Bayesian Variational Regularisation for Dark Matter Reconstruction with Uncertainty Quantification

Matthew Alexander Price

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Space and Climate Physics, MSSL
University College London

December 15, 2021

I, Matthew Alexander Price, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work. The academic work undertaken during the STFC funding period associated with this thesis has led to the following academic articles

- **M. A. Price**, J. D. McEwen, X. Cai, T. D. Kitching, and C. G. R. Wallis, "Sparse Bayesian mass-mapping with uncertainties: hypothesis testing of structure", *Monthly Notices of the Royal Astronomical Society*, vol. 506, no. 3, pp. 3678-3690, September 2021, arXiv:1812.04014
- **M. A. Price**, X. Cai, J. D. McEwen, M. Pereyra, T. D. Kitching, and LSST Dark Energy Science Collaboration, "Sparse Bayesian mass mapping with uncertainties: local credible intervals", *Monthly Notices of the Royal Astronomical Society*, vol. 492, no. 1, pp. 394-404, Dec. 2019, arXiv:1812.04017
- **M. A. Price**, J. D. McEwen, X. Cai, T. D. Kitching, and LSST Dark Energy Science Collaboration, "Sparse Bayesian mass mapping with uncertainties: peak statistics and feature locations", *Monthly Notices of the Royal Astronomical Society*, vol. 489, no. 3, pp. 3236-3250, Dec. 2019, arXiv:1812.04018
- **M. A. Price**, J. D. McEwen, L. Pratley, and T. D. Kitching, "Sparse Bayesian mass-mapping with uncertainties: Full sky observations on the celestial sphere", *Monthly Notices of the Royal Astronomical Society*, vol. 500, no. 4, pp. 5436-5452, Jan. 2021, arXiv:2004.07855
- **M. A. Price**, L. Pratley, J. D. McEwen, "Sparse image reconstruction on the sphere: a general approach with uncertainty quantification", *submitted to IEEE Transactions on Image Processing*, 2021., arXiv:2105.04935
- **M. A. Price**, and J. D. McEwen, "Bayesian variational regularization on the ball", *submitted to IEEE Signal Processing Letters*, 2021, arXiv:2105.05518

which the content of this thesis are based upon. I have also contributed to academic articles in other areas, including:

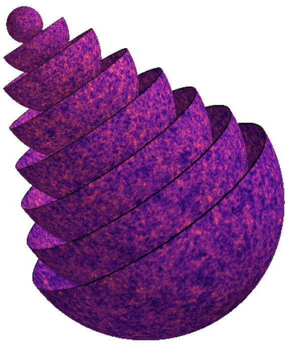
- J. D. McEwen and **M. A. Price**, "Scale-discretised ridgelet transform on the sphere", *in 27th European Signal Processing Conference (EUSIPCO)*, 2019, arXiv:1510.01595
- J. D. McEwen, C. G. R. Wallis, **M. A. Price**, and M. M. Docherty, "Machine learning assisted marginal likelihood estimation: learnt harmonic mean estimator", *submitted to Statistics and Computing*, 2021, arXiv:2111.12720
- O. J. Cobb, C. G. R. Wallis, A. N. Mavor-Parker, A. Marignier, **M. A. Price**, M. d'Avezac, and J. D. McEwen, "Efficient Generalized Spherical CNNs", *International Conference on Learning Representations (ICLR)*, 2021, arXiv:2010.11661
- C. G. R. Wallis, **M. A. Price**, J. D. McEwen, T. D. Kitching, B. Leistedt, and A. Plouviez, "Mapping dark matter on the celestial sphere with weak gravitational lensing", *Monthly Notices of the Royal Astronomical Society*, vol. 509, no. 3, pp. 4480-4497, Nov. 2021, arXiv:1703.09233

in addition to various ongoing projects.

Abstract

Despite the great wealth of cosmological knowledge accumulated since the early 20th century, the nature of dark-matter, which accounts for $\sim 85\%$ of the matter content of the universe, remains illusive. Unfortunately, though dark-matter is scientifically interesting, with implications for our fundamental understanding of the Universe, it cannot be directly observed. Instead, dark-matter may be inferred from *e.g.* the optical distortion (lensing) of distant galaxies which, at linear order, manifests as a perturbation to the apparent magnitude (convergence) and ellipticity (shearing). Ensemble observations of the shear are collected and leveraged to construct estimates of the convergence, which can directly be related to the universal dark-matter distribution. Imminent stage IV surveys are forecast to accrue an unprecedented quantity of cosmological information; a discriminative partition of which is accessible through the convergence, and is disproportionately concentrated at high angular resolutions, where the echoes of cosmological evolution under gravity are most apparent.

Capitalising on advances in probability concentration theory, this thesis merges the paradigms of Bayesian inference and optimisation to develop hybrid convergence inference techniques which are scalable, statistically principled, and operate over the Euclidean plane, celestial sphere, and 3-dimensional ball. Such techniques can quantify the plausibility of inferences at one-millionth the computational overhead of competing sampling methods. These Bayesian techniques are applied to the hotly debated Abell-520 merging cluster, concluding that observational catalogues contain insufficient information to determine the existence of dark-matter self-interactions. Further, these techniques were applied to all public lensing catalogues, recovering the then largest global dark-matter mass-map. The primary methodological contributions of this thesis depend only on posterior log-concavity, paving the way towards a, potentially revolutionary, complete hybridisation with artificial intelligence techniques. These next-generation techniques are the first to operate over the full 3-dimensional ball, laying the foundations for statistically principled universal dark-matter cartography, and the cosmological insights such advances may provide.



Simulated 3-dimensional weak lensing convergence field. Each spherical shell is at progressively increasing distance (redshift).

Impact Statement

(1) Impacts on theoretical methodology

This thesis merges the paradigms of Bayesian inference and applied mathematics (optimisation) to develop truly next-generation hybrid Bayesian inference methodology. Such hybrid techniques are broadly applicable, retain the principled statistical interpretation of Bayesian analysis, and inherit the remarkable computational efficiency of optimisation approaches. These techniques are abstracted to operate over not only standard Euclidean planes, but also curved manifolds such as the 2-sphere and the 3-dimensional ball, for which pre-existing algorithms are few and far between. In fact, this thesis is the first body of research to consider variational techniques over the 3-dimensional ball. As these methodological advances are dependent only on posterior log-concavity, they are primed for further hybridization with artificial intelligence techniques, *e.g.* learnt prior distributions, which has the potential to revolutionize computational imaging over the coming years.

(2) Impacts on the reconstruction of cosmological dark-matter

Methodological developments in this thesis are applied to the field of gravitational lensing, in which dark-matter is reconstructed: routinely over Euclidean tangent planes, occasionally over the celestial sphere, and most naturally over the 3-dimensional ball. Though a variety of methods exist to perform this inference over tangent planes, this thesis also considers the inference of dark-matter over the celestial sphere, which is lacking in the existing literature. Notably, abstraction to the 3-dimensional ball provides the methodological and computational foundations upon which the universal dark-matter distribution may be reconstructed. In each case, the approach developed in this thesis recovers the most accurate dark-matter inferences to date and, for the first time, can quantify the uncertainty in such inferences at one-millionth the computational overhead of competing sampling methods[1].

This thesis introduces a variety of novel techniques with which one may quantify the uncertainty in both local dark-matter sub-structures and global cosmological statistics. This hybrid methodology was applied to the hotly debated Abel-520 merging cluster, believed to contain spurious dark cores[2] which would have significant implications for dark-matter self-interactions. Despite reconstructing concurrent dark-matter distributions, the existence of such dark cores is statistically inconclusive

---

[1] Note that the vast majority of pre-existing approaches simply do not support uncertainty quantification to which one may compare.

[2] A dark core is a seemingly over-dense region with no optical counterpart.

at 99% confidence. The spherical methodology developed in this thesis was applied to all public weak lensing data-sets producing, what was at the time, the most comprehensive dark-matter mass-map to date. The work presented in this thesis advances the field in several key ways, priming the community to construct principled scientific statements in light of next-generation, now imminent, stage IV weak lensing surveys.

(3) Impacts on the broader scientific community

Hybrid Bayesian inference and uncertainty quantification methods developed throughout this thesis depend only on posterior log-concavity, and are thus widely applicable. For example these methods may be directly applied to magnetic resonance imaging, determining both the plausibility that apparent medical abnormalities are real, and the accuracy of their location. Moreover, as these inference methods are abstracted to the sphere and ball they are applicable to a broad range of non-Euclidean inverse problems, *e.g.* diffusion tensor imaging, computer vision, radio interferometry, geophysical modeling, and beyond.

Acknowledgements

To all those that encouraged, inspired, and tolerated me.

It is difficult to overstate my gratitude towards Professor Jason McEwen, my principal supervisor, as his mentorship throughout my doctoral studies has been exemplary. I am particularly appreciative of his advice when writing academic articles and proposals, which has greatly improved the quality of my work, and the many rigorous academic discussions we have shared. I also thank Professor Tom Kitching for providing his extensive knowledge of the field of gravitational lensing. There are, of course, many more people with which I have shared fruitful academic discussions, both at UCL and further abroad. In particular, I am indebted to: Dr. Xiaohao Cai, for introducing me to many core themes of this thesis; Dr. Luke Pratley, for his many insights and academic camaraderie; Dr. Peter Taylor, for stimulating coffee break conversations; and Associate Professor Marcelo Pereyra, Dr. Christopher Wallis, and Dr. Mayeul d'Avezac, for various collaborative efforts.

Though interactions over my doctoral years are perhaps the most pertinent, there are a great many people that have shaped the person I am today; ranging from the excellent set of supervisors I had the privilege of experiencing at Cambridge (Fitzwilliam College) to my high-school teachers, in particular Mr. Clare, Dr. Exley, and Ms. Dawson. On a broader scale, I thank the LSST-DESC collaboration, to which I have been a member these few short years, for reviewing various articles and granting me a global perspective on the scientific process. Academic endeavours are not without, often substantial, administrative overhead. In this regard I thank Philippa Elwell and Samantha Babister for their help (and patience) arranging conference and travel funds. Of course, none of this would be have been possible without the MSSL Astrophysics group, and the funding provided by STFC, for which I am grateful.

Closer to home, I am eternally grateful to my family for the support, encouragement, and edification they have provided me thus far. To my parents, who taught me the value of intellectual rigour, curiosity, and imagination, I am singularly indebted. It would be remiss of me not to take this opportunity to thank my partner Abi, for her endless support and for patiently lending an ear to topics which are, no doubt, of only passing interest to her.

Contents

List of Figures

List of Tables

Part I

Theoretical background

Chapter 1

Introduction and historical context

How did the Universe, as we know it, come to be? This fundamental question as to the nature of the cosmos has been pondered for millennia, despite which it remains as compelling as it is involuted to this day. Only within the past century have the necessary tools with which such a question may begin to be addressed been developed. In recent decades cosmologists have converged to a robust theoretical understanding of the state of the Universe as it was roughly 13.8 billion years ago. This consensus understanding has now been corroborated by substantial observational evidence, extracted, in large part, from the cosmic microwave background (CMB). Contrasting the largely homogeneous CMB to the highly structured cosmic web we observe here and now, would appear to indicate substantial cosmological evolution under gravity, which must be understood. To this end, cosmologists are effectively attempting to unravel the narrative arc of the Universe, having access only to its first and final chapter; requiring a comprehensive understanding of the individual components of the Universe, and the interactions between such components, *e.g.* matter and gravity.

> *"Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house."*
>
> Henri Poincaré (1902)

One of the most direct avenues through which cosmologists may study both gravity and (in particular dark) matter is through the cosmological phenomenon of gravitational lensing, which is the primary topic of this thesis. In this first part a holistic background is provided, setting the scene for the novel research provided in later chapters. This chapter summarizes the historical context within which gravitational lensing exists, before proceeding to a description of the current state of the field. Emphasis is placed on major theoretic and experimental strides made during the mid to late 20th century moving into the early 21st century. In tandem to the astrophysical discussion a more general discussion of inverse problems within science is then broached before presenting the context within which this thesis is set. The themes raised within this introduction are introduced so as to motivate extended mathematical background presented in chapters 2 through 4.

1.1 Early history of gravitational lensing

Newton (1704)

One may attribute the earliest reference to the concept we now know as gravitational lensing to none other than Sir Isaac Newton in Book 4, Part 1 of his renowned 'Opticks' published in 1704, which reads

> *"Do not bodies act upon light at a distance and by their action bend its ray's; and is not this action (caeteris paribus) strongest at the least distance?"*
>
> Sir Isaac Newton (1704)

This statement is somewhat surprising as, under a Newtonian interpretation of gravity, the notion of light bending within a gravitational field is somewhat confused. Nevertheless, the consensus today is that this was the first, albeit incorrect (perhaps more diplomatically incomplete), conception of gravitational lensing. Shortly following the 1919 observation of gravitational lensing, the then president of the Royal Society, J.J. Thomson, noted that Newton had "in fact suggested this very point in the first query". However, Newton never returned to this topic and so the first formal calculation of such a deflection of light cannot strictly be attributed to him.

Soldner (1804) and Cavendish (1784)

This first published calculation of light deflection by a gravitating body was written by Johann Georg Soldner a century later in 1804 (Soldner, 1804). However in 1921, a further century later documents dated to 1784 by the eminent Henry Cavendish were uncovered, within the collection of the Duke of Devonshire, explicitly stating the influence of gravity on light (Dyson, 1921) – perhaps the most relevant under the title 'On the bending of light by gravitation'. Both Soldner and Cavendish's papers address gravitational interactions of light rays passing close to massive bodies, however it is interesting to note that certain boundary conditions within Soldner's derivation were incorrect within a Newtonian setting. Fortuitously, any erroneous factors cancelled in the penultimate stages of the derivation, and so the results, to linear-order, remained unchanged. It was in this contribution that Soldner explicitly calculated the gravitational deflection of a light ray passing the limb of the Sun to be ~ 0.84 arc-seconds.

Einstein (1905-1916)

Often referred to as Albert Einstein's *annus mirabilis*, 'miracle year', 1905 marked the beginning of the theory of relativity, within which Einstein published several seminal papers. The first paper translates roughly to 'Concerning a heuristic point of view towards the emission and transformation of light' (Einstein, 1905c) in which the photoelectric effect was first postulated, winning Einstein the physics Nobel prize in 1921 after a, somewhat sceptical, Robert Millikan confirmed the theoretic predictions in 1916 (Millikan, 1916). Einstein's second paper presents a theory of Brownian motion (Einstein, 1905a) which is largely unrelated to gravitational lensing, however has found wide-spread use *e.g.* within probability theory, as discussed in later chapters. Einstein's third (Einstein, 1905d) and fourth (Einstein, 1905b) papers, the former being titled 'On the electrodynamics of Moving Bodies', presented a truly novel interpretation of the nature of gravity. Where Newton conceived of gravity as an attractive force between masses, Einstein drew on the field of differential geometry to introduce the notion of gravity as curvature of an underlying Riemannian manifold. Further to this, the revolutionary mass-energy relation was derived. Using both these novel concepts, within a relativistic framework it is entirely consistent that light, just as matter, will deflect when passing close to massive bodies. This gravitational deflection, or *lensing*, of light was explicitly stated in a review published in the *Jahrbuch der Radioactivität und Elektronik* in 1907.

Subsequently in 1911 Einstein further published a theoretical linear-order deflection value (Einstein, 1911), consistent with that previously derived by Soldner and Cavendish, though derived from a starkly different interpretation of gravity. Importantly, at this point the theory of general relativity was in its infancy and incomplete, hence Einstein had considered spatial curvature alone, neglecting temporal curvature. Of course, for any theory to be accepted it must first be confirmed through observation which, at the time, raised some concern. The problem was that the angular resolution required to observe such sub arc-second deflections around the Sun was, at the time, perceived to be unachievable. Furthermore, observations were fundamentally predicated on a total solar eclipse. Such eclipses occur roughly 2-4 times per annum, however only a ~ 50 mile diameter circular patch of the Earth's surface experiences totality, per solar eclipse. Therefore, to recover the requisite observations within a time-frame of years, observers would be required to travel to remote locations around the globe. At the time such exploration was notoriously expensive, not to mention time consuming.

Nonetheless, Erwin Finlay-Freundlich and William Wallace Campbell set out to just such a total solar eclipse located in Feodosiya, Crimea in 1914. Unfortunately, before the 1914 eclipse occurred the western world spiralled into a global conflict, now known as the first world war. All experimental equipment was confiscated and expedition members were taken into custody. The observation was not made. In hindsight, this delay perhaps benefited relativity as it allowed time for the theory to mature, and the theoretic predictions to be adjusted. By 1916 Einstein's theory of general relativity was complete (Einstein, 1916) and a new, corrected, theoretical prediction of the solar deflection angle was calculated to be ~ 1.7 arc-seconds – precisely a factor of 2 more than the Newtonian prediction.

Dyson, Eddington and Davidson (1919)

In May 1919 a Royal Astronomical Society expedition, purportedly conceived by Frank Watson Dyson, set out to perform observations of Mercury's solar transit, during a total solar eclipse. On the globe there were only two locations at which the eclipse was total: Princípe (off the west coast of Africa), and Sobral (in Brazil). Sir Arthur Eddington led the expedition to Princípe, whilst simultaneously Andrew Crommelin and Charles Davidson travelled to Sobral. After six and a half minutes of totality, and substantial post-processing and analysis of astro-graph plates, Eddington reached a deflection of 1.61 ± 0.30 arc-seconds (Dyson et al., 1920). The Sobral plates however, being out of focus, gave measurements of 0.93 arc-seconds which were subsequently accounted for in the statistical analysis.

> *"Thus the results of the expeditions to Sobral and Princípe can leave little doubt that a deflection of light takes place in the neighborhood of the Sun and that it is of the amount demanded by Einstein's general theory of relativity, as attributed to the Sun's gravitational field."*
>
> Sir Joseph John Thomson (1919)

At this juncture, physicists had both a working relativistic theory of gravity, in which light naturally deflects around massive bodies, and observational confirmation thereof. Hence, the birth of the field of gravitational lensing can arguably be attributed to this date.

