}(0, \sigma^2) \in \mathbb{C}^M$, then the observations are modeled by $\mathbf{y} = \Phi_{\text{WL}}\mathbf{x} + \mathbf{n}$ for measurement operator

$$\Phi_{\text{WL}} = \mathbf{D}_2 \mathbf{Y}^{-1} \mathbf{W}_0 \mathbf{Y} \quad \text{and} \quad \Phi_{\text{WL}}^\dagger = {}_0 \mathbf{Y}^\dagger \mathbf{W}_2 (\mathbf{Y}^{-1})^\dagger \mathbf{D}^\dagger, \quad (25)$$

for self-adjoint harmonic space multiplication \mathbf{W} with the axisymmetric kernel W_ℓ , masking and projection operators $\mathbf{D}, \mathbf{D}^\dagger$, and spin- s forward and inverse spherical harmonic transforms ${}_s \mathbf{Y}, {}_s \mathbf{Y}^{-1}$ respectively.

Since principled statistical interpretation is crucial for this science application, one may consider the unconstrained formulation of this inverse problem which we solve here with the proximal forward-backward algorithm in the analysis setting, with univariate Gaussian likelihood (data-fidelity) and sparsity promoting Laplace type spherical wavelet prior (regularizer),

$$\mathbf{x}^{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{C}^{N_{S^2}}}{\text{argmin}} \left\{ \lambda_\Psi \|\Psi^{-1} \mathbf{x}\|_1 + \frac{1}{2\sigma^2} \|\Phi_{\text{WL}} \mathbf{x} - \mathbf{y}\|_2^2 \right\}, \quad (26)$$

for automatically marginalized regularization parameter $\lambda \in \mathbb{R}^+$ (see Section III-A). Note that sparse priors are often adopted in the weak lensing setting, recovering state-of-the-art results [1], [33], [57]. Images from this experiment can be seen in Figure 5. A more in depth application of the methods developed in this paper to dark matter reconstruction can be found in [1], in which global hypothesis testing (leveraging the techniques of Section IV-B) is performed to determine whether two reconstruction methods produce commensurate estimates.

VI. CONCLUSION

We present and discuss a flexible, general framework for variational imaging on the sphere. We consider different formulations of inverse problems as either constrained or unconstrained problems [27] in both the analysis and synthesis settings [21], [24]. The implications, advantages, and disadvantages of each choice within the context of imaging on the sphere is considered both qualitatively and quantitatively. Crucially, we highlight the direct relationship between the unconstrained setting and Bayesian inference. We combine this realization with recent developments in the field of probability density theory [30] to demonstrate how one can perform rapid, statistically principled uncertainty quantification on reconstructed signals (building upon work in [30]–[35]). Furthermore, we demonstrate mathematically how one may exploit linearity and general inequality relations to dramatically accelerate such uncertainty quantification techniques in all settings. It is shown that in a variety of interesting cases these uncertainty quantification techniques reduce to computationally trivial 1-dimensional P^{th} -order polynomial root finding problems, which can often be solved analytically. While such computational savings are key for scalable, statistically principled spherical imaging, they are likely also of use for standard 2-dimensional Euclidean imaging.

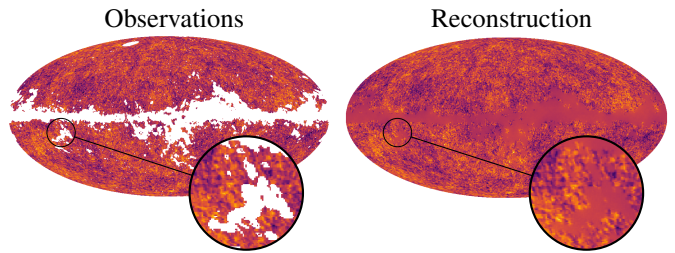


Fig. 4. **Left:** Simulated Gaussian random field generated from lambda cold dark matter (Λ -CDM) best fit power spectrum, masked by a Planck survey mask [3] and polluted by 30dB *i.i.d.* Gaussian noise. **Right:** Unconstrained reconstruction using an ℓ_2 -norm Wiener prior solved by the proximal forward-backward algorithm, in the synthesis setting (for computational efficiency). The purpose of this reconstruction is to observe recovered low- ℓ , large-scale information into the masked region (see e.g. [51]).