Khwolson (1924), Einstein (1936), and Zwicky (1937)

Orest Khwolson, sometimes written Chwolson, in 1924 published a short note in *Astronomische Nachrichten*, astronomical notes, demonstrating that certain geometric configurations of source and lens can produce multiple images of a single source (Khvolson, 1924), an effect which we now refer to as strong gravitational lensing – an extreme example of which is an Einstein-Khwolson ring (see *e.g.* Kochanek et al., 2001). Khwolson was not alone in this realization; both Oliver Lodge (Lodge, 1919), and Eddington present similar works at that time, though all agreed the effect would not be observable. In 1936, it is reasonable to conclude that Einstein held much the same opinion, believing the lensing effect to be unobservable in practice. However, purportedly with urging from Rudi W. Mandl, Einstein published a short note in Science presenting explicit formulae for certain geometric factors relating to gravitational lensing (Einstein, 1936). Notably, Einstein included the formula for the magnification factor in addition to introducing the concept of micro-lensing, since coined by Paczynski, in which the fluctuation in variable intensity stars could be explained by intervening massive bodies. Little did Einstein know that such micro-lensing events would, a century later, be leveraged to detect far flung exoplanets (Gaudi, 2012), and upon which further great discoveries may yet depend.

Building on his then recent detection of a mysterious dark matter component in the Coma cluster (Zwicky, 1933), Fritz Zwicky published two short articles in Physical Review. In the first of these articles Zwicky argues that galaxies, or galaxy clusters, provide far more effective gravitational lenses than stars (Zwicky, 1937a). Importantly, this conclusion was met by assuming a galactic mass of $\mathcal{O}(10^2)$ times larger than what was, at the time, the expectation – a value justified by Zwicky's recent discovery of dark matter. In the second of these articles Zwicky determined that, if one considers a dark[1] component within massive clusters, the probability that nebulae acting as gravitational lenses will be found is "practically a certainty" (Zwicky, 1937b).

1.2 Modern history of gravitational lensing

Liebes, Klimov, Refsdal and Zel'dovich (1962 - 1963)

Between 1937 and the early 1960s Fritz Zwicky's vision of dark matter cartography through gravitational lensing (Zwicky, 1937a,b) remained incomplete, indeed in many regards it simply wasn't addressed. Despite the obvious reason that the world had been plunged once again into global conflict, the angular deflections predicted by gravitational lensing were simply too small to be feasibly measured. Bluntly put, the rapid progression of the theory of gravity, and the theoretical predictions such advances entail, had dramatically outpaced the development of astronomical imaging technology. In a somewhat morbid sense, the second world war led to great technological advances in many scientific disciplines; particularly in the field of radio communications — which, at the time, was inextricably linked to astronomical imaging. This, along with steady advances in computing made throughout the middle of the 20th century, placed astronomy in a far better technical state than it had been just a few short decades prior.

Gravitational lensing had something of a renaissance, leading to a string of theoretical publications, which further developed the underlying geometric framework (Klimov, 1963; Liebes Jr, 1964; Refsdal and Bondi, 1964; Zel'dovich, 1964). Sjur Refsdal further demonstrated various important characteristics exhibited by variable intensity light sources, *e.g.* quasars, when gravitationally lensed (Refsdal, 1964). Refsdal highlighted that observation of such events would provide a geometric means by which one might infer the expansion rate of the universe – postulated many years prior by Hubble (Hubble, 1929) or Friedmann (Friedmann, 1922). Of particular interest, Zel'dovich presented an analysis of the effect of gravitational lensing on cosmological scales (Zel'dovich, 1964), reinforcing the foundations upon which the study of gravitational lensing and the cosmic microwave background rely today.

---

[1] In this context dark refers to matter which cannot be observed, but for which we may observe downstream effects *e.g.* perceived gravitational attraction to visually empty regions of space.

Boyle and Smith (1969)

For modern cosmological surveys to become a reality, a sufficiently precise imaging device was required, one which ideally was not restricted to the large unwieldy dishes of radio telescopes. In the mid 1960s, Bell Labs were working on precisely such a device, though for entirely separate applications. In 1969, Willard Boyle and George E. Smith noticed that two of their primary projects, a picture phone and semi-conductor bubble memory, could be combined into what was referred to as a 'charge bubble device' – known now as a charge coupled device (CCD). Originally, an electrical impulse across the surface of a semi-conductor would trigger memory storage within said surface, however Boyle and Smith realized that such memory storage could equally be triggered *via* Einstein's photoelectric effect (Einstein, 1905c; Boyle and Smith, 1970). Due to industrial development time, CCD imaging only became widespread within astronomical imaging in the 1980s, narrowly missing out on the first detection of gravitational lensing in 1979, made on a traditional radio telescope. Both Boyle and Smith shared the physics Nobel prize in 2009 for the 'invention of an imaging semiconductor circuit - the CCD sensor'.

Walsh *et al* (1979)

In 1979 Dennis Walsh, Bob Carswell, and Ray Weymann made the first observation of a multiply imaged galactic source on the Kitt Peak National Observatory 2.1 meter radio telescope (Walsh et al., 1979). In a remarkably reserved abstract, published in Nature, the authors announced the detection of a pair of blue stellar objects that were, in fact, multiple images of the same, gravitationally lensed, quasar labelled SBS 0957+561 A/B — therefore crediting the team with the first confirmed detection of a strongly gravitationally lensed source image.

Irwin *et al* (1989)

In 1989 Mike J. Irwin and Geraint F. Lewis published a paper titled "Photometric variations in the Q2237+0305 System: First detection of a micro-lensing event" in which they present, as the title suggests, the first detection of a micro-lensing event (Irwin et al., 1989). Interestingly, they also quote an upper-bound on the lensing mass as ≤ 0.1 solar masses and illustrate that subsequent monitoring of the aforementioned quasar system could provide an elegant probe of the intervening mass distribution.

Tyson, Valdes and Wenk (1990)

Within a year of the first detection of gravitational micro-lensing, J. A. Tyson, F. Valdes and R. A. Wenk detected the first signatures of galaxy cluster lensing in two, high-velocity dispersion *i.e.* the highly dynamical, clusters A1689 and CL 1409+52 (Tyson et al., 1990). Observations of background galaxy elipticities were averaged to remove residual intrinsic effects, and a so-called ‘alignment statistic’ was computed for a grid of possible lens locations. The most likely alignment statistic was determined, from which the lensing mass centre was computed and compared against the luminosity profile. Tyson *et al*, found that the lensing distortions were roughly correlated to both the light centre and profile, leading the authors to conclude that the observations supported theories of baryonic dark matter – or at least theories with dissipatively coupled dark matter.

Kaiser and Squires (1993)

In 1993 N. Kaiser and G. Squires released, a now classic, paper in which they considered, for the first time, what researchers now refer to as mass-mapping (Kaiser and Squires, 1993): the reconstruction of the integrated mass distribution field from coherent distortions of background galaxies. In this paper, Kaiser & Squires determine analytic relations between linear-order lensing components and the integrated lensing potential, from which the lensing forward model is derived. This forward model allows one to calculate an analytic relation between the unobservable magnification and the observable distortions of distant galaxies. One may invert this relation to compute estimates of the magnification. Such a lensing inversion method is referred to as Kaiser-Squires (KS) inversion. As will be discussed in many areas of this thesis, this KS approach is sub-optimal for modern weak lensing mass-mapping (see *e.g.* Lanusse et al., 2016; Jeffrey et al., 2018; Price et al., 2021a), however at the time this was a seminal paper, and is still widely used today.

Brainerd *et al* (1996)

Subsequently, weak gravitational lensing of distant galaxies was detected in a paper by Tereasa G. Brainerd, Roger D. Blandford and Ian Smail, accepted to the Astrophysical Journal in 1996 (Brainerd et al., 1996), within which the polarization of weakly lensed images was measured. Brainerd *et al* conclude that the weak lensing on galactic scales provides “a viable and potentially powerful probe of the outer parts of normal galaxies”. Additionally, the authors discussed the possibility of corroborating measurements made using deep Hubble space telescope (HST) data archives.

Bacon, Refregier and Ellis (2000)

At the turn of the millennium David J. Bacon, Alexandre R. Refregier, and Richard S. Ellis reported a 3.4σ detection of weak gravitational lensing by large scale structure (Bacon et al., 2000), from measurements made on the William Herschel Telescope (WHT), a phenomenon now referred to as ‘cosmic shear’. Analysing this detection, the group found agreement with theoretical predictions of cluster-normalized cold dark matter (CDM) and the current consensus Λ-CDM cosmology, with a matter density parameter $\Omega_m = 0.3$. Further to this ground breaking detection, the authors discuss the future of cosmic shear observations with ground-based telescopes. It should also be noted that later that same year two other groups, spearheaded by Van Waerbeke (Van Waerbeke et al., 2000) and Wittmann (Wittman et al., 2000), also reported successful detections of cosmic shear, corroborating the conclusion drawn by Bacon *et al.*

Massey *et al* (2007)

Until now researchers had largely, though not wholly, only considered 2D projected maps of dark matter, hence it is fitting to end this overview of the historical context with the first attempt at a 3D dark matter map. In 2007 Massey *et al*, for the first time, conducted a full 3-dimensional analysis of an observational cosmic shear data-set (Massey et al., 2007b), acquired through the Hubble Space Telescope COSMOS survey. In a paper published in Nature, the group recovered a Euclidean 3-dimensional map of the dark-matter distribution over a 1 deg^2 patch of the celestial sphere, out to a redshift of $z = 1$ (Massey et al., 2007a). Despite the limited resolution of Massey's 3D reconstruction, its existence demonstrates that the cosmological community is finally approaching the vision, proposed by Fritz Zwicky seven decades earlier, of universal dark matter cartography.

1.3 Gravitational lensing in the 21^{st} century

In the wake of first detection (Bacon et al., 2000) weak gravitational lensing has rapidly advanced, both in theoretical knowledge and technical capacity, engendering a staggering number of important contributions over the past two decades. A large factor driving this expansion is the runaway growth of computing bandwidth, as famously predicted by Moore (Moore, 1965). Many developments build upon: advances in computational algorithms, probabilistic methodology, and data reduction techniques. In particular, over roughly the past decade, astrophysics has rapidly evolved, and is evolving still, into the so-called era of big data, in which scientists have far more data than ever before. In a very real sense the tables have now turned: where once astronomers were limited by insufficient data, in today's age astrophysicists are struggling to process the data into useful information. Reverberations of such a paradigm shift are particularly egregious within the field of cosmology, where large scale wide-field sky surveys are staged to gather hundreds of petabytes of raw astrophysical data. Such data must be transformed into scientifically meaningful quantities, from which cosmological information may be extracted. This process is often extremely computationally taxing, in certain cases prohibitively so *e.g.* for high resolution cosmological imagery – within which interesting, and potentially decisive, lensing information is forecast to be stored (see *e.g.* Munshi et al., 2008; Copeland et al., 2018; Taylor et al., 2018).

Weak gravitational lensing, at linear order, manifests itself into two fields; the perturbation of the apparent magnification (convergence) and the ellipticity (shearing) of distant galaxies. As is detailed in chapter 4, the convergence is not an observable quantity, however it can be mathematically related to the observable shear. Hence, lensing surveys collect observational catalogues of cosmic shear, which are subsequently leveraged to infer the convergence. Prototypical weak lensing analysis pipelines can roughly be decomposed into various stages, several of which are experimental in nature, and are thus susceptible to a variety of epistemic uncertainties which are typically difficult to account for (see *e.g.* Mandelbaum, 2017). The convergence field, from which potentially decisive cosmological information may readily be extracted (see *e.g.* Munshi et al., 2008; Peel et al., 2018), is typically inferred from cosmic shear, through theoretically motivated relations. Colloquially this inverse problem is referred to as 'mass-mapping' and is often seriously ill-posed, introducing significant uncertainty which has previously not been accounted for (Price et al., 2021a,b). Though a large proportion of the cosmological information is encoded in 2^{nd}-order statistics of the cosmic shear (see *e.g.* Alsing et al., 2016; Taylor et al., 2018), the weak lensing community has become increasingly interested in extracting higher order, non-Gaussian information from the convergence field (Takada and Jain, 2004; Munshi et al., 2012; Kayo et al., 2013; Lin, 2016; Munshi and Coles, 2017; Peel et al., 2017b). Extraction of scientifically relevant, non-Gaussian information from weak lensing convergence fundamentally requires a principled understanding of any and all uncertainties introduced during the lensing pipeline. After all, scientific statements of belief are of limited use when one does not have a principled understanding of the degree to which such beliefs are plausible – see chapter 3 for a discussion on this topic.

At the time of writing this thesis, a variety of mass-mapping techniques are routinely adopted throughout the weak lensing community, several of which are discussed in section 4.2.4. Perhaps most pervasive is that of Kaiser-Squires (Kaiser and Squires, 1993) and variants thereof, which are considered the industry standard throughout the literature. Under certain conditions the KS estimator for the convergence is equivalent to the maximum likelihood estimator which, by the Gauss-Markov theorem, is of minimum variance. However, such estimators are patently sub-optimal in realistic settings: in which noise overwhelmingly dilutes the lensing signal, the survey area is incomplete, and is of non-trivial geometry. Moreover, as the decomposition of spin signals (Newman and Penrose, 1966; Goldberg et al., 1967) over bounded manifolds is degenerate the mass-mapping problem is inherently ill-posed (see *e.g.* Bunn et al., 2003). With this in mind, direct inversion methods, *e.g.* the KS estimator, lack sufficient regularity properties, and thus often exhibit poor reconstruction fidelity. More complex mass-mapping techniques have been developed (*e.g.* VanderPlas et al., 2011; Alsing et al., 2016; Lanusse et al., 2016), which share a commonality in that they attempt to regularise this ill-posed inversion, with varying degrees of success[2].

---

[2] This topic is discussed in sections 2.1.3, 3.1.3, and 4.2.4.

Though these techniques can be shown, both practically and theoretically, to produce robust convergence estimators they have, until recently, lacked a principled means by which the uncertainty in such estimators may be quantified (Price et al., 2021a). Those that are constructed within a statistical framework typically require strong assumptions of Gaussianity, which inadvertently degrades the quality of cosmologically interesting non-Gaussian information[3], at times severely. Methods which are statistically principled and do not rely on Gaussianity, *e.g.* sampling methods, are computationally infeasible for the high angular resolutions at which the discriminatory lensing information exists. At the time of this thesis, mass-mapping is very much an open problem.

As is evident from the breadth of research around mass-mapping, and the study of inverse problems more generally, it is clear that research in this space is partitioned into (at least) three, regrettably rigid, schools of thought: probabilistic (Bayesian) inference, traditional optimisation, and artificial intelligence (machine learning). Each of these mindsets boasts a distinctly contrasting set of merits and drawbacks, despite which hybrid approaches are few and far between. Divisions such as this are glaringly sub-optimal, as scientific cross-pollination may well mitigate any individual drawbacks, particularly for technically challenging applications, such as cosmology. At its root, cosmology is a maximally statistically limited field of study, *i.e.* astrophysicists can recover observations of only a single Universe, hence cosmological inference problems benefit greatly from a Bayesian perspective, which does not fundamentally depend on the repeatability of outcomes (see *e.g.* Robert, 2001). Bayesian approaches are often reliant on sampling methods (discussed in section 3.2.1), which are computationally challenging in complex and/or high-dimensional settings. Furthermore, there is an inherent level of both complexity and ambiguity associated with the selection of the prior function, as discussed in section 3.1.3. Hence, although Bayesian approaches may appear ideal in theory, they are often restricted by prohibitive computational overhead. Conversely, one may adopt more efficient approaches, *e.g.* machine learning, which typically provide rapid solutions to highly complex (potentially non-linear) problems, with no obvious avenue through which to quantify the plausibility of such solutions[4]. Equivalently one may adopt image processing techniques, which often require minimal computational overhead, have very recently been shown to support approximate notions of plausibility, but are fundamentally limited to linear, low complexity problems.

At its core, the work presented in this thesis builds upon recent developments in the field of probability concentration theory (Pereyra, 2017), to bridge between the image processing and Bayesian schools of thought. In this way, one may perform

---

[3] The late universe is inherently non-Gaussian due to, *e.g.* baryonic interactions.

[4] It is certainly worth mentioning that research into artificial intelligence techniques which support principled uncertainty quantification is currently a growing interest within the machine learning community.

remarkably efficient Bayesian inferences which scale to high dimensions, retaining the computational efficiency of optimisation techniques, whilst gaining the ability to perform statistically principled uncertainty quantification (Price et al., 2021a,b). Though this thesis is grounded within the application domain of cosmology, the majority of the techniques developed are applicable to a large class of inverse problems, *e.g.* diffusion magnetic resonance imaging (dMRI), geophysics, and 360° computer vision. This highlights a somewhat novel aspect of the rapidly evolving field of astro-informatics: as research in the realms of astro-informatics and astro-statistics can often be abstracted, developed techniques invariably find a diverse range of tangential applications. Hence, research in the field of astro-informatics is of importance to, and has far reaching impact in, many fields of research as opposed to one, which has historically largely been the case in science.

> *"Astro-informatics includes a set of naturally-related specialities including data organization, data description, astronomical classification taxonomies, astronomical concept ontologies, data mining, machine learning, visualization, and astrostatistics"*
>
> Kirk D. Borne 2009

Overview of background chapters

The content of this thesis is configured as follows. **Chapters** 2-4 provide the information and context necessary for one to properly appreciate the later research chapters. In **chapter** 2 we introduce inverse problems both conceptually and mathematically, with an emphasis on how such problems become ill-posed and/or ill-conditioned. Subsequently, we discuss how one may solve such problems *via* iterative optimisation methods, drawing on the field of convex optimisation. Finally, we highlight aspects of proximal analysis and how they may be leveraged to solve optimisation problems over non-differentiable functions (which are increasingly popular throughout the literature). In **chapter** 3 we consider the discipline of probabilistic inference, primarily as an avenue by which one may approach inverse problems discussed in the previous section. A holistic overview of Bayesian methodology, including the consideration of posterior definition, is provided. We conclude this section with an overview of sampling methods, *maximum a posteriori* inference, credible regions, and recent advances in probability concentration theory — which underpin much of the original research discussed in later chapters. **Chapter** 4 provides a mathematically rigorous dissection of gravitational lensing (the science application domain in which this thesis is grounded). It is demonstrated mathematically how one may progress from linear-order cosmological perturbations to the, typically assumed, relations which describe the deflection of photons as they propagate through space-time. The weak gravitational lensing inverse problem is defined from these relations, and current methods adopted by the scientific community to solve said problem are contrasted — setting the scene for the research of this thesis.