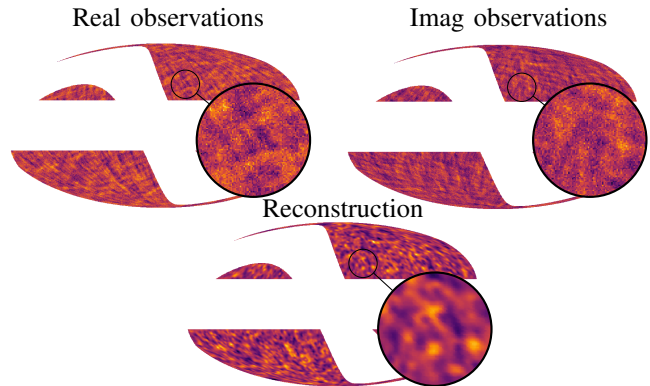


Fig. 5. **Top:** Simulated weak lensing shear field generated from a ground truth N-body simulation [58] signal, which was further contaminated with 5dB *i.i.d.* Gaussian noise and masked using realistic pseudo-Euclid survey mask. **Bottom:** Unconstrained reconstructed dark matter mass-map using ℓ_1 -norm wavelet sparsity prior solved through using the proximal forward-backward algorithm in the analysis setting. See related works [1] for a comprehensive analysis of this application, with uncertainty quantification.

The aforementioned techniques are demonstrated on an extensive suite of numerical experiments, which simulate a diverse set of typical use-cases. Specifically, we consider a spread of deconvolution, inpainting, and de-noising problems, e.g. from resolving blurred 360° camera images, to imaging the dark matter distribution on the celestial sphere. It is found that that optimality of problem formulation (constrained *versus* unconstrained) and setting (analysis *versus* synthesis) is highly situationally dependent on the sphere. The authors make the scalable, open-source spherical reconstruction software developed during this work (S2INV), publicly available.

ACKNOWLEDGMENT

MAP is supported by the Science and Technology Facilities Council (STFC). This work was also supported in part by the Leverhulme Trust. LP is supported by the Dunlap Institute, which is funded through an endowment established by the David Dunlap family and the University of Toronto.

REFERENCES

- [1] M. Price, J. McEwen, L. Pratley, and T. Kitching, “Sparse bayesian mass-mapping with uncertainties: Full sky observations on the celestial sphere,” *MNRAS*, vol. 500, no. 4, pp. 5436–5452, 2021.
- [2] C. G. R. Wallis *et al.*, “Mapping dark matter on the celestial sphere with weak gravitational lensing,” *ArXiv e-prints*, Mar. 2017.