Overview of doctoral research chapters

With the necessary mathematical concepts introduced, the research content of this thesis is covered in **chapters** 5-10. These chapters are largely based on a proper sub-set of 6 first-author publications, completed during the funding period of my doctoral studies. Where applicable, the relationship to specific publications, and the necessity of any additional *a priori* expertise (beyond that discussed in chapters 2-4), are clarified. Throughout these chapters the 'royal we' is adopted so as to avoid confusion. In **chapter** 5 the novel sparse Bayesian planar mass-mapping algorithm, and scalable uncertainty quantification techniques, developed in Price et al. (2021a) are presented. The efficacy of super-resolution operators is demonstrated, and the algorithm is leveraged to reconstruct observations of the Abel-520 merging cluster, which is considered by some to provide evidence of a non-zero dark-matter self-interaction cross-section. In **chapter** 6 the scalable Bayesian uncertainty quantification methodology, presented in chapter 5, is leveraged to effectively recover pixel-level Bayesian error bars, or 'local credible intervals', of recovered maps. These efficient uncertainty quantification techniques are benchmarked against state-of-the-art proximal sampling methods, so as to quantify the magnitude of potential approximation discrepancies. In **chapter** 7 the aforementioned Bayesian algorithm is leveraged to quantify the uncertainty in the location of recovered features, *e.g.* dark-matter haloes, and the aggregate peak statistic (Lin, 2016).

Having considered the mass-mapping inverse problem over Euclidean tangent planes, we extend this discussion to encompass wide fields of view, over which the lensing inverse problem is naturally spherical. **Chapter** 8 presents the work of Price et al. (2021c), in which a holistic analysis of inverse problems defined over spherical manifolds is considered in detail. Highly optimized C++ software developed during this work is leveraged to solve a diverse set of spherical inverse problems, and the uncertainty quantification techniques of chapters 5 and 6 are extended to $\mathbb{S}^2$. Penultimately, in **chapter** 9 the spherical techniques considered in chapter 8 are leveraged, so as to extend the sparse Bayesian mass-mapping algorithm of chapter 5 to the full celestial sphere. Such an extension is strongly motivated both experimentally (see *e.g.* Wallis et al., 2021) and theoretically by *Gauss's Theorema Egregium*[5]. This spherical mass-mapping algorithm is demonstrated on highly realistic N-body simulations, before being applied to generate, what was at the time, the largest, most comprehensive map of the celestial dark-matter distribution.

Finally, in **chapter** 10 the discussion is abstracted from inverse problems over spherical manifolds to the full 3-dimensional ball, *i.e.* $\mathbb{B}^3 := \mathbb{R}_{>0} \times \mathbb{S}^2$. Rotationally

---

[5] A fundamental theorem of differential geometry which states that curvature is an immutable characteristic of a manifold.

symmetric and radially distributed inverse problems, which are legion, are most naturally defined over $\mathbb{B}^3$. This chapter, for the first time, considers such problems which until recently have been fundamentally limited to *ad hoc* approximate methods, *e.g.* spherical tomography, laying principled foundations upon which the complete 3-dimensional lensing problem can be considered (see *e.g.* Massey et al., 2007a,b, for a Euclidean attempt). Next generation, non-Euclidean signal processing techniques, such as those developed throughout this thesis, may well constitute one of the major developments necessary to truly realize Fritz Zwicky's vision of universal dark-matter cartography. Moreover, the degree of plausibility in such cosmological cartography is quantifiable, which is of paramount importance for principled scientific progress *i.e.* progress that is self-consistent, unbiased, and accurately reflects any statistical limitations of observed data.

Chapter 2

Inverse problems

This chapter introduces the concept of an inverse problem, providing mathematical and conceptual discussions of both what is meant by ill-posedness of an inverse problem, and why regularisation is often not only beneficial, but required. A simple, yet highly informative example is provided in this regard. Following this we introduce the notion of convexity, a mathematical description of convex optimisation in its various incarnations, and demonstrate how gradient based algorithms can be extended to include non-differentiable (but Lipschitz continuous) functions. Finally, we introduce the concept of sparsity in the context of compressive (often 'compressed') sensing, before concluding with a mathematical discussion of naturally sparsifying dictionaries on non-Euclidean manifolds, *e.g.* the celestial sphere.

> *"All models are wrong, but some are useful."*
> George E. P. Box (1976)

2.1 Concept of an inverse problem

Suppose a pebble is dropped into a small bounded pond, which inevitably sends ripples propagating across the surface and reflecting from the boundaries. Suppose you can see that the surface is still and flat before you drop the pebble, that you know the mass of the pebble and the height it falls from before hitting the surface. Without dropping the pebble, it is trivial to predict exactly what form the ripples will take, how fast they will propagate, and what the distribution of ripples will be after a time t. This is the 'forward model'. Suppose instead you have no *a priori* knowledge of the pebble, and instead observe the pond at a time t after the pebble was dropped. Given the distribution of ripples apparent at time t what was the mass of the pebble? This is the 'inverse problem'.

2.1.1 Introduction to inverse problems

In the majority of scientific scenarios, observations of a quantity are made from which the question is asked, what underlying process led us to observe the quantity to take a certain value — and so science is, in a great many cases, attempting to solve the inverse problem. How does one actually solve problems of this type? Typically the process is to first construct a physically motivated *forward model* which accurately models the observations that we see at time t. The question is then: given a forward model, informed by our understanding of the underlying physics, which model parameters re-produce the results we observe. Canonically, such inverse problems are considered to be probabilistic in nature, resulting in the widespread adoption of frequentist or Bayesian mindsets — which both address similar problems in distinctly different ways (see *e.g.* Robert, 2001).

In the aforementioned case of the pebble and pond, observing the well-defined distribution of ripples on a pond after an extended time, t, theoretically contains sufficient information to precisely predict where the pebble was dropped, and even the energy with which the pebble hit the surface of the water. Provided we construct an accurate[1] *forward model* which describes the dynamical system of ripples on the surface of a pond, and provided (somewhat idealistically) that our measurement device is perfect, we can fully deduce the position and energy of the pebble's impact. Therefore, should the *inverse problem* be formulated to infer the position and energy 'parameters' then the problem is determined to be *well-posed* — which is to say, we have enough information to solve this problem precisely and uniquely (*i.e.* find a unique solution to the initial position).

Suppose instead the *inverse problem* is attempting to infer the mass of the pebble dropped. Once the pebble hits the surface this particular information is lost — a large pebble dropped from lower will produce the same energy as a small pebble dropped from higher, or in other words, the solution space is degenerate. Therefore this particular inverse problem is *ill-posed* — which is to say, we do not have enough information to solve this problem uniquely, but we can reasonably recover bounds (upper and lower limits) on what the mass could be.

It is perhaps apparent that there are, at the least, two further factors which must be considered. In practice observations of said ripples are imperfect, and thus in realistic scenarios the data with which one attempts to infer the model parameters (*e.g.* the mass, position *etc.*) has an inherent degree of uncertainty. Therefore, even given a perfect forward model one can, for instance, determine only a likely region over which the pebble was initially dropped rather than a precise point. Throughout we have implicitly assumed a pebble was dropped which, though correct by design, is by no means a certainty. In this sense such problems are inherently uncertain, and inferences as to the characteristic of model parameters must be carefully considered. This heuristic introduction is surprisingly informative. However, practical application of these ideas requires a formal understanding of the underlying mathematical framework of inverse problems — a comprehensive general review of which can be found in various articles, a good example being Tarantola (2005).

---

[1] In practice, forward-models need only be sufficiently accurate.

2.1.2 Ill-posedness and ill-conditioning

Consider the canonical case in which one collects observations $y \in \mathcal{H}_1$ (*i.e.* the distribution of ripples at time t) which can be related to the parameters of interest $x \in \mathcal{H}_2$ (*i.e.* the characteristics of the dropped pebble) by a well defined forward model $F : \mathcal{D}(F) \subseteq \mathcal{H}_2 \mapsto \mathcal{H}_1$, where $\mathcal{D}(\cdot)$ denotes the domain over which a function is defined and $\mathcal{H}_{1,2}$ are standard Hilbert spaces with associated norms $\|\cdot\|$ and inner products $\langle \cdot, \cdot \rangle$. Further define the range of the forward model to be $\mathcal{R}(\mathcal{H}_1)$. Mathematically an inverse problem is said to be *ill-posed* in the sense of Hadamard (1902) if it fails to satisfies any of the following three conditions:

(i) A solution exists.

(ii) The solution is unique.

(iii) The solution is a continuous function of the input data.

In this thesis we will consider bounded linear operators $F \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$, where $\mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$ is the set of linear operators between Hilbert spaces $\mathcal{H}_{1,2}$, for which condition (i) requires that the forward model is surjective $\mathcal{R}(F) = \mathcal{H}_1$ (ii) requires that F is injective *i.e.* the nullity $\mathcal{N}(F) = \{\mathbf{0}\}$ (iii) requires that the inverse of F is bounded. Throughout this thesis, we refer to ill-posedness implicitly in the sense of Hadamard, though there are other complimentary definitions throughout the literature.

Alternatively, one can consider an inverse problem to be ill-posed in the sense of Nashed if the set of observations and data is open, *i.e.* the range of the forward operator is not a closed subset of $\mathcal{H}_2$ (Schuster et al., 2012), or ill-posed in the sense of Tikhonov over set M if the solution does not converge to the true solution in the limit of the approximation order going to zero, *i.e.* the solution does not exhibit conditional stability (Kabanikhin, 2008). A concise example for which F is ill-posed is the differential operator $F = \partial_x$. Consider $f(x), f(x+\delta) \in \mathbb{C}([0,1])$, the error between datum is given by $\|f(x) - f(x+\delta)\| \leq \delta$ and so is bounded and may be small but the error on the differential $\|\partial_x f(x) - \partial_x(f(x+\delta)\|$ is unbounded, may be large and is not necessarily a continuous function of the data. As such, the differential operator contradicts condition (iii) of Hadamard, and is therefore ill-posed.

Another key concept worth discussing is the notion of conditioning, which simply put considers by how much a small variation in the input of a function relates to a variation in the output. Under translation $y \mapsto y + \Delta y$ such that $F(y) \mapsto F(y + \Delta y) = \Delta F(y)$ the relative condition number is given by

$$\lim_{\epsilon \to 0} \sup_{\|\Delta y\| \leq \epsilon} \frac{\|\Delta F(y)\| / \|F(y)\|}{\|\Delta y\| / \|y\|}. \tag{2.1}$$

If the relative condition number is large, a small perturbation Δy to datum y (*i.e.*

the presence of measurement error) can result in exceedingly large perturbations in the solution space. In this sense the operator F is considered to be ill-conditioned (numerically unstable), conversely small relative condition number corresponds to well-posed operators, F. Note that not all ill-conditioned problems need necessarily be ill-posed and *vice versa* though this is often the case in practice.

2.1.3 Regularisation

Often the inverse problem in question is ill-posed and/or ill-conditioned, motivating one to inject prior knowledge as to the nature of the problem so as to stabilize the inversion process — or to *regularise* the inverse problem. Such stabilizing terms are colloquially referred to as 'regularisation functionals' in the signal processing community, though these are somewhat analogous to the prior function considered within the Bayesian community (see *e.g.* Robert, 2001; Sivia and Skilling, 2006). In later chapters these terms are used somewhat interchangeably, despite originating from distinctly contrasting mind-sets. One should notice that the way in which regularisation is considered in this thesis is application oriented, and external to this thesis regularisation is a far more general mathematical, and in many ways philosophical, concept. In broad strokes: regularisation of an ill-posed and/or ill-conditioned inverse aims to infuse the problem with superior regularity properties, *e.g.* the regularised problem may be characterized by a lower condition number (see equation 2.1), or the solution space may be restricted to a small subset $\mathcal{H}_1' \subseteq \mathcal{H}_1$ to encourage convergence, avoid discontinuities, or break redundant degeneracies of the solution space. This thesis primarily considers regularisation functions which restrict or re-weight the solution space $\mathcal{H}_1$, so as to bias towards solutions which exhibit a known, or at least *a priori* assumed, structure of x. Specifically we leverage regularisation functions which promote sparsity (see *e.g.* Donoho, 2006), which will be a recurring theme throughout.

Returning to the case of the pebble and the pond: a basic form of prior knowledge could be that the pebble mass is necessarily positive. Further, one may sample the pebbles surrounding the pond, and find them to be on average of mass m, with some approximate distribution about this value (*e.g.* Gaussian distributed). In such a case, it would be sensible to leverage this prior knowledge to re-weight the degenerate solution space, biasing towards pebbles of masses similar to those surrounding the pond, thus regularizing the inverse problem. Furthermore, given that the pebble was dropped by a person standing on the ground, it is highly unlikely the pebble was dropped from much more than three metres or less than a metre. Hence, one may include additional restrictions on the feasible set of solutions, further reducing the degeneracy of the solution space. With these additional constraints on the pebble mass, the solution space is now significantly less degenerate, increasing the regularity of the inverse problem.

Sparsity

A form of regularisation increasingly often adopted throughout the literature is that of a sparsity promoting function. Conceptually, sparsity promoting functions bias, or for certain implementations restrict, the space of possible solutions towards solutions which can be expressed by the fewest coefficients in a given dictionary[2] (Elad, 2010; Gkioulekas, 2010). A dictionary in which the solution is *a priori* assumed sparse, is referred to as a sparsifying dictionary. In this sense, leveraging sparsity to increase regularity can be thought of as a mathematical implementation of the philosophical notion of *Occam's Razor*, or equivalently the *principle of parsimony* (Jefferys and Berger, 1992).

In practice, dictionaries localized entirely in the temporal domain, *e.g.* the Fourier basis, are often ineffective – as they do not effectively capture sharp, sudden features (Graps, 1995; Valens, 1999). Conversely, dictionaries localized solely in the real domain, *e.g.* the Dirac basis, are often ineffective as they capture only such sharp, sudden features. Throughout this thesis we will adopt wavelet dictionaries, see *e.g.* Figure 2.1, which are, by construction, highly localized in both the temporal and real domains (Lee and Yamamoto, 1994; Meyer, 1995; McEwen et al., 2015a). Such wavelets are widely effective, as a large class of physical signals are sparse when projected onto their atoms (see *e.g.* Daubechies, 1988, 1992; Mallat, 2008; Baldi et al., 2009; Starck et al., 2010; Carrillo et al., 2012; Leistedt et al., 2013; Chan et al., 2017; McEwen and Price, 2019).

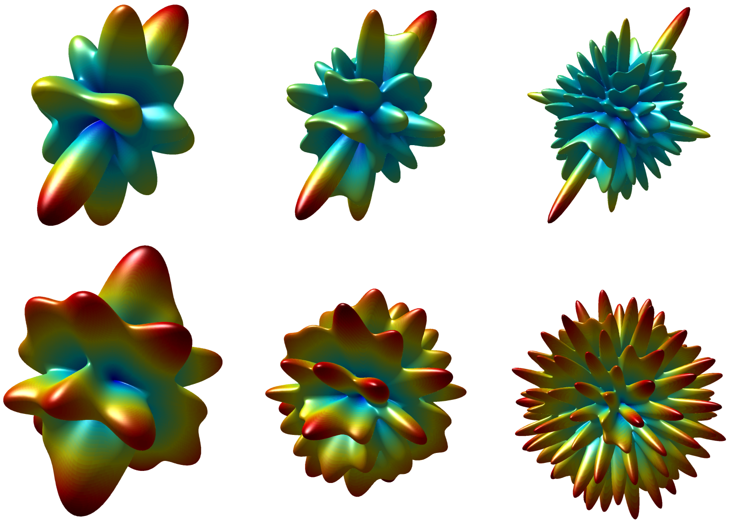


Figure 2.1: Top & Bottom: Spherical ridgelet and wavelet dictionary coefficients respectively of a diffusion tensor imaging (DTI) signal at scales $j \in [3, 5]$. Notice the frequency and pixel space localization of both dictionaries, and the relative degree to which the DTI signal is sparse when projected into each dictionary (McEwen and Price, 2019).

---

[2] A dictionary is the term attributed to a spanning set of basis functions which need not necessarily be orthogonal to one another.

2.1.4 Example ill-posed inverse problem

At this juncture it is informative to consider a simple example. One of the most simple imaging inverse problems for which regularisation is necessary is a noisy deconvolution problem. This example closely follows an example by Farrens (2018).

Suppose one captures an image $y \in \mathcal{H}_2 = \mathbb{R}^N$ with $N \in \mathbb{Z}_{>0}$ pixels, using a camera whose lens introduces a point spread function $h \in \mathbb{R}^N$. In theory our image is related to the true image $x \in \mathcal{H}_1 = \mathbb{R}^N$, where for simplicity we assume that no pixels are missing or faulty, through the forward model given by the convolution $\circledast$ equation

$$y = x \circledast B = \langle x, \mathcal{T} h \rangle_{\mathbb{R}^N} = \mathcal{F}^{-1}\Big[\mathcal{F}[x] \cdot \mathcal{F}[h]\Big], \tag{2.2}$$

where $\mathcal{F}[\cdot]$ represents the Fourier transform, $\mathcal{T}$ denotes the straightforward Euclidean translation $x \mapsto x - \tau$, and the entire equation is given by the convolution theorem on Euclidean space. Provided the observed image y is collected perfectly, the problem is well-posed in the sense of Hadamard and the true image x can be safely computed by naively inverting the forward model such that

$$\tilde{x} = \mathcal{F}^{-1}\Big[\frac{\mathcal{F}[y]}{\mathcal{F}[h]}\Big] = x. \tag{2.3}$$

Consider a more realistic situation in which, perhaps due to statistical variability in the number of photons observed per camera pixel during the shutter exposure, the observed image y is instead contaminated with low-level Gaussian noise $n \sim \mathcal{N}(0, \sigma) \in \mathbb{R}^N$ with $\sigma \ll \max(x)$, such that the measurement equation (the combination of the forward model and any factors external to the forward model *e.g.* the presence of noise) is given by

$$y = x \circledast h + n \Rightarrow \tilde{x} = \mathcal{F}^{-1}\Big[\frac{\mathcal{F}[y] + \mathcal{F}[n]}{\mathcal{F}[h]}\Big], \tag{2.4}$$

where the second equation is the modified, naive inversion. This inverse problem is now ill-posed in the sense of Hadamard, and equally importantly is severely ill-conditioned — which is to say the presence of very low-level fluctuations (noise) in the observations results in a complete scrambling of the estimate of x. Hence, direct inversion approaches are typically highly ineffective. Images for each stage of this example are presented in Figure 2.2.