- [3] P. Collaboration, "Planck 2018 results. I. Overview and the cosmological legacy of Planck," vol. 641, p. A1, Sept. 2020.
- [4] J. D. McEwen and M. A. Price, "Scale-discretised ridgelet transform on the sphere," in *27th European Signal Processing Conference (EU-SIPCO)*, 2019.
- [5] D. S. Tuch, "Q-ball imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 52, no. 6, pp. 1358–1372, 2004.
- [6] J. Ritsema, a. A. Deuss, H. Van Heijst, and J. Woodhouse, *Geophysical Journal International*, vol. 184, no. 3, pp. 1223–1236, 2011.
- [7] E. J. Candès *et al.*, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [8] E. J. Candès *et al.*, "Compressive sampling," in *Proceedings of the international congress of mathematicians*, vol. 3, 2006, pp. 1433–1452.
- [9] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [10] L. Pratley *et al.*, "A Fast and Exact w-stacking and w-projection Hybrid Algorithm for Wide-field Interferometric Imaging," *The Astrophysical Journal*, vol. 874, no. 2, p. 174, Apr. 2019.
- [11] A. Onose *et al.*, "Scalable splitting algorithms for big-data interferometric imaging in the SKA era," *MNRAS*, vol. 462, no. 4, pp. 4314–4335, Nov 2016.
- [12] J. D. McEwen *et al.*, "Sparse Image Reconstruction on the Sphere: Implications of a New Sampling Theorem," *IEEE Transactions on Image Processing*, vol. 22, pp. 2275–2285, June 2013.
- [13] F. J. Narcowich *et al.*, "Localized tight frames on spheres," *SIAM Journal on Mathematical Analysis*, vol. 38, no. 2, pp. 574–594, 2006.
- [14] P. Baldi *et al.*, "Asymptotics for spherical needlets," *The Annals of Statistics*, vol. 37, no. 3, pp. 1150–1171, 2009.
- [15] J.-L. Starck *et al.*, "Wavelets, ridgelets and curvelets on the sphere," *Astronomy & Astrophysics*, vol. 446, no. 3, pp. 1191–1204, 2006.
- [16] Y. Wiaux, J. D. McEwen, P. Vandergheynst, and O. Blanc, "Exact reconstruction with directional wavelets on the sphere," *MNRAS*, vol. 388, pp. 770–788, Aug. 2008.
- [17] B. Leistedt *et al.*, "S2LET: A code to perform fast wavelet analysis on the sphere," *aap*, vol. 558, p. A128, Oct. 2013.
- [18] J. D. McEwen, P. Vandergheynst, and Y. Wiaux, "On the computation of directional scale-discretized wavelet transforms on the sphere," in *SPIE Wavelets and Sparsity XV, invited contribution*, vol. 8858, 2013.
- [19] J. D. McEwen, B. Leistedt, M. Büttner, H. V. Peiris, and Y. Wiaux, "Directional spin wavelets on the sphere," *ArXiv e-prints*, Sept. 2015.
- [20] J. D. McEwen, C. Durastanti, and Y. Wiaux, "Localisation of directional scale-discretized wavelets on the sphere," *Applied and Computational Harmonic Analysis*, vol. 44, no. 1, pp. 59 – 88, 2018.
- [21] C. G. R. Wallis, Y. Wiaux, and J. D. McEwen, "Sparse Image Reconstruction on the Sphere: Analysis and Synthesis," *IEEE Transactions on Image Processing*, vol. 26, pp. 5176–5187, Nov. 2017.
- [22] H. Rauhut and R. Ward, "Sparse Legendre expansions via ℓ_1 minimization," *arXiv e-prints*, p. arXiv:1003.0251, Mar. 2010.
- [23] P. Abrial *et al.*, "Morphological component analysis and inpainting on the sphere: Application in physics and astrophysics," *Journal of Fourier Analysis and Applications*, vol. 13, pp. 729–748, 12 2007.
- [24] M. Elad *et al.*, "Analysis versus synthesis in signal priors," in *2006 14th European Signal Processing Conference*, 2006, pp. 1–5.
- [25] N. Cleju, M. G. Jafari, and M. D. Plumbley, "Choosing analysis or synthesis recovery for sparse reconstruction," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 869–873.
- [26] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cospase analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.
- [27] S. Boyd *et al.*, *Convex optimization*. Cambridge university press, 2004.
- [28] C.-P. Robert, *The Bayesian Choice*, 2001, vol. 340.
- [29] S. M. Feeney *et al.*, "Sparse inpainting and isotropy," *JCAP*, vol. 2014, no. 1, p. 050, Jan. 2014.
- [30] M. Pereyra, "Maximum-a-posteriori estimation with bayesian confidence regions," *SIAM Journal on Imaging Sciences*, vol. 10, no. 1, pp. 285–302, 2017.