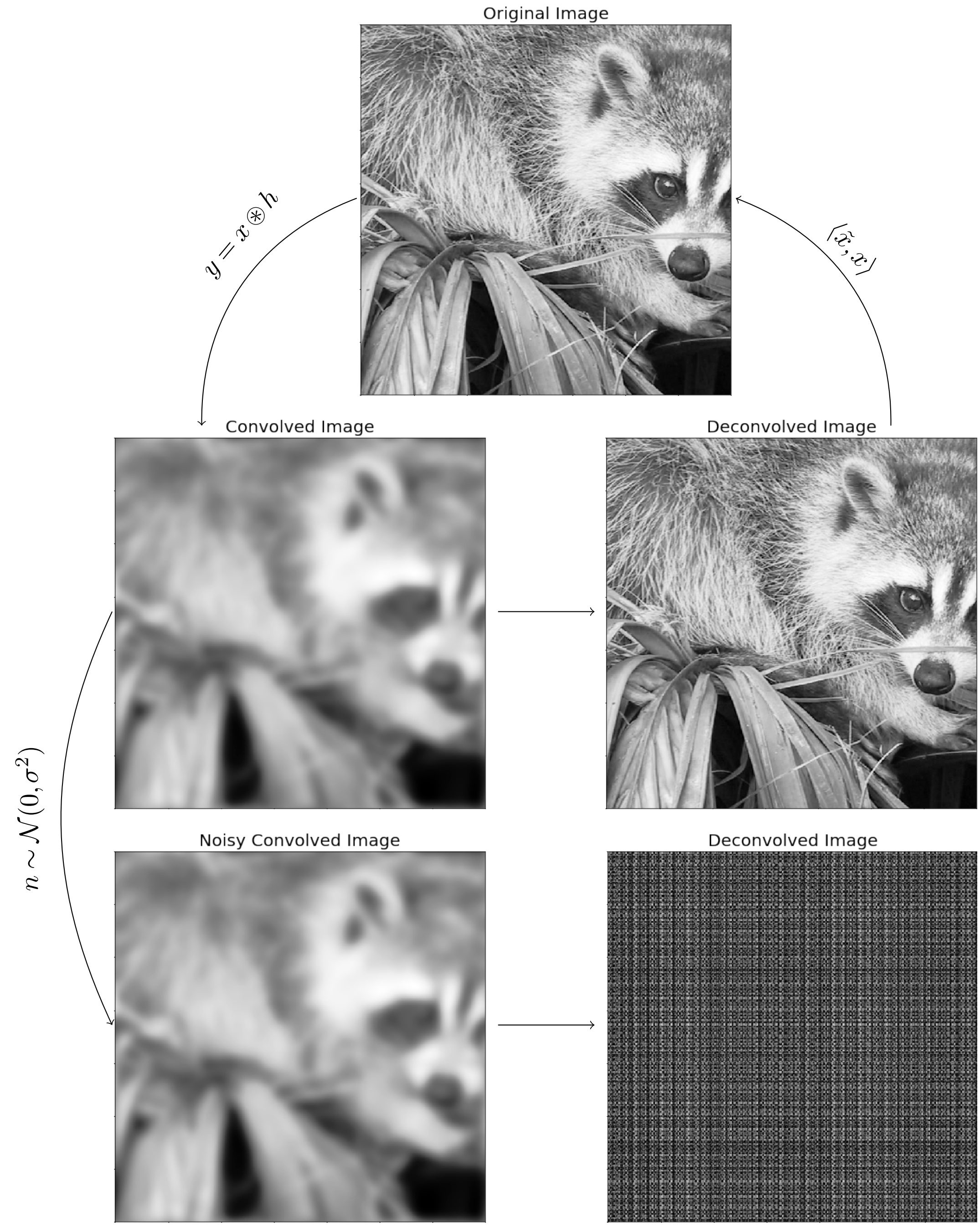


Figure 2.2: Considers both the clean and noisy naive deconvolution of an image $y \in \mathcal{H}_2 = \mathbb{R}^N$ for $N \in \mathbb{Z}_{>0}$ modeled to have been captured under a blurring *via* convolution with a point spread function h, *i.e.* $y = x \circledast h$ where $x \in \mathcal{H}_1 = \mathbb{R}^N$ is the true image. Notice that both the noisy and clean convolved images are indistinguishable to the human eye, whereas the naively de-convolved images are highly dissimilar. Note that we denote the mean squared error by the shorthand $\langle \tilde{x}, x \rangle$.

2.2 Convex optimisation

Drawing again on the pebble and pond analogy; suppose one can quantify one's *a priori* assumptions as to the nature of the pebble into a regularisation function $f(x)$, and model any aleatoric uncertainty by a data-fidelity function $g(x)$. Then the problem becomes location of solutions which extremise some combination of these functions, typically through iterative optimisation techniques. This section provides a mathematical and conceptual overview of (convex) optimisation, including details of convex tools for dealing with non-differentiable terms such as proximal calculus. Additionally a discussion of wavelets, the concept of sparsity and the functional mechanics of a variety of proximal operators is discussed.

2.2.1 Affinity, linearity, and convexity

Functions

It is informative to first lay down some definitions which will clarify further statements. A *linear function* between Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ with cardinality $|\mathcal{H}_1| = m$ and $|\mathcal{H}_2| = n$ is a function $f : \mathcal{H}_1 \mapsto \mathcal{H}_2$ such that for $f(x \in \mathcal{H}_1) = Ax \in \mathcal{H}_2$ for rank-2 tensor $A \in \mathcal{H}_2 \times \mathcal{H}_1$. An *affine function* is defined as above but that $f(x) = Ax + t$ where $t \in \mathcal{H}_2$ is a translation vector. Therefore an affine function is similar to a linear function, but conserves the relative distance between points (Boyd, 2004).

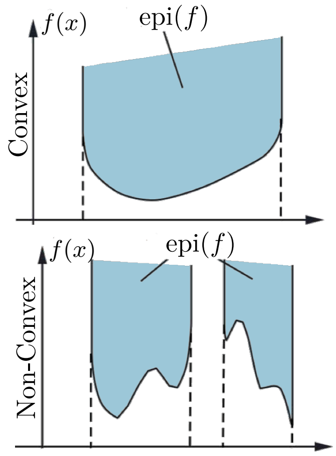


Figure 2.3: Illustration of the epigraph of convex and non-convex functions, thus describing both convex sets and functions. The dashed line corresponds to $\mathcal{I} \subseteq \mathcal{H}$, over which we are considering the convexity of $f(x)$. In terms of sets: as the line segment joining any two points is contained by the blue region, the top blue region is a convex set (adapted from MIT, 2012)

A *convex function* is a function which, between any distinct points $x_1, x_2 \in \mathcal{H}$, lies below the secant line connecting points x_1 and x_2. Formally, a function $f(x) : \mathcal{H} \mapsto \mathbb{R}$ is convex over the interval $\mathcal{I} \subseteq \mathcal{H}$ if *Jensen's inequality* (Jensen, 1906),

$$\underbrace{f(ax_1 + (1-a)x_2)}_{\text{evaluation at point}} \leq \underbrace{af(x_1) + (1-a)f(x_2)}_{\text{line segment}}, \quad (2.5)$$

$\forall\, a \in [0,1], x_1, x_2 \in \mathcal{I} \subseteq \mathcal{H}$ is satisfied. If the equality of equation 2.5 is prohibited the function is said to be strictly convex. Equivalently, this can be explained, perhaps more satisfyingly, by considering the epigraph of a function, *i.e.* the set of points above $f(x)$ contained within $\mathcal{I}$, as in Figure 2.3. The *epigraph* of a function $f(x) : \mathcal{I} \subseteq \mathcal{H} \mapsto \mathbb{R}$ is given by

$$\text{epi}(f) = \{\, x, r \,|\, x \in \mathcal{I},\, r \in \mathbb{R},\, f(x) \leq r \}. \quad (2.6)$$

A function f is convex if, and only if, epi(f) is a convex set. The 'only if' can be seen by first supposing that f is convex, and considering epigraph elements $x_i, y_i \in \text{epi}(f)$. For any $z_i \in [0,1]$ such that $\sum_i z_i = 1$ we can write

$$\begin{aligned} \Rightarrow (x,y) &= z_i \sum_i (x_i, y_i) = (\sum_i z_i x_i, \sum_i z_i y_i), \\ \Rightarrow y &= \sum_i z_i y_i \leq \sum_i z_i f(x_i) \leq f\Big(\sum_i z_i x_i\Big) = f(x), \end{aligned} \tag{2.7}$$

and thus the epigraph of a convex function is a convex set, as laid out above. Equally this logic can be inverted to prove the 'if', showing that if epi(f) is a convex set then necessarily $f(x)$ belongs to the space of convex functions. Further to this, a convex function is *proper* if and only if epi$(f) \neq \emptyset$ and does not contain a line parallel to r. Some common examples of convex functions include: all affine functions, all norms, and positive semi-definite quadratic functions. Importantly, by definition a convex function has a point $x_{\text{min}} \in \mathcal{I}$ which is a global minimum, and so optimisation algorithms can be applied effectively due to the absence of multimodality.

Sets

Consider now sets rather than functions. An *affine set* $\mathcal{A} \subseteq \mathbb{R}^n$ with cardinality $n \in \mathbb{Z}_{>0}$ is defined such that the straight line through any two distinct points $a, b \in \mathcal{A}$ is contained within $\mathcal{A}$, *i.e.* $\forall a, b \in \mathcal{A}\, a + b = c \in \mathcal{A}$. In general this is not limited to only two points *i.e.* $\forall a_1, a_2, ..., a_i \in \mathcal{A}$ for $i \in \mathbb{Z}_{\leq n}$ we have $\sum_{j=1}^{i} a_j = c \in \mathcal{A}$. With this more general approach an affine set contains all affine combinations of its constituent components. Further define the *affine hull* of $\mathcal{A}$ to be the smallest affine set which contains $\mathcal{A}$.

A *convex set* $\mathcal{C} \subseteq \mathbb{R}^n$ is a set such that the line segment connecting any two points within the set is also contained within the set. Clearly, all affine sets are necessarily convex though the opposite is not true. Again, we can generalize to n-points within $\mathcal{C}$ and state that a set is convex if and only if it contains all convex combinations of its components (Boyd, 2004). The *convex hull* is the set of all convex combinations of points within a set $\mathcal{C}'$, and is equivalent to the smallest convex set which contains $\mathcal{C}$.

2.2.2 The primal problem

The general form of a constrained (regularised) optimisation problem is given by,

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \; i = 1, \ldots, m \\ & h_i(x) = 0, \; i = 1, \ldots, p, \end{aligned} \tag{2.8}$$

where the *optimisation variable* $x \in \mathcal{H}$ is the variable we are optimizing with respect to, $f_0(x) : \mathcal{H} \mapsto \mathbb{R}$ is the *objective function* (sometimes cost or loss function) which one minimizes such that constraints $f_i(x)$ and $h_i(x)$ are met — these constraints are in effect regularizing the optimisation. This form of the optimisation is typically referred to as either the constrained optimisation or the *primal problem*. As $f_i(x) : \mathcal{H} \mapsto \mathbb{R}$ is an inequality, it is commonly referred to as an *inequality constraint*, similarly $h_i(x) : \mathcal{H} \mapsto \mathbb{R}$ is considered an *equality constraint*.

If there are no constraints $h_i(x)$ or $f_i(x)$ the domain of the problem is the domain of $f_0(x)$. If there are constraints, the domain of the problem is typically restricted and can be given by the intersection of the domains of all functions involved

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathcal{D}(f_i) \cap \bigcap_{i=1}^{p} \mathcal{D}(h_i). \tag{2.9}$$

A highly interesting and widely applicable subset of optimisation problems, is the set of *convex optimisation* problems (Nesterov, 2013), where now $f_i(x) \in \mathcal{C}^n$ belong to the subspace of convex functions $\mathcal{C}^n$, and $h_i(x) \in \mathcal{A}$ is necessarily affine. Further to this, consider the case in which one restricts both the objective function and any inequality constraints to be convex. Under such considerations the domain $\mathcal{D}_{\mathcal{C}}$ of the solution space is given by

$$\mathcal{D}_{\mathcal{C}} = \bigcap_{i=0}^{m} \mathcal{D}(f_i), \tag{2.10}$$

demonstrating that the feasible set is also convex. Hence, by definition, the problem has a unique global minimum which belongs to the set. It is informative to note that, if instead the functions $f_i(x)$ are concave and we maximize rather than minimize, then this too is a convex optimisation problem (which is discussed in section 3.2.2). A typical problem in the constrained setting, *i.e.* with constraints, could be given by,

$$x^{\text{optimal}} = \underset{x \in \mathcal{H}}{\operatorname{argmin}} \Big[f(x) \Big] \quad s.t. \quad h(x) \leq \epsilon, \tag{2.11}$$

where $\epsilon \in \mathbb{R}_{>0}$ is contour defining a level-set of h. Such level-sets are typically projections onto balls of the data likelihood, thus enforcing hard constraints on solution data fidelity.

2.2.3 Duality and the Lagrangian

One can instead consider a weighted linear combination of the initial objective function $f_0(x)$ and all constraints $f_i(x)$ and $h_i(x)$, with coefficients λ_i and μ_i respectively. Formally this is equivalent to defining a *Lagrangian* $\mathcal{L} : \mathcal{H} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ such that,

$$\mathcal{L}(x,\lambda,\mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x), \tag{2.12}$$

where the domain of the optimisation is now given by $\mathcal{D}(\mathcal{L}) = \mathcal{H} \times \mathbb{R}^m \times \mathbb{R}^p$. Analogous to the standard method of Lagrangian multipliers, the variables λ_i and μ_i represent Lagrangian variables and are often referred to as *dual variables* (Boyd, 2004; Nesterov, 2013).

From the Lagrangian $\mathcal{L}$ we can now define the *Lagrangian dual function* $g(\lambda,\mu)$ as (Kuhn and Tucker, 1951),

$$g(\lambda,\mu) = \inf_{x\in\mathcal{D}} \Big[\mathcal{L}(x,\lambda,\mu) \Big] = \inf_{x\in\mathcal{D}} \Big[f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \Big], \tag{2.13}$$

where the *infimum* is the greatest lower bound on the $\mathcal{L}$. Note that the dual function g is necessarily concave as the arguments of the infimum are affine functions. Thus, even if the original problem is non-convex, the dual problem can be solved uniquely for optimal dual variables λ^{opt} and μ^{opt}. Interestingly, the solutions of the Lagrangian dual problem provide a lower-bound on the optimal solution p^{opt} of the primal problem. To see this, consider a feasible point $x \in \mathcal{D}$, then for $\lambda_i, \mu_i \in \mathbb{R}_{\geq 0} \,\forall i$ and $(\lambda,\mu) \in \mathcal{D}(g)$, *i.e.* such that λ and μ are *dual feasible*, we have

$$\begin{aligned}
&\overbrace{\sum_{i=1}^{m} \lambda_i f_i(x)}^{\text{non-positive}} + \overbrace{\sum_{i=1}^{p} \mu_i h_i(x)}^{=0} \leq 0 \\
\Rightarrow\ & \mathcal{L}(x,\lambda,\mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \leq f_0(x) \\
\Rightarrow\ & g(\lambda,\mu) = \inf_{x\in\mathcal{D}} \Big[\mathcal{L}(x,\lambda,\mu) \Big] \leq \mathcal{L}(x,\lambda,\mu) \leq f_0(x).
\end{aligned} \tag{2.14}$$

Hence, the optimal value of $g(\lambda,\mu)$ provides a natural lower bound on the optimal value of $f_0(x)$ (Boyd, 2004). The real question is then: given that solutions of the dual problem for g provide lower bounds on the solutions of the original problem for f, what is the optimal lower bound which can be constructed? This in itself forms a maximization problem over $g(\lambda,\mu)$ such that we wish to

$$\text{maximize} \quad g(\lambda,\mu) \quad s.t. \quad \lambda,\mu \in \mathbb{R}_{\geq 0}, \tag{2.15}$$

which is often referred to as the *Lagrangian dual problem* and is a convex problem. The optimal solution g^{opt} is the greatest lower bound on the optimal value of the primal problem p^{opt}, and as such the difference $p^{\text{opt}} - g^{\text{opt}} \leq 0$ can be constructed which is known as the *(optimal) duality gap*. If the optimal duality gap is zero, then the lower bound has *strong duality*, which is to say the optimal solution of both the primal and dual problems is the same — or colloquially the solution is *primal dual optimal*.

Often, but by no means always, if the primal problem is convex then the primal dual optimal solution will exhibit strong duality. In fact, this property is not solely attributable to convexity, instead strong duality is attributed based on the constraint qualifications, *e.g. Slater's theorem*, which are beyond the scope of this thesis, yet worth noting. Nonetheless, throughout this thesis we consider cases for which duality is at least close to strong, and discuss any dissimilarities incurred by deviations from strong duality. The Lagrangian dual problem is often referred to, somewhat confusingly, as the unconstrained optimisation problem; constructed by locating the infimum (in this case the minimizer) of the Lagrangian, such that the problem is written as

$$x^{\text{optimal}} = \underset{x \in \mathcal{D}}{\operatorname{argmin}} \Big[\mathcal{L}(x, \lambda) \Big], \tag{2.16}$$

where we have simplified to the setting in which only a single inequality constraint is considered. Colloquially the weighting factor λ, which is a dual variable, is referred to as the *regularisation parameter*, optimal selection of which is a subject of current research in this area and is by no means a solved problem. The practical difference between the primal and dual problems is subtle, and may seem esoteric and/or superficial, but it is certainly worth considering. Within the constrained setting (the primal problem) the domain $\mathcal{D}_p$ of the solution space is strictly restricted to the intersection of the constraints and the objective function, whereas within the unconstrained problem (the dual problem), the domain $\mathcal{D}_d$ is open but a scalable bias factor λ penalizes solutions which stray far from the expected 'center' of $\mathcal{D}_d$. In this sense, the unconstrained problem supports a more reasonable, probabilistic approach to constraining the solution space – this links with the discussion of section 3.1.3.

Consequently, within the image-processing community the primal problem is typically considered, as the restrictions imposed on the solution space often result in greater reconstruction fidelity, rate of convergence, and do not require selection or computation of dual variables. Conversely the unconstrained setting supports a principled Bayesian interpretation, and thus can be shown to support uncertainty quantification (see section 3.2.2). As such, the dual problem is often chosen for scientific applications, in which one typically places more weight on uncertainty quantification than reconstruction fidelity.

2.2.4 Moreau-Yosida envelopes and proximal calculus

Having discussed how one may formulate inverse problems as optimisation problems, how does one actually compute optimal solutions to such problems? Generally, for differentiable $f_i(x)$ and $h_i(x)$ there exists a comprehensive set of algorithms which can be leveraged to locate x^{opt}, typically leveraging 1^{st}-order gradient information through differential operators to converge to extremal solutions. Suppose instead at least one of $f_i(x), h_i(x)$ does not belong to the space of continuously differentiable functions — for instance if one adopts a ℓ_1-norm sparsity regularisation functional $f_i(\cdot) = \|\cdot\|_1$, which is increasingly common throughout the literature (Combettes and Pesquet, 2011). One then requires mathematical tools with which to access analogous 1^{st}-order information of non-differentiable functions, which we discuss in this section.

First let us define a *lower semicontinuous* (l.s.c.) function to be a function $g : \mathcal{H} \mapsto \mathbb{R}$ such that $\mathcal{H}$ is a closed subset of $\mathbb{R}^n$. Additionally the gradient of g, denoted ∇g, is *Lipschitz continuous* with *Lipschitz constant* $\beta_{\text{Lip}} \in \mathbb{R}_{\geq 0}$ if

$$\|\nabla g(x) - \nabla g(x')\| \leq \beta_{\text{Lip}} \|x - x'\|, \qquad \forall (x, x') \in \mathcal{H} \times \mathcal{H}. \tag{2.17}$$

Typically, Lipschitz continuity is a stronger constraint than 'just continuous' but a weaker constraint than continuously differentiable.

With the concept of lower semicontinuity introduced, let us define a convex function $f(x) \in \mathcal{C}$ for $x \in \mathcal{H}$, and the *Moreau-Yosida* regularised approximation of f, the *Moreau-Yosida envelope*, given by the function $f_\mu(x)$ such that (Moreau, 1962)

$$f_\mu(x) = \inf_{x' \in \mathcal{H}} \Big[\underbrace{f(x') + \frac{1}{2\mu} \|x - x'\|_2^2}_{\text{Jointly Convex}} \Big], \tag{2.18}$$

where the bracketed term is the Lagrangian dual $\mathcal{L}(x, x')$ of a constrained minimization of f. The Moreau-Yosida envelope is given as an *infimal convolution*[3] of the non-differentiable function $f(x)$ and a smooth function. As the Lagrangian $\mathcal{L}$ is convex and the epigraph of the infimum is a projection of a convex set, then the Moreau-Yosida envelope is necessarily a convex function (Yang, 2015).

Consider now the primal problem for the Moreau-Yosida regularised minimiza-

---

[3] Convolution with a function over the infimum operator, corresponding to addition under the infimum.

tion,

$$\begin{aligned} f_\mu(x) &= \frac{1}{2\mu}\|x\|^2 - \frac{1}{\mu}\sup_{x'\in\mathcal{H}}\Big[x^T x' - \mu f(x') - \frac{1}{2}\|x'\|^2\Big] \\ &= \frac{1}{2\mu}\|x\|^2 - \frac{1}{\mu}\Big[\mu f(x') + \frac{1}{2}\|x'\|^2\Big]^* \\ \nabla f_\mu(x) &= \frac{x}{\mu} - \frac{1}{\mu}\operatorname*{argmax}_{x'\in\mathcal{H}}\Big[x^T x' - \mu f(x') - \frac{1}{2}\|x'\|^2\Big], \end{aligned} \tag{2.19}$$

where * represents the complex conjugation. Now the *proximity operator*, often denoted $\text{prox}(\cdot)$, is given by the dual problem over x' on the RHS. Noting that the proximal operator is generally defined to be

$$\text{prox}_{\mu f}(x) = \operatorname*{argmin}_{x'\in\mathcal{H}}\Big[\frac{1}{2\mu}\|x - x'\|^2 + f(x')\Big], \tag{2.20}$$

we can rearrange the final line of equation 2.19 to relate the proximal operator and the gradient of the Moreau-Yosida envelope by

$$\Rightarrow \text{prox}_{\mu f}(x) = x - \mu\nabla f_\mu(x). \tag{2.21}$$

This proximity operator is therefore solving for the gradient of the smooth Moreau-Yosida envelope which is a differentiable approximation to the gradient of the enveloped function $f(x)$, which need not be differentiable, only Lipschitz continuous with Lipschitz constant $\beta_{\text{Lip}} = \mu^{-1}$ (Moreau, 1962; Combettes and Pesquet, 2011)

Finally, consider a minimization problem of the general form $\min f(x)$ *s.t.* $x \in \mathcal{H}$ for solution domain $\mathcal{D}$ as previously defined. Suppose $f(x) \in \mathcal{C}$ has domain $\mathcal{D}(f) \neq \emptyset$ then a solution x^{opt} exists, is unique, and minimizes $f(x)$ over $\mathcal{D}(f)$ if, and only if, it also minimizes the Moreau-Yosida envelope,

$$\begin{aligned} \inf_x f_\mu(x) &= \inf_{x\in\mathcal{H}}\inf_{x'\in\mathcal{H}}\Big[f(x') + \frac{1}{2\mu}\|x - x'\|^2\Big], \\ &= \inf_{x'\in\mathcal{H}}\inf_{x\in\mathcal{H}}\Big[f(x') + \frac{1}{2\mu}\|x - x'\|^2\Big], \\ \Rightarrow \inf_{x\in\mathcal{H}} f_\mu(x) &= \inf_{x'\in\mathcal{H}} f(x'), \end{aligned} \tag{2.22}$$

where in the second line the infimum over x is trivially given by $\delta_{x,x'}$ and thus the final line follows. Therefore, introducing a Moreau-Yosida envelope minimization problem — *i.e.* substituting the differential for the corresponding proximal projection — around non-differentiable terms $f(x)$ in a convex optimisation problem for which strong duality holds necessarily converges to the same solution (Yang, 2015).

Proximal projection of the ℓ_1-norm

As discussed previously in section 2.1.3, one often wishes to adopt sparsity promoting regularisation functionals. The function which most naturally measures sparsity is given by the ℓ_0-norm, or the *Hamming distance*, however the ℓ_0-norm is non-convex and typically far from smooth, resulting in optimisations which are at best, highly complex, and at worst NP-complete. One can adopt the ℓ_0-norm and attempt to locate global minima through locally optimal choices, *i.e.* through greedy algorithms (Temlyakov, 1999; Donoho et al., 2012), though such approaches do not support desirable certificates of optimality.

One might reasonably consider adopting the ℓ_1-norm $\|x\|_1 \in \mathcal{C}$ for $x \in \mathcal{H} \subseteq \mathbb{R}^n$, which is defined as

$$\ell_1\text{-norm} \equiv \|x\|_1 \equiv \sum_{i=0}^{n} |x_i|, \tag{2.23}$$

to approximate the ℓ_0-norm. In fact, for sufficiently sparse x (Donoho, 2006), solutions to optimisations regularised by the ℓ_0 and ℓ_1-norm functions are concurrent. This approximation, colloquially referred to as *convex relaxation*, is a useful tool which allows one to enforce sparse regularisation whilst retaining convexity of the objective. However, the ℓ_1-norm is discontinuous, and therefore non-differentiable, at the origin and thus one must consider the proximal projection to access the gradient information required to locate extremal solutions.

The proximal projection of $\lambda\|\cdot\|_1$, with positive semi-definite regularisation coefficient $\lambda \in \mathbb{R}_{\geq 0}$ can be derived by forming the Moreau-Yosida envelope (Moreau, 1962)

$$\text{Prox}_{\lambda\|\cdot\|_1}(x) = \underset{x' \in \mathcal{H}}{\text{argmin}} \left[\frac{1}{2}\|x - x'\|_2^2 + \lambda\|x'\|_1 \right], \tag{2.24}$$

which in this case is block separable and thus may be solved for each index $i \in \mathbb{Z}_{\geq 0}$ by

$$\mathbf{0} \in \nabla\left[\frac{1}{2}(x_i - x_i')^2\right] + \partial\left[\lambda|x_i'|\right] = x_i' - x_i + \lambda\partial\{|x_i'|\}\Big]. \tag{2.25}$$

As the modulus sub-gradient is given by $\partial\{|x_i'|\} = \text{sign}(x_i') \, \forall \, x_i' \neq 0$ the optimal solution to equation 2.25 $x_i^\star$ is given by $x_i^\star = x_i - \lambda \text{sign}(x_i'^{,*})$. Notice from this equation that $x_i^\star < 0 \Rightarrow x_i < -\lambda$ and correspondingly $x_i^\star > 0 \Rightarrow x_i > \lambda$. Therefore $\forall x_i' \neq 0$ one has $|x_i| > \lambda$ and $\text{sign}(x_i) = \text{sign}(x_i')$ and thus $x_i^\star = x_i - \lambda \ \text{sign}(x_i)$. Note that the sign function for complex numbers is adjusted to $x_i/|x_i|$ when $|x_i| > 0$. Consider now the case $x_i' = 0$ where the modulus sub-gradient is the interval $[-1,1]$ and thus the solution to equation 2.25 is given by

$$\mathbf{0} \in -x_i + \lambda[-1,1] \Rightarrow x_i \in [-\lambda, \lambda] \Rightarrow |x_i| \leq \lambda, \tag{2.26}$$

and therefore leads us to the proximal projection of the ℓ_1-norm defined generally to be

$$\mathrm{Prox}_{\lambda\|\cdot\|_1}(x_i) := \begin{Bmatrix} 0 \text{ if } |x_i| \leq \lambda, \\ x_i(1-\lambda/|x_i|) \text{ if } |x_i| > \lambda. \end{Bmatrix} = \mathrm{Soft}_\lambda(x) \tag{2.27}$$

This operator is commonly known as the piecewise soft-thresholding operator (Boyd, 2004; Parikh et al., 2014). Note further that if instead one considers $\lambda\|\Psi(\cdot)+b\|_1$ for a unitary linear transformation $\Psi : \mathcal{H} \rightarrow \mathcal{H}'$ *s.t.* $\Psi^\dagger\Psi = (1/d)\mathbb{I}$, where $b \in \mathcal{H}'$, and $d \in \mathbb{R}_{\geq 0}$ the proximal projection is given by

$$\mathrm{Prox}_{\lambda\|\Psi(\cdot)+b\|_1}(x) = x + d\Psi^\dagger\left(\mathrm{Soft}_{\lambda/d}(\Psi x + b) - \Psi x - b\right), \tag{2.28}$$

which follows straightforwardly from the fundamental properties of proximal calculus *e.g.* pre/postcomposition, affine addition (Combettes and Pesquet, 2011). This facilitates derivation of optimisation iterations required for analysis setting, in which Ψ takes the form of some sparsifying dictionary (a set of basis functions which need not necessarily be orthogonal) *e.g.* a wavelet dictionary. In certain cases the proximal operator does not support an analytic form, hence further sub-iterations are required which are beyond the scope of this discussion (see *e.g.* Cai et al., 2018b).

Proximal projection of the ℓ_2-norm

In cosmology, amongst other application domains, *e.g.* texture reconstruction, one might reasonably regularise an inverse problem by biasing towards solutions with particular statistical properties, which may be expected from theory. By construction, such functions exhibit distinct similarities to the logarithm of Gaussian distributions (see *e.g.* Hiller and Chin, 1990; Horowitz et al., 2019; Kodi Ramanah et al., 2019), resulting in the ℓ_2-norm function $\|x\|_2 \in \mathcal{C}$ for $x \in \mathcal{H}$, which is defined as

$$\ell_2\text{-norm} \equiv = \|x\|_2 \equiv \sqrt{\sum_{i=0}^{n} |x_i|^2}. \tag{2.29}$$

The proximal projection of $\|\cdot\|_2$ with regularisation coefficient $\lambda \in \mathbb{R}_{\geq 0}$ approximates the gradient and can be derived by forming the Moreau-Yosida (Moreau, 1962) envelope

$$\mathrm{Prox}_{\lambda\|\cdot\|_2}(x) = \underset{x' \in \mathcal{H}}{\mathrm{argmin}} \left[\frac{1}{2}\|x - x'\|_2^2 + \lambda\|x'\|_2\right]. \tag{2.30}$$

Now consider the components of $x' \in \mathcal{H}$ parallel and perpendicular to x, *e.g.* a substitution such as $x' = \frac{cx}{\|x\|_1} + y$ for $y \in \mathcal{H}$ *s.t.* $x^T y = 0$ and for scaling constant $c \in \mathbb{R}_{\geq 0)}$, then the envelope becomes the 2-dimensional minimization

$$\underset{y \in \mathcal{H}, c \in \mathbb{R}_{\geq 0}}{\mathrm{argmin}} \left[\frac{1}{2}\|y\|_2^2 + \frac{1}{2}(c - \|x\|_2)^2 + \lambda\sqrt{c^2 + \|y\|_2^2}\right], \tag{2.31}$$

which is clearly minimized for $y = 0$. At this minimum the problem reduces for

$\|x\|_2 > \lambda$ to the 1-dimensional minimization

$$
\begin{aligned}
\mathrm{Prox}_{\lambda\|\cdot\|_2}(x) &= \underset{c\in\mathbb{R}_{\geq 0}}{\mathrm{argmin}}\left[\frac{1}{2}(c-\|x\|_2)^2 + c\lambda\right], \\
\nabla\left[\frac{1}{2}(c-\|x\|_2)^2 + c\lambda\right] &\Rightarrow c^\star - \|x\|_2 + \lambda = 0 \\
&\Rightarrow c^\star = \|x\|_2 - \lambda.
\end{aligned} \tag{2.32}
$$

When $\|x\|_2 \leq \lambda$ the scalar $c^\star$ which minimizes the objective is clearly $c^\star = 0$. Together this simply gives us the proximal projection of $\|\cdot\|_2$ as

$$
\mathrm{Prox}_{\lambda\|\cdot\|_2}(x) = \max\left\{\|x\|_2 - \lambda, 0\right\}\frac{x}{\|x\|_2} \tag{2.33}
$$

Proximal projection of the indicator function of the ℓ_2-ball

---

For the primal problem, *i.e.* constrained optimisation, the data fidelity term $f_i(x) \in \mathcal{C}$ is included as a hard constraint which constitutes an indicator function on the f_i-ball $\mathbb{1}_{\mathcal{B}^\delta_{f_i}}$ which is a projection onto a sub-space $\mathcal{B}^\delta_{f_i} \subset \mathcal{H}$ such that

$$
\mathbb{1}_{\mathcal{B}^\delta_{f_i}} := \begin{cases} 0 & x \in \mathcal{B}^\delta_{f_i}, \\ +\infty & x \notin \mathcal{B}^\delta_{f_i}, \end{cases} \tag{2.34}
$$

for f-ball $\mathcal{B}^\delta_{f_i}$ centred at $z \in \mathcal{H}$ with radius $\delta \in \mathbb{R}_{\geq 0}$, defined by $\mathcal{B}^\delta_{f_i}(z) := \{x : f_i(x) \leq \delta\}$. Forming a Moreau-Yosida envelope (Moreau, 1962) as before we get

$$
\begin{aligned}
\mathrm{Prox}_{\mathbb{1}_{\mathcal{B}^\delta_{f_i}}}(x) &= \underset{x'\in\mathcal{H}}{\mathrm{argmin}}\left[\|x-x'\|_2^2 + \mathbb{1}_{\mathcal{B}^\delta_{f_i}}(x')\right], \\
&= \underset{x'\in\mathcal{B}^\delta_{f_i}}{\mathrm{argmin}}\left[\|x-x'\|_2^2\right] = \mathrm{Proj}_{\mathcal{B}^\delta_{f_i}}(x)
\end{aligned} \tag{2.35}
$$

i.e. the $\mathrm{Prox}_{\mathbb{1}_{\mathcal{B}^\delta_{f_i}}}$ operator is simply the projection onto the f_i-ball $\mathrm{Proj}_{\mathcal{B}^\delta_{f_i}}(x)$ which is given intuitively, which is to say geometrically, for $f_i = \|\cdot\|_2^2 = \ell_2^2$ by

$$
\mathrm{Prox}_{\mathbb{1}_{\mathcal{B}^\delta_{\ell_2^2}}}(x) = \begin{cases} x & x - z \in \mathcal{B}^\delta_{\ell_2^2}(z = \mathbf{0}), \\ \frac{x-z}{|x-z|}\delta + z & x - z \notin \mathcal{B}^\delta_{\ell_2^2}(z = \mathbf{0}), \end{cases} \tag{2.36}
$$

where $\mathcal{B}^\delta_{\ell_2^2}(z = \mathbf{0})$ is the ℓ_2-ball centred on the origin.

2.2.5 Proximal algorithms

Having demonstrated how proximal operators may be leveraged to access gradient information, it is informative to highlight how this information can be incorporated into classical optimisation algorithms to form proximal optimisation algorithms. As a large array of classical algorithms has been developed, many of which facilitate proximal versions, there are in principle a wide variety of proximal algorithms to choose from, *e.g.* proximal primal-dual (Parikh et al., 2014), proximal ADMM (Boyd et al., 2011; Parikh et al., 2014), some of which have recently been developed to support artificial intelligence techniques (Adler and Öktem, 2018). Throughout this thesis the proximal forward-backward splitting algorithm (Combettes and Pesquet, 2011) is typically adopted, though it is worth stressing that the contributions presented in this thesis are largely independent of the choice of algorithm.

Proximal Forward-Backward Splitting

Consider again the canonical optimisation problem of minimising an unconstrained objective composed of the sum of a proper *l.s.c.* convex function $f(x)$ and a convex β_{Lip}-Lipschitz differentiable convex function $g(x) \in \mathcal{C}$, for $x \in \mathcal{H}_1$. In this case the objective can be *split* into differentiable and non-differentiable components to define the *proximal gradient method*, which for our choice of functions f, g gives the *proximal forward-backward* iterations

$$x^{t+1} = \underbrace{\text{Prox}_{\lambda^t g}}_{\text{backward}} \underbrace{(x^t - \lambda^t \nabla f(x^t))}_{\text{forward}}, \tag{2.37}$$

where the forward step is a standard gradient descent over the differentiable function f, and the backward step is a proximal projection over the non-differentiable function g. The proximal gradient method can be interpreted as: a majorisation-minimisation or fixed point iteration, or (perhaps more satisfyingly) as an Euler discretized integration of a gradient flow differential equation (Parikh et al., 2014). Regardless of the choice of interpretation the forward-backward iteration steps can be seen in the pseudo-code of Algorithm 1. A full derivation and consideration of these iteration steps is beyond the scope of this thesis and can be found in external articles (Daubechies et al., 2004; Combettes and Wajs, 2005; Komodakis and Pesquet, 2015).

Algorithm 1 Proximal Forward-backward Splitting

```
    Input: $f(x)$, $g(x)$, $\mu \in \mathbb{R}_{>0}$, $x_{(0)}$
    Output: $x^{\text{opt}} \in \mathbb{C}^N$.
Do:
    $\nu^{(t)} \leftarrow x^{(t)} - \mu \nabla g(x^{(t)})$,
    $x^{(t+1)} \leftarrow \nu^{(t)} + \lambda(\text{prox}_{\mu f} \nu^{(t)} - x^{(t)})$,
    $t \leftarrow t + 1$.
Until: Stopping criterion satisfied.
```

Chapter 3

Probabilistic inference

This chapter provides an overview of Bayesian methodology with an emphasis on how it can be leveraged to produce scientific statements and, perhaps more importantly, quantify the degree to which such statements are plausible, *i.e.* the uncertainty inherent in such statements. In section 3.1.1 we introduce the notion of Bayes' theorem, before discussing, in detail, the terms associated with Bayes' theorem in sections 3.1.2 and 3.1.3. After exploring how one formulates inference (inverse) problems in a Bayesian sense, we discuss how to solve such problems in section 3.2, focussing primarily on sampling methods in section 3.2.1 and optimisation methods in section 3.2.2. Finally, in sections 3.2.3 onwards we show how one uses the results from aforementioned methods to quantify the degree to which solutions (scientific statements) are plausible, in a principled Bayesian manner.

> *"Probability theory is nothing but common sense reduced to calculation."*
>
> Pierre-Simon Laplace (1819)

3.1 Probability theory & scientific reasoning

Probability theory, a mathematical manifestation of logical reasoning (Jaynes, 2003), underpins a substantial portion of scientific reasoning today, including *e.g.* Bayesian inference and frequentist statistics. This becomes apparent when one notices that many, if not most, scientific problems can be described generally as inverse problems. Such problems are typically ill-posed, often severely so[1], leading to degenerate latent spaces, over which one wishes to reason. As such, they are most naturally solved in a probabilistic manner, hence aspects of probability theory are a near necessity for many, if not all, scientific disciplines. Interestingly, the theory of probability is isomorphic to the provably singular consistent set of axioms for plausible reasoning (Cox, 1946). If mathematics is the language of science, then the theory of probability is perhaps its most common vernacular.

Much of this thesis is focussed on Bayesian methodology, a sub-domain of probability theory which is prevalent throughout the cosmological scientific community. Bayesian inference provides a principled statistical framework, within which

---

[1] See section 2.1.2 for details in this regard.

quantification of the uncertainty in scientific propositions, in light of observed data and reasoned *a priori* expectations, exists naturally (Robert, 2001). In a Bayesian paradigm, once a probabilistic model is defined, the resulting probabilities directly describe the degree to which a given scientific proposition is plausible. Furthermore, such inferences are principled and easily interpretable. Contrasting this, frequentist type approaches fundamentally depend on the repeatability of outcomes, recovering direct statements of plausibility only in the limit of infinite realizations. In many areas of science, in particular the study of the Universe, it is typically difficult to observe very many realizations, hence frequentist reasoning in this regard is somewhat limited.

3.1.1 Bayes' theorem

The prototypical inverse problem is structured as follows: suppose one recovers observations y of some underlying quantity of interest x which may be degraded in some manner, *e.g.* through instrumental noise; given such data, under some assumed *forward model* $\mathcal{M}$ relating y with x, and with some *a priori* assumptions as to the nature of x. In such a paradigm, what scientific inferences can be made of x, and to what degree are these inferences plausible?[2] Such inverse problems can be described concisely by Bayes' theory

$$P(x|y,\mathcal{M}) = \frac{P(y|x,\mathcal{M})P(x)}{P(y|\mathcal{M})}, \tag{3.1}$$

where $P(A|B)$ denotes the conditional probability of event A occurring, given that B is known to have occurred. This equation is straightforwardly derived by considering that $P(A) + P(\bar{A}) = 1$ and $P(A,B) = P(A|B)P(B)$, *i.e.* the sum and product rules of probability theory respectively. The left-hand term of equation 3.1 is colloquially referred to as the *posterior distribution*, or *density*, and is the key quantity of interest for many aspects of Bayesian inference (see *e.g.* Sivia and Skilling, 2006). The posterior distribution is equivalent to the product of: the *likelihood* distribution $\mathcal{L}(x) = P(y|x,\mathcal{M})$ which encodes data-fidelity; the *prior* distribution $P(x)$ which encodes the practitioner's *a priori* assumptions as to the nature of x; and the reciprocal of the *Bayesian evidence* $z = P(y|\mathcal{M}) = \int dx P(y|x,\mathcal{M})$ which is a normalization factor which can be leveraged to conduct, *e.g.* model comparison, amongst other statistical tests (see *e.g.* Robert, 2001, for further details in this regard).

So far we have commented on how Bayes' theorem allows one to recast the conditional probability, *i.e.* the plausibility, of parameter configuration x under model $\mathcal{M}$ being the 'true' parameters, in light of observable data y. This process is colloquially referred to as *parameter inference* and is concerned with probabilistic

---

[2] Formal mathematical definitions of these objects is avoided as the discussion is quite general.

objects such as $P(x|y,\mathcal{M})$. Alternatively, one may be fundamentally interested in the probabilistic object $P(\mathcal{M}|y)$ *i.e.* a direct assessment of the plausibility of a model $\mathcal{M}$ in light of observations y (Cox, 1946; Good, 1962). In this thesis we are primarily concerned with parameter inference, hence much of the following discussion will focus on this application of Bayes' theorem.

3.1.2 Likelihood & data-fidelity

The first readily computable term $\mathcal{L}(x) = P(y|x,\mathcal{M})$, the likelihood distribution, is a conditional probability which encodes the odds of having observed data y under model $\mathcal{M}$, given some parameter configuration x. Typically one constructs $\mathcal{L}$ by first considering the forward model $\mathcal{M}$ to generate simulated observations $\tilde{y} = \mathcal{M}x$, before evaluating the potential sources of degradation in observations y *e.g.* the distribution of instrumental noise. In such cases the likelihood, more often than not, takes the form

$$\mathcal{L}(x) = \mathcal{F}(y - \tilde{y}, \mathcal{P}), \tag{3.2}$$

where $\mathcal{F}$ is a generic function which maps the difference between observed and simulated data $(y - \tilde{y})$ onto an assumed probability distribution $\mathcal{D}$, from which it is assumed the noise on observations y has been drawn, *i.e.* $n \sim \mathcal{D}$.

Note that $\mathcal{F}$ may, or may not, require additional free variables $\mathcal{P}$ which are often introduced to ensure $\mathcal{L}(x)$ can accurately model the data acquisition system. Further note that the function $\mathcal{F}$ may be difficult to formulate mathematically, and can instead be approximated through machine learning techniques (Gutmann et al., 2016; Alsing et al., 2019). There are therefore, at least, two components which require thought; these being specification of a model $\mathcal{M}$ and an error distribution $\mathcal{D}$. For the purposes of this thesis we will operate under the assumption that the forward model $\mathcal{M}$ is known to be correct, and thus are only concerned with specification of the distribution $\mathcal{D}$ from which our observation degradation is presumed to have been sampled.

The multivariate Gaussian likelihood

A central limit theory argument for the Gaussianity of $\mathcal{D} \sim \mathcal{N}(\mu, \Sigma)$ is often introduced such that the likelihood distribution is given by the exponential of the Mahalanobis distance (Mahalanobis, 1936)

$$\mathcal{L}(x) = \mathcal{F}(y - \tilde{y}, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|^{\frac{1}{2}}}} \exp\big(-\frac{1}{2}[y - \tilde{y} - \mu]^\dagger \Sigma^{-1} [y - \tilde{y} - \mu]\big), \tag{3.3}$$

where observations y are k-dimensional, μ is the mean vector, and Σ is the covariance of y modelling both the variance and inter-dependence of each observation y_i, for index $i \in \mathbb{Z}_{\leq k}$. For computational simplicity, one is often required to compute the natural log-likelihood which in this case is given by

$$\ln\left(\mathcal{L}(x)\right) = -\frac{k}{2}\ln(2\pi) - \frac{1}{4}\ln|\Sigma| - \frac{1}{2}(y - \tilde{y} - \mu)^{\dagger}\Sigma^{-1}(y - \tilde{y} - \mu) \tag{3.4}$$

In many cases Σ is taken to be diagonal, with elements $\Sigma = \text{diag}(\sigma^2)$, *i.e.* observations y are taken to be independent, therefore the log-likelihood reduces neatly to

$$\ln\left(\mathcal{L}(x)\right) = \overbrace{-\frac{k}{2}\ln(2\pi) - \frac{1}{4}\ln(\prod_i \sigma_i)}^{\text{constant}=C} - \frac{1}{2}\sum_i \overbrace{\Big(\frac{y_i - \tilde{y}_i - \mu_i}{\sigma_i}\Big)}^{z_i}{}^2 = C - \frac{1}{2}\|z_i\|_2^2, \tag{3.5}$$

which clearly demonstrates that a multivariate Gaussian likelihood (a Bayesian concept) reduces under the natural logarithm to a quadratic loss (an optimisation concept) over whitened[3] random variables. Furthermore, for independent random variables, *i.e.* diagonal covariances, the log-likelihood is straightforwardly given by the ℓ_2-loss. Due to the monotonicity of the logarithm function, should a practitioner choose to minimise the ℓ_2-loss to recover an estimate of x, *i.e.* ordinary least squares (OLS) optimisation, they are implicitly recovering the *maximum likelihood* solution (as the prior is implicitly uniform).

From the Gauss-Markov theorem, both OLS and maximum likelihood estimation are the minimum variance estimators of the set of linear unbiased estimators, and are thus perhaps 'optimal' in a somewhat naive sense. For inverse problems of interest, which are typically seriously ill-posed and or ill-conditioned, the adoption of a uniform prior (correspondingly a constant regularisation function over a potentially infinite domain) is patently sub-optimal. For such problems the information content of the data is often of too poor a quality to effectively constrain the posterior, and thus *a priori* bias can be introduced through weakly informative priors, so as to recover more desirable estimators overall. In this sense, a trade-off between bias and variance is implicit to all inverse problems, solved in such a manner. Though this similarity between optimisation functions and Bayesian likelihoods is somewhat obvious it is interesting nonetheless: when a ℓ_2-loss function is selected, whether it be for *e.g.* image-processing or machine learning, it is an implicit assumption of independent multivariate Gaussianty of the loss argument. Of course this avenue of comparison between loss functions and Bayesian likelihoods (and further Bayesian posteriors) is not limited to Gaussian distributions, though for brevity this discussion will be truncated here.

---

[3] A whitened random multivariate object is one in which the variance over each component is uniform, and often set to 1. This can be achieved, as in this case, by simply dividing each component by the corresponding standard deviation.

3.1.3 Prior specification & regularisation

The second term highlighted in the numerator of Bayes' theorem (equation 3.1) is the prior function $P(x)$, which captures any *a priori* assumed knowledge as to the nature of x. Conceptually, this knowledge codifies our *a priori* beliefs of the system in question, and can be predicated on past experimental results, *i.e.* through empirical Bayes (Casella, 1985), or simply prejudices inherited from theory (see *e.g.* Sivia and Skilling, 2006). The classical collections of prior distributions, from which one typically selects, are outlined below.

It should be noted that Bayesian priors have, until recently, typically been restricted to mathematical forms which can be implemented numerically. Such approximate priors are typically of limited efficacy for complex inference. Over the past decade, Bayesian inference problems have become increasingly complex, motivating researchers to develop more representative prior distributions (functions). Recently, researchers have developed methods to integrate so called *plug-and-play* machine learning techniques to learn far more realistic prior distributions directly from the data, drastically enhancing the state-of-the-art in Bayesian inference techniques (see *e.g.* Venkatakrishnan et al., 2013; Laumont et al., 2021).

Ignorance priors

In the absence of justifiable prior knowledge of, or prejudice regarding, x an uninformative prior is often adopted to communicate a lack of insight. One might reasonably interpret this as a statement of maximal ignorance of x, a concept which is rationalized in accordance with Laplace's *principle of indifference*. The idea of epistemic ignorance has since been discussed by other prominent academics, *e.g.* Henri Poincaré (Poincaré, 1912) and John Maynard Keynes (Keynes, 1921), and remains an area of mathematical philosophy to date.

One might reasonably, albeit naively, presume that being indifferent across the posterior space, *i.e.* assigning each possible solution equal weight, communicates a state of maximal *a priori* ignorance. Such a prior density is referred to as a uniform (sometimes flat) distribution, and is in some cases a reasonable choice. However, on closer inspection such priors can fail in a variety of ways: *e.g.* they are inconsistent under re-parameterization (see *e.g.* Trotta, 2017), as in the Borel-Kolmogorov paradox (Jaynes, 2003); they often suffer from the *concentration of measure* phenomenon, in which uniform sampling over n-dimensional spaces becomes asymptotically concentrated onto an infinitely thin $(n-1)$-dimensional shell as $n \rightarrow \infty$ (Giannopoulos and Milman, 2000), which is highly discriminate; and due to being improper, uniform prior distributions can lead to seemingly paradoxical results (see *e.g.* Stone, 1976).

Such failure modes are antithetical to intuitive inductive reasoning, and more broadly conflict with the basic desideratum of coherence. There exist a diverse spread of formal rules and principles by which ignorance priors, which are robust in this regard, may be constructed — the most widely accepted is the principle of transformation groups. This principle equates the state of 'total confusion' and 'complete ignorance', which is to say that one's hypothetical *a priori* degree of belief $f(\theta)$ is invariant under the action of a group $\mathcal{G}$, related to inherent symmetries of the space on which θ lives (see *e.g.* Jaynes, 1968; Kass and Wasserman, 1996). Consider a hypothetical observation of $X \sim \mathcal{N}(\theta, 1)$ for location family parameter $\theta \in \mathbb{R}$. Suppose an independent hypothetical observation Y was made such that $Y = X + a$, *i.e.* the observation is translated by $\tau_a : \mathbb{R} \mapsto \mathbb{R}$ such that $\tau_a(x) = x + a$. The group of all such translations is $\mathcal{G}_1 = \{\tau_a | a \in \mathbb{R}\}$ for $\tau_a \tau_b = \tau_{a+b}$. A prior density $\pi(x)$ is invariant under the action of $\mathcal{G}_1$ if, and only if, $\pi(\tau_a X) = \pi(X)$, which for location parameters is satisfied uniquely by the uniform prior.

Suppose instead $X \sim \mathcal{N}(\theta_1, \sigma_1^2)$, then τ_a becomes the affine transformation $\mathcal{A}_{a,b} : \mathbb{R} \mapsto \mathbb{R}$ such that $\mathcal{A}_{a,b}(x) = a + bx$, thus defining the affine transformation group $\mathcal{G}_2 = \{\mathcal{A}_{a,b} | a \in \mathbb{R}, b \in \mathbb{R}_{>0}\}$. Additionally, consider transformations $g_{a,b} : \mathbb{R} \times \mathbb{R}_{>0} \mapsto \mathbb{R} \times \mathbb{R}_{>0}$ over the parameter space such that $g_{a,b}(\theta, \sigma) = (a + b\theta, b\sigma)$, hence defining the parameter space transformation group $\mathcal{G}_3 = \{g_{a,b} | a \in \mathbb{R}, b \in \mathbb{R}_{>0}\}$. Under such conditions, should $X \sim \mathcal{N}(\theta_1, \sigma_1^2)$ and $Y = \mathcal{A}_{a,b} X$ then $Y \sim \mathcal{N}(\theta_2, \sigma_2^2)$ where $\theta_2, \sigma_2 = g_{a,b}(\theta_1, \sigma_1)$, hence the model is said to be invariant under $\mathcal{G}_3$. An ignorance prior which is invariant under left multiplication by $\mathcal{G}_3$, *i.e.* $\pi_L(g_{a,b} X) = \pi_L(X)$, is given by $\pi_L(\theta, \sigma) \propto \sigma^{-2}$. Equivalently, should one instead consider right multiplication, *i.e.* $\pi_R(X g_{a,b}) = \pi_R(X)$, then $\pi_R(\theta, \sigma) \propto \sigma^{-1}$. These two prior densities π_L, π_R correspond to the left and right invariant Haar measures respectively, the optimal selection of which is a topic far beyond the scope of this thesis (see *e.g.* Kass and Wasserman, 1996). Deriving invariant densities is, in general, non-trivial, motivating the relaxation to relative invariance, *i.e.* $\pi(X) \propto \pi(g_{a,b} X)$, which represent a larger class of priors which are more straightforward to locate (Hartigan, 1964).

Reference priors

Alternatively, and perhaps more intuitively, one may communicate a state of ignorance by maximizing the contribution of the data to the posterior (Bernardo, 1979). Conceptually, this is achieved by maximizing the shared information between the prior and the posterior, *i.e.* in a sense minimizing the new information introduced by the choice of prior. The mutual information $\mathcal{I}(x)$ between prior $P(x)$ and posterior distribution $P(x|y, \mathcal{M})$ is given by (Shannon, 1948)

$$\mathcal{I}(x) = \mathbb{E}_{y|\mathcal{M}} \Big[\int P(x|y, \mathcal{M}) \log \frac{P(x|y, \mathcal{M})}{P(x)} dx \Big], \tag{3.6}$$

where $\mathbb{E}_{y|\mathcal{M}}$ denotes the expectation under the likelihood, and the bracketed term is the Kullback-Leibler divergence (Kullback, 1997), which measures the relative entropy between two probability distributions. The reference prior $P(x)$ is consequently selected, either numerically or *via* variational methods, such that it maximizes $\mathcal{I}(x)$. For cases in which no nuisance parameters exist, the reference prior is given by Jeffreys' prior, *i.e.* the square root of the determinant of the Fisher information (see *e.g.* Jeffreys, 1946; Jaynes, 1968), which is invariant under re-parameterization. For scale parameters the Jeffreys' prior is equivalent to the aforementioned left invariant Haar prior (see *e.g.* Ghosh, 2011).

Conjugate priors

Conceptually, conjugate priors lie somewhat orthogonal to ignorance priors, in that they are not explicitly constructed with the principle of indifference in mind. Instead, conjugate priors are largely selected such that the posterior and prior densities belong to the same family (Schlaifer and Raiffa, 1961; Lindley, 1972). In this sense, the families to which the prior and posterior belong are said to be *conjugate*. Conjugate priors are primarily advantageous as they reduce a, potentially highly complex, Bayesian inference problem to an iterative modification of the (hyper-) parameters of the prior. This characteristic is convenient for both theoretical interpretation and computational implementation, particularly when the dimensionality of the prior is exceedingly large (Fink, 1997).

It has been shown that all exponential family probability distributions support conjugate priors (Diaconis and Ylvisaker, 1979), one such family being that of the Gaussian distribution, which features in later areas of this thesis. Suppose one draws likelihood from a zero-mean multivariate Gaussian distribution, *i.e.* $\mathcal{D} \sim \mathcal{N}(0, \Sigma_{\mathcal{L}})$, as given by equation 3.3. Further, suppose one holds *a priori* belief that x should be well approximated by a Gaussian random field; in such a case one may reasonably assign a Gaussian prior $P(x|\Sigma_{\mathcal{P}}) \sim \mathcal{N}(0, \Sigma_{\mathcal{P}})$, again with the functional form presented in equation 3.3. Note that priors selected in this way, though conjugate, are also typically weakly informative. Under these conditions the posterior distribution $P(x|y, \mathcal{M}, \Sigma_{\mathcal{L},\mathcal{P}})$ is given by

$$P(x|y, \mathcal{M}, \Sigma_{\mathcal{L},\mathcal{P}}) = \frac{\exp\Big[-\frac{1}{2}\big([y-\mathcal{M}x]^{\dagger}\Sigma_{\mathcal{L}}^{-1}[y-\mathcal{M}x] + x^{\dagger}\Sigma_{\mathcal{P}}^{-1}x\big)\Big]}{\sqrt{(2\pi)^{k+m}|\Sigma_{\mathcal{L}}\Sigma_{\mathcal{P}}|^{\frac{1}{2}}}}. \tag{3.7}$$

As the posterior still belongs to the family of Gaussian distributions, it is entirely described by its mean $\mu_{x|y}$ and covariance matrix

$$\Sigma_{x|y} \equiv \big[-\partial_x^2 \log P(x|y, \mathcal{M}, \Sigma_{\mathcal{L}}, \Sigma_{\mathcal{P}})\big]^{-1}. \tag{3.8}$$

A possible solution $x^{\star}$ of interest is that which, from the monotonicity of the logarithm function, extremizes (see *e.g.* MacKay, 2003) the log-posterior

$$\log P(x|y,\mathcal{M},\Sigma_{\mathcal{L},\pi}) \propto (y-\mathcal{M}x)^{\dagger}\Sigma_{\mathcal{L}}^{-1}(y-\mathcal{M}x)+x^{\dagger}\Sigma_{\mathcal{P}}^{-1}x, \tag{3.9}$$

which recovers the solution which maximizes the posterior odds, the *maximum a posteriori* (MAP) solution, given by

$$x^{\star} = (\Sigma_{\mathcal{P}}^{-1}+\mathcal{M}^{\dagger}\Sigma_{\mathcal{L}}^{-1}\mathcal{M})^{-1}\mathcal{M}^{\dagger}\Sigma_{\mathcal{L}}^{-1}y. \tag{3.10}$$

Hence, a Bayesian inference problem in which the prior is conjugate to a multivariate Gaussian likelihood returns, as a MAP estimator, the classic Wiener filter (Wiener et al., 1964). Wiener filters are routinely adopted throughout the disciplines of cosmology and astrophysics (see *e.g.* Elsner, F. and Wandelt, B. D., 2013; Jeffrey et al., 2018; Kodi Ramanah et al., 2019; Starck, J.-L. et al., 2021), however the inversion of matrices in equation 3.10, especially for modern complex datasets, can be substantially problematic, *e.g.* as they are rarely sparse. Alternatively, one may recover the Wiener filtering result by instead iteratively minimizing the negative log-posterior (see *e.g.* Elsner, F. and Wandelt, B. D., 2013; Horowitz et al., 2019), before leveraging either Laplace's approximation (see *e.g.* MacKay, 2003, pages 341-343) or the covariance of equation 3.8 to quantify the plausibility of such solutions.

Weakly informative priors

Priors which do not attempt to communicate a state of indifference, but instead attempt to introduce *a priori* assumed knowledge as to the nature of a given system are referred to as *weakly informative priors* – as they inform the posterior inference (see *e.g.* Lemoine, 2019). This class of prior is extremely broad and includes, *e.g.* regularizing priors and symmetry priors (sometimes structural priors). Regularisation priors attempt to promote solutions which exhibit problem specific regularity properties, typically stabilizing posterior inference. As such the discussion around them is inextricably linked to discussions around regularisation in optimisation (see section 2.1.3).

In the purist Bayesian sense, regularizing priors are often considered a frequentist construct (Rubin, 1984), and as such constitute a branch along which research surrounding Bayesian inference and optimisation collide. A popular example of a regularizing prior, is that of the Laplacian distribution $P(x|\mu,\lambda) \sim \text{Laplace}(\mu,\lambda)$, sometimes referred to as the double exponential distribution, which can be written as

$$P(x|\mu,\lambda) = \frac{\lambda}{2}\exp\big(-\lambda|x-\mu|\big), \tag{3.11}$$

for positive $\lambda \in \mathbb{R}_{>0}$, and mean μ. As the Laplace distribution can be considered the product of two Gaussian distributed variables, it exhibits a greater prior concentration around μ. Hence, for $\mu = 0$ a prior drawn from the Laplacian distribution biases towards solutions $x^\star$ for which many coefficients are zero-valued. In words, a Laplace prior biases towards solutions which are sparsely distributed (Tibshirani, 1996), as discussed in section 2.1.3. This becomes apparent when one notices that the negative log-prior for a Laplacian distribution is given by the ℓ_1-norm, the closest convex relaxation of the Hamming distance (Donoho, 2006) (see section 2.2.4).

3.1.4 Marginalization of nuisance parameters

A question outstanding from the previous subsections is how one should handle additional parameters introduced to the inference which, though only of tangential interest, can dramatically affect our final scientific statements. Such parameters z are called *nuisance parameters* and include, *e.g.* the scaling parameter λ used to parameterize the Laplacian distribution (see weakly informative priors in section 3.1.3). It is often unclear how to select such parameters, leading to inherent uncertainties which are accounted for by jointly inferring both x and z, before marginalizing (integrating) over z, *i.e.*

$$P(x|y,\mathcal{M}) = \int P(x,z|y,\mathcal{M})dz = \int \frac{P(y|x,z,\mathcal{M})P(x,z|\mathcal{M})}{P(y|\mathcal{M})}dz \tag{3.12}$$

where the final equality is simply a substitution of Bayes theorem given in equation 3.1. Formally, this is the Bayesian approach for dealing with nuisance parameters, however for an increasingly large set of inference (inverse) problems obtaining normalizations required to evaluate such integrals is intractable (Gelman and Meng, 1998). In such cases, more computationally efficient prior specific approaches are strongly motivated (see *e.g.* Pereyra et al., 2015).

3.2 Bayesian algorithms & uncertainty quantification

Having discussed how one constructs the posterior distribution, we will now discuss the mechanics of how the mathematical definition of the posterior distribution can be mapped onto scientific statements, with corresponding plausibilities. Though many approaches may be adopted, we will limit our discussion to the two most popular: Markov chain Monte Carlo (MCMC) sampling (see *e.g.* Robert and Casella, 2013, for a review), and Bayesian optimisation methods (see *e.g.* Robert, 2001, chapter 4). Furthermore, we will, to a reasonable extent, ground our discussion within the application of Bayesian methodology to image processing.

3.2.1 Sampling the posterior density

Markov chain Monte Carlo (MCMC) methods are designed to sequentially gather a large number of posterior samples, which under certain conditions can be shown to trace the underlying posterior distribution (Sivia and Skilling, 2006; Robert and Casella, 2013). Such sampling methods, as the name implies, are a composition of Markov methods – which describe the evolution of probability density in a dynamic system of states (Hastings, 1970; Gilks et al., 1995) – and Monte-Carlo methods – which recover numerical results through large scale (pseudo-) simulation.

A Markov chain is an n-sequence of ordered states $X = \{x_i \,|\, i \in \mathbb{Z}_{\leq n}\}$ which depend probabilistically on the previous state configuration x_{i-1}. For MCMC methods, a chain of states is sequentially generated such that the transition probability $\mathcal{T}(x'|x)$ between states x, x' satisfies detailed balance (Gilks et al., 1995)

$$P(x|y,\mathcal{M})\mathcal{T}(x'|x) = P(x'|y,\mathcal{M})\mathcal{T}(x|x'), \tag{3.13}$$

such that the limiting distribution of the Markov chain tends to the posterior $P(x|y,\mathcal{M})$, *i.e.* the process must be ergodic (see *e.g.* MacKay, 2003). The exact details of how one constructs such transition probabilities and state update schemes is the core concept from which the discipline of sampling methods has, at least in large part, been born. Often, though not always, such methods define functions $\mathcal{T}(x'|x)$ which exploit 1^{st}-order gradient information to explore the posterior space; typically *via* diffusion (Roberts et al., 1997). Those that are not reliant on gradient information, *e.g.* Metropolis-Hastings (Hastings, 1970), are typically computationally ineffective in high dimensional spaces (see *e.g.* Katafygiotis and Zuev, 2008).

Though gradient-based methods have been applied widely with resounding success, they are by construction limited to differential log-posteriors, hence they do not support current state-of-the-art image-processing priors, *e.g.* sparsity promoting Laplace distributions (see sections 3.1.3, 2.1.3). In recent years, such methods have been modernized (Pereyra, 2016; Durmus et al., 2018) by leveraging aspects of proximal analysis (see 2.2.4) to accommodate non-differentiable priors. These state-of-the-art MCMC methods are, somewhat unimaginatively, referred to as *proximal sampling methods.*

Proximal sampling methods

One such proximal MCMC algorithm adopted later in this thesis is the proximal modernization of the Metropolis-adjusted Langevin algorithm (MALA) (Robert and Casella, 2013), called the Px-MALA algorithm (Pereyra, 2016). Consider the Langevin diffusion (Roberts et al., 1997) of Lipchitz differentiable prior density $\pi \in \mathcal{C}^1$ for $x \in \mathbb{C}^n$,

$$d\mathcal{L}(t) = \frac{1}{2}\nabla \log \pi\left[\mathcal{L}(t)\right]dt + d\mathcal{W}(t), \tag{3.14}$$

for Brownian motion $\mathcal{W}(t)$. In the infinite t limit $d\mathcal{L}(t)$ approaches the posterior density, and as such a Langevin diffusion can be used to continuously sample the posterior density π. Discrete sampling is a practical necessity, hence a Euler-Maruyama approximation of the Markov chain which, under certain conditions

(Durmus et al., 2018), retains ergodicity is adopted

$$x^{(i+1)} = x^{(i)} + \frac{\delta}{2}\nabla \log \pi(x^{(i)}) + \sqrt{\delta}\omega^{(i+1)}, \tag{3.15}$$

where $\delta \in \mathbb{R}_{\geq 0}$ is the step-size of the discretization, and $\omega^{i+1} \sim \mathcal{N}(0, \mathbb{I}_n)$ is an n-dimensional sequence of Gaussian random variables. Equation 3.15 contains the term $\nabla \log \pi(x^{(i)})$ which is not defined if the density π is not strictly differentiable everywhere.

Consider the general case in which $\pi = f + g$ for lower-semi continuous log-prior $f \notin \mathcal{C}^1$ and Lipchitz-continuous log-likelihood $g \in \mathcal{C}^1$ which are both convex (see section 2.2.1). Under such conditions, and introducing a Moreau-Yosida (Moreau, 1962) envelope discussed in section 2.2.4, the Markov chain described by equation 3.15 becomes

$$x^{(i+1)} = \Big(1 - \frac{\delta}{\mu}\Big)x^{(i)} + \frac{\delta}{\mu}\text{prox}_{\mu f}(x^{(i)}) - \delta\nabla g(x^{(i)}) + \sqrt{2\delta}\omega^{(i)}, \tag{3.16}$$

which is named the MYULA (Moreau-Yosida unadjusted Langevin algorithm) Markov chain (Durmus et al., 2018). Such a chain is scalable, and exhibits competitive convergence properties (Roberts and Tweedie, 1996), however an asymptotic estimation bias exists. To mitigate this bias, at the cost of marginally reduced computational efficiency, a Metropolis-Hastings accept-reject step (Chib and Greenberg, 1995) is introduced, under which a proposed state x' is accepted with probability

$$p = \min\Big[1, \frac{\mathcal{T}(x|x')\pi(x')}{\mathcal{T}(x'|x)\pi(x)}\Big], \tag{3.17}$$

where $\mathcal{T}(x'|x)$ is the MYULA transition probability kernel (Pereyra, 2016).

Proximal algorithms, such as Px-MALA provide asymptotically exact avenues through which one can sample from, under the aforementioned conditions, posterior densities which are not strictly differentiable, and have been adopted for various applications (see *e.g.* Cai et al., 2018a; Price et al., 2019a). However, broadly speaking, sampling methods become computationally taxing in high dimensions as the number of samples required to achieve sufficient convergence increases dramatically. This motivates research into alternate, perhaps approximate or accelerated, methods with which one can recover, in a Bayesian sense, scientific statements with principled measures of plausibility.

3.2.2 Maximum a posteriori estimation

Once a sufficiently large set of posterior samples has been collected, the question becomes: which solution is the best? The irreducible representation of this question is simply: how does a Bayesian define best? Unfortunately this is a rather open ended question with countably infinite answers. Nonetheless, two straightforward potential answers come to mind. Suppose one simply averages over the posterior samples one recovers the posterior mean, which may seem reasonable: on average, this is the solution one may expect. Alternatively, one may select the solution which maximizes the posterior odds: this is the most plausible solution, referred to as the *Bayes estimator*. Effectively these two potential solutions correspond to the mean and maximum of the posterior distribution. It is not necessarily clear which point estimate best characterizes the posterior, hence the ambiguity around optimality.

In this thesis we will focus on methods which search for the *Bayes estimator* which maximizes the posterior odds, for quasi-concave functions (Boyd, 2004) (on which we are primarily focussed); this is given by the *maximum a posteriori* (MAP) solution. Given a posterior as defined generally by equation 3.1 one can locate the MAP solution by locating extremal solutions, *i.e.* solutions for which

$$\partial_x \big[P(x|y,\mathcal{M}) \big] = \partial_x \Big[\frac{P(y|x,\mathcal{M})P(x)}{P(y|\mathcal{M})} \Big] \propto \partial_x \big[P(y|x,\mathcal{M})P(x) \big] = 0, \tag{3.18}$$

where the proportionality follows from the realization that the evidence $P(y|\mathcal{M}) \in \mathbb{R}_{>0}$, and therefore acts as a scaling coefficient which does not affect the position of the solution. Typically this problem is solved over the log-posterior, as it is often more straightforward numerically. From the monotonicity of the logarithm function one can instead locate extremal solutions

$$\partial_x \big[\log P(x|y,\mathcal{M}) \big] \propto \partial_x \log \big[P(y|x,\mathcal{M})P(x) \big] = \partial_x \big[\log P(y|x,\mathcal{M}) + \log P(x) \big] = 0. \tag{3.19}$$

Under the canonical substitution $f(x) = \log P(x)$ and $g(x) = \log P(y|x,\mathcal{M})$, and considering the process of nullity under ∂_x to be isomorphic with standard minimisation, computing the MAP estimator reduces to the Lagrangian dual problem

$$x^{\text{MAP}} = \underset{x}{\operatorname{argmin}} \Big[h(x) = f(x) + g(x) \Big] \tag{3.20}$$

where $h(x)$ is the negative log-posterior, referred to as the *objective* or *loss* function. Notice that this equation takes the form seen in equation 2.16 implying that MAP estimation is unconstrained optimisation, however with the added benefit of a principled statistical interpretation (Kaipio and Somersalo, 2006). Estimation of the MAP solution has seen wide-spread adoption for high-dimensional problems in which $h(x)$ belongs to the space of convex functions, for which large-scale convex optimisation algorithms can be leveraged (Green et al., 2015), as discussed in section 2.2. Solving such problems *via* sampling methods, as discussed in section 3.2.1, is often computationally expensive, motivating the adoption of rapid convex optimisation techniques.

The reader may have noticed the somewhat glaring issue: though MAP estimation can be an effective tool for locating (in some sense) optimal solutions for Bayesian inference problems, it is unclear how one may quantify the plausibility of such solutions — a necessity for principled scientific inquiry. Leveraging recent results (Pereyra, 2017), in the following subsections, and as a core concept explored in this thesis, I will demonstrate how one can quantify the plausibility of such MAP solutions in a fully consistent, principled manner.

3.2.3 Posterior credible regions

Having now covered how to construct the Bayesian posterior we will now discuss the plausibility of scientific statements one may make using the posterior. Perhaps the most useful quantity one can calculate from the posterior is accessed through *credible regions* of the posterior (see *e.g.* Robert, 2001). Suppose one attempts to make scientific statements about parameters $x \in \mathbb{R}^n$, in light of observations $y \in \mathbb{R}^m$. A posterior credible region C_α at $100(1-\alpha)\%$ confidence is given by

$$p(x \in C_\alpha | y) = \int_{x \in \mathbb{R}^n} p(x|y, \mathcal{M}) \mathbb{I}_{C_\alpha} dx = 1 - \alpha, \tag{3.21}$$

where $\mathbb{I}_{C_\alpha}$ is the usual set indicator function defined by $\mathbb{I}_{C_\alpha}(x) = 1 \; \forall x \in C_\alpha$ and 0 elsewhere. In words, this is saying that a Bayesian credible region at $100(1-\alpha)\%$ confidence is a closed region of the posterior hyper-volume which contains $100(1-\alpha)\%$ of the probability, *i.e.* the statement that the true solution $x^{\text{true}} \in C_\alpha$ is assigned $100(1-\alpha)\%$ confidence. There are clearly many regions one may construct which satisfy equation 3.21, hence there is an infinite degeneracy as to the question of which region a Bayesian should select.

HPD-credible region

The decision-theoretical optimal credible region, in the sense of minimum volume, is referred to as the *highest posterior density* (HPD) credible region (Robert, 2001). The HPD credible region is defined concisely by the set

$$C_\alpha := \{x | P(x|y, \mathcal{M}) \leq \epsilon_\alpha\}. \tag{3.22}$$

where ϵ_α is an iso-contour of the posterior density, sometimes referred to as the level-set threshold, which satisfies $\int_{C_\alpha} P(x|y, \mathcal{M}) dx = 1 - \alpha$. Conceptually, one may quantify the plausibility of scientific statements in a relatively straightforward manner: recover a (pseudo-) optimal estimator for x, *e.g.* the MAP estimator as discussed in section 3.2.2; and locate the boundary values $x + \delta x$ for which the posterior $P(x + \delta x | y, \mathcal{M}) \in C_\alpha$ and saturates the level-set ϵ_α. One may then conclude that the true value of x is contained within the region spanned by δx with $100(1-\alpha)\%$ confidence.

However, to compute the level-set threshold ϵ_α associated with such posterior iso-contours requires evaluation of an n-dimensional integral over the parameter space. Consequently, for high-dimensional inferences the true HPD credible region is impractical to compute. As such, one is either restricted to low-dimensional parameter spaces, or approximate techniques must be developed.

Approximate HPD-credible region

A powerful approximate method which has recently been derived (Theorem 3.1 in Pereyra, 2017), which will be leveraged throughout this thesis, is that of the *approximate HPD credible region*, C'_α. This approximation is derived from probability concentration theory, and can be used to approximate the level-set threshold ϵ_α without the need to evaluate any, often computationally intractable, high-dimensional integrals. Suppose the posterior density $P(x|y,\mathcal{M})$ is log-concave over $\mathbb{R}^n$, and is composed of a prior and likelihood which both belong to the exponential family. Under such presuppositions, the posterior density can be written generally as $P(x|y,\mathcal{M}) = \exp[-h(x)]/z$ where $h(x)$ is as defined previously, *i.e.* the objective function (optimisation) or the product of the log-prior and log-likelihood (Bayesian), and z is the evidence. It has been shown (Theorem 1.2 in Bobkov and Madiman, 2011a) that the concentration of log-concave densities obeys

$$P\Big(\frac{1}{\sqrt{n}} \,|\, \log P(x|y,\mathcal{M}) - \mathbb{E}\big[\log P(x|y,\mathcal{M})\big] \,|\, \geq t\Big) \;\leq\; 3e^{-ct^2}, \tag{3.23}$$

where $\mathbb{E}$ is the standard expectation, $t \in [0, 2\sqrt{n}]$, and constant $c \in \mathbb{R}_{>0}$ which can, without loss of generality, be set to $1/16$. Consider the substitution $t = \tau\sqrt{n} \in [0,2]$, under which it can be shown (Lemma 3.1 in Pereyra, 2017) that

$$P\big(\,|\,h(x) - \mathbb{E}[h(x)]\,|\, \geq \tau n\big) \;\leq\; 3e^{-n\tau^2/16}, \tag{3.24}$$

This relation is a consequence of the concentration properties of log-concave random vectors, which implies that as the dimensionality n grows the probability becomes increasingly concentrated around the $(n-1)$-dimensional hyper-spherical shell $\{x : h(x) = \mathbb{E}[h(x)]\}$. This result is not dissimilar, albeit in a different domain, from the concentration of measure phenomenon (Giannopoulos and Milman, 2000) discussed previously. As per the definition of a credible region, see equation 3.21, the RHS term is equivalent to the confidence level. Hence the substitution parameter τ_α can be calibrated such that $\alpha = 3e^{-n\tau_\alpha^2/16} \Rightarrow \tau_\alpha = \sqrt{16\log(3/\alpha)/n}$. With these substitutions, and rearranging the expectation of the probability inequality of equation 3.23, one finds

$$P\big(h(x) \;\geq \mathbb{E}[h(x)] + \tau_\alpha n\big) \;\leq \alpha. \tag{3.25}$$

Leveraging proposition 1.2 (Bobkov and Madiman, 2011b), it can be shown (Pereyra, 2017) that for log-concave functions, with MAP solution x^{MAP}, the computationally intractable expectation of equation 3.25 can be bounded by $\mathbb{E}[h(x)] \leq h(x^{\text{MAP}}) + n$.

Substituting this expression for the expectation into equation 3.25, allows us to define the approximate credible set $C'_\alpha = \{x : h(x) \leq \epsilon'_\alpha\}$, with level-set threshold $\epsilon'_\alpha = h(x^{\text{MAP}}) + \tau_\alpha \sqrt{n} + n$. Importantly, by construction $C_\alpha \subseteq C'_\alpha$ making this approximation conservative, hence the degree of plausibility in a given statement of belief will never be over-estimated — which is a prudent safeguard against reaching incorrect conclusions. Furthermore, this approximation requires only that one recovers the computationally inexpensive MAP solution, from which one may quantify the plausibility of scientific statements in a principled manner (see *e.g.* Pereyra, 2017; Cai et al., 2018a,b; Price et al., 2019a,b, 2021a,b).

Of course this computational advantage comes at the cost of approximation error, which we necessarily must consider. As the approximation is conservative the error is positive, and thus bounded below at 0. Derivation of the the upper-bound (see Pereyra, 2017) is drastically more complicated and beyond the scope of this discussion. Nevertheless, the inequality which constrains the approximation error is given by

$$0 \leq \epsilon'_\alpha - \epsilon_\alpha \leq \eta_\alpha \sqrt{n} + n, \tag{3.26}$$

where $\eta_\alpha = \sqrt{16 \log(3/\alpha)} + \sqrt{1/\alpha}$. In high dimensional settings, *i.e.* $n \rightarrow \infty$, the upper bound of this error may appear large, however in practice the error is relatively small (as found in *e.g.* Cai et al., 2018b; Price et al., 2019a). In fact, in later chapters of this thesis it will be demonstrated that this approximation error is very small indeed. Perhaps most crucially, this approximation circumvents calculation of the high dimensional integral present in equation 3.21, and so can be computed efficiently in arbitrarily high dimensional settings.

Chapter 4

Gravitational lensing & dark matter

In this chapter we will consider, in some detail, the relevant mathematical background underpinning the field of gravitational lensing, with a particular focus on how the quantities of interest emerge from theory and how they can be modelled on the sky. The structure of the chapter is as follows. To begin our discussion, in section 4.1 we will introduce core Cosmological concepts *e.g.* in subsection 4.1.1 we discuss space-time metrics and geodesics before using perturbation theory to derive the fundamental lens equations underpinning gravitational lensing in section 4.1.4. Following this, in section 4.2.4 we will explicitly draw relations between theoretical variables and physically observable fields, and highlight how one can recover cosmological information from such observations through solving severely ill-posed inverse problems (see section 8.3 for a discussion on this topic). Finally, in section 4.2.4 we explore the current leading algorithms developed by the weak lensing community for solving such inverse problems. Though substantial mathematical discussion is presented throughout this chapter, review articles are recommended for a broader discussion of the field (see *e.g.* Bartelmann and Schneider, 2001; Bartelmann, 2010).

> *"The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve"*
>
> Eugene Wigner (1960)

4.1 Weak gravitational lensing

Originating from Newton's *'Naturalis Principia Mathematica'* published in 1687, modern Cosmology is based on the fundamental assumption that the Universe as we observe it is *homogeneous* and *isotropic*. In other words we sit at no special position in space or time, often referred to as the *'Copernican Principle'* despite widely being attributed to Giordano Bruno (before being burnt at the stake for 'heresy' in 1600). *Homogeneity* is a property which implies that a vector field is identical in magnitude and direction at every co-ordinate, *i.e.* that the structure of the Universe is translation invariant. It is perhaps obvious that the late universe is, in fact, far from homogeneous, however deviations from homogeneity on a universal scale are substantially damped, and so to a good approximation the Universe can be considered homogeneous (Planck Collaboration, 2018). Interestingly, inhomogeneities are currently believed to be the late Universe manifestations of quantum level fluctuations, dramatically magnified during a postulated inflationary period during the early evolution of the cosmos (see *e.g.* Challinor and Peiris, 2009). *Isotropy* is the property of a vector field which states that the statistical properties of the Universe are independent of direction, *i.e.* the physical laws of the Universe are rotation invariant. Importantly, isotropy manifests itself in the aggregate statistics of the universe rather than the absolute distribution – so viewed in a given direction the overall shape on different scales is, on average, uniform (Planck Collaboration, 2018). Again, on small scales the universe we observed today is far from isotropic, however on a cosmic scale these anisotropies are comparatively negligible (Saadeh et al., 2016).

The fabric of such a Universe is perhaps best described mathematically by (pseudo-) Riemannian manifolds, upon which all matter and energy exists, and can move and interact freely under the known forces of nature. Theory suggests that matter, amongst other potential factors, dictates the topology of the manifold upon which it exists, which in turn determines the equations of motion by which the matter must abide. Conversely, massless particles, *e.g.* photons, are forced to travel along null geodesics determined by the Riemannian metric of a prescribed manifold interpretation of the Universe (see *e.g.* Hobson et al., 2006). As the path of massless particles depends on the metric, which in turn is dependent on the dynamical distribution of matter in the Universe, one can construct inverse problems (see chapter 2) or equivalently inference problems (see chapter 3) through which cosmological information may be extracted.

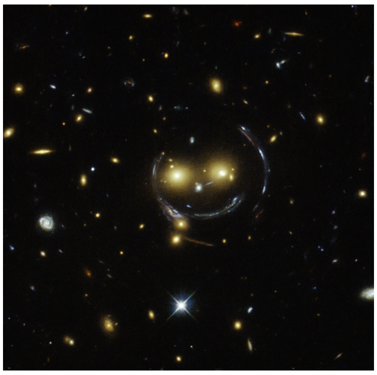


Figure 4.1: Hubble Space Telescope image of the SDSSCGB:8842.3 and 8842.4 galaxies within *Ursa Major*. Note the arc like structures surrounding the central halos: these are caustic curves along which images of galaxies within an Einstein radius have been spread. These events are referred to as *strong lensing events*. *Credit to NASA & ESA.*

Specifically, gravitational lensing is the term ascribed to the study of the apparent deflection of photons as they travel from a distant source to observers here and now. Such deflections indicate curvature of the underlying manifold, which in turn can be used to infer properties of, amongst other quantities, the intervening matter distribution – both observable and dark. The gravitational lensing effect is sensitive over many scales; ranging from micro-lensing which can be used *e.g.* for exoplanet detection (Gaudi, 2012), to strong-lensing (see Figure 4.1) which can be used for the study of *e.g.* dark-matter profiles (see *e.g.* Kochanek, 1991; Koopmans and Treu, 2003) or dark-matter self-interactions (Markevitch et al.,

2004). Weak gravitational lensing is concerned
with distortions to images of distant galaxies, in particular linear order perturbations to their ellipticity (third-flattening) and apparent magnitude (brightness). As the name implies, such perturbations are particularly small, hence in practice the effect is integrated over billions of galaxies prior to the extraction of cosmological information. In this sense, weak gravitational lensing is expanding into the so called 'big data' regime, particularly with next generation surveys on the not too distant horizon. In the following sections we will discuss how the weak gravitational lensing phenomenon emerges as a consequence of general relativity; beginning by outlining the space-time metric, before deriving the core relations which underpin the study of gravitationally lensed images. Subsequently, the discussion will cover how observations over the celestial sphere here and now can be modelled, from which cosmological information may be inferred.

4.1.1 Space-time metrics: the Robertson-Walker metric

General relativity is built upon differential geometry, the analysis of fields defined over manifolds. Fundamentally, such analysis requires an, ideally analytic, measure of the distance between points, referred to as the Riemannian metric. The manifold interpretation of the universe was first introduced to Cosmology in the early 20th century (Friedmann, 1922; Lemaître, 1933; Robertson, 1935; Walker, 1937), and has led to the current adoption of the term *space-time manifold.* In the absence of time, the space-time metric reduces to the Euclidean measure

$$dl^2 = dx^2 + dy^2 + dz^2, \tag{4.1}$$

where there is clearly no time dependence. When a fourth, time-like, dependence is introduced the flat $3+1$-dimensional space-time measure (3 space-like dimensions and 1 time-like dimension) is simply given by the Minkowski metric

$$ds^2 = c^2 dt^2 - \underbrace{dx^2 - dy^2 - dz^2,}_{dl^2} \tag{4.2}$$

where c is the speed of light in a vacuum. Consider further the case where the universe has some intrinsic curvature K, under which the spatial measure dl^2 (projected onto spherical polar co-ordinates $x, y, z \mapsto r, \theta, \phi$, for mathematical simplicity) transforms to (Hobson et al., 2006)

$$\Rightarrow ds^2 = c^2 dt^2 - \left(\frac{dr}{\sqrt{1 - Kr^2}} \right)^2 - r^2 d\theta^2 - r^2 \sin^2(\theta) d\psi^2, \tag{4.3}$$

from which, should one transform to co-moving distance $r \rightarrow r/a(t)$ and note that $K = k/a^2(t)$, one recovers the *co-moving Robertson-Walker* measure

$$ds^2 = c^2 dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 \underbrace{\left(d\theta^2 + \sin^2\theta d\psi^2\right)}_{\equiv d\Omega^2}\right]. \tag{4.4}$$

Let $d\chi \equiv dr/\sqrt{1-kr^2}$, from which we can define,

$$f_K(\chi) = \begin{cases} \sin(\chi) & \text{if} \quad k = 1 \Rightarrow \mathbb{S}^3, \\ \chi & \text{if} \quad k = 0 \Rightarrow \mathbb{R}^3, \\ \sinh(\chi) & \text{if} \quad k = -1 \Rightarrow \mathbb{H}^3, \end{cases} \tag{4.5}$$

for spherical, flat, and hyperbolic universes respectively. This expression allows us to simplify the measure to

$$ds^2 = dt^2 - a^2(t)\left[d\chi^2 + f_K^2(\chi)d\Omega^2\right] = a^2(\tau)\left[d\tau^2 - (d\chi^2 + f_K^2(\chi)d\Omega^2)\right]. \tag{4.6}$$

Finally, adopting the full tensor notation we can write the measure as $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, where the Einstein summation convention is adopted and $g_{\mu\nu}$ is a rank-2 tensor called the space-time metric or often just metric. The Robertson-Walker metric (Robertson, 1935; Walker, 1937) can then be read from the measure

$$g_{\mu\nu} = a^2(\tau)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -f_K^2(\chi) & 0 \\ 0 & 0 & 0 & -f_K^2(\chi)\sin^2(\theta) \end{bmatrix}. \tag{4.7}$$

4.1.2 Perturbation theory and gauge transformations

For clarity, we will now discuss how a general space-time metric is perturbed within the framework of cosmological perturbation theory. In a flat FLRW homogeneous background space-time (Friedmann, 1922; Lemaître, 1933; Robertson, 1935; Walker, 1937) the unperturbed metric is given by

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = a^2(\tau)\left[d\tau^2 - \delta_{ij}dx^i dx^j\right], \tag{4.8}$$

with Greek indices $\in [0,3]$ and roman indices $\in [1,3]$ with the 0^{th} component being time-like (the standard convention). Suppose this metric includes tensor perturbations A, B_i, and h_{ij} of varying rank such that (Weinberg, 2008)

$$ds^2 = a^2(\tau)\left[(1+2A)d\tau^2 - 2B_i dx^j d\tau - (\delta_{ij} + h_{ij})dx^i dx^j\right], \tag{4.9}$$

where A are rank-0 scalar perturbation, B_i are rank-1 vector perturbations of rank 1, and h_{ij} are rank-2 tensor perturbations. Consider now the SVT (scalar, vector, tensor) decomposition of these perturbed components (Stewart, 1990). As A is a scalar perturbation it consists only of a scalar component, whereas B_i contains both scalar and vector components such that $B_i = \partial_i B + \hat{B}_i$. Correspondingly, the rank-2 symmetric tensor h_{ij} can be decomposed into scalar, vector and tensor components such that,

$$\begin{aligned} h_{ij} &= 2C\delta_{ij} + 2\partial_{\langle i}\partial