- [31] X. Cai, M. Pereyra, and J. D. McEwen, "Uncertainty quantification for radio interferometric imaging: II. MAP estimation," *MNRAS*, vol. 480, no. 3, pp. 4170–4182, Nov. 2018.
- [32] A. Repetti, M. Pereyra, and Y. Wiaux, "Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimization," *ArXiv e-prints*, Mar. 2018.
- [33] M. A. Price, X. Cai, J. D. McEwen, T. D. Kitching, and C. G. R. Wallis, "Sparse Bayesian mass-mapping with uncertainties: hypothesis testing of structure," *submitted to MNRAS*, 2018.
- [34] M. A. Price, X. Cai, J. D. McEwen, M. Pereyra, and T. D. Kitching, "Sparse Bayesian mass mapping with uncertainties: local credible intervals," *MNRAS*, vol. 492, no. 1, pp. 394–404, 12 2019.
- [35] M. A. Price, J. D. McEwen, X. Cai, and T. D. Kitching, "Sparse Bayesian mass mapping with uncertainties: peak statistics and feature locations," *MNRAS*, vol. 489, no. 3, pp. 3236–3250, 08 2019.
- [36] R. E. Carrillo, J. D. McEwen, D. Van De Ville, J.-P. Thiran, and Y. Wiaux, "Sparsity Averaging for Compressive Imaging," *IEEE Signal Processing Letters*, vol. 20, no. 6, pp. 591–594, Jun 2013.
- [37] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging," *MNRAS*, vol. 426, pp. 1223–1234, Oct. 2012.
- [38] J. D. McEwen and Y. Wiaux, "A Novel Sampling Theorem on the Sphere," *IEEE Transactions on Signal Processing*, vol. 59, pp. 5876–5887, Dec. 2011.
- [39] E. T. Newman and R. Penrose, "Note on the bondi-metzner-sachs group," *Journal of Mathematical Physics*, vol. 7, no. 5, pp. 863–870, 1966.
- [40] J. N. Goldberg *et al.*, "Spin spherical harmonics," *Journal of Mathematical Physics*, vol. 8, no. 11, pp. 2155–2161, 1967.
- [41] E. U. Condon, E. Condon, and G. Shortley, *The theory of atomic spectra*. Cambridge University Press, 1951.
- [42] D. Marinucci *et al.*, "Spherical needlets for cosmic microwave background data analysis," vol. 383, pp. 539–545, 2008.
- [43] J. Y. H. Chan *et al.*, "Second-Generation Curvelets on the Sphere," *IEEE Transactions on Signal Processing*, vol. 65, no. 1, pp. 5–14, Jan. 2017.
- [44] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.
- [45] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [46] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The annals of Statistics*, pp. 1135–1151, 1981.
- [47] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A sure approach for digital signal/image deconvolution problems," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4616–4632, 2009.
- [48] A. F. Vidal and M. Pereyra, "Maximum likelihood estimation of regularisation parameters," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 1742–1746.
- [49] M. Pereyra *et al.*, *Maximum-a-posteriori estimation with unknown regularisation parameters*, 12 2015, pp. 230–234.
- [50] A. D. Hillery *et al.*, "Iterative wiener filters for image restoration," *IEEE Transactions on Signal Processing*, vol. 39, no. 8, pp. 1892–1899, 1991.
- [51] D. Kodi Ramanah, G. Lavaux, and B. D. Wandelt, "Wiener filtering and pure E/B decomposition of CMB maps with anisotropic correlated noise," *MNRAS*, vol. 490, no. 1, pp. 947–961, Nov. 2019.
- [52] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*. Springer, 2011, pp. 185–212.
- [53] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [54] N. Komodakis and J. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.
- [55] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. Vũ, "A forward-backward view of some primal-dual optimization methods in image recovery," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 4141–4145.
- [56] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.
- [57] F. Lanusse, J. L. Starck, A. Leonard, and S. Pires, "High resolution weak lensing mass mapping combining shear and flexion," *Astronomy & Astrophysics*, vol. 591, p. A2, June 2016.
- [58] R. Takahashi *et al.*, "Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments," *The Astrophysical Journal*, vol. 850, no. 1, p. 24, Nov 2017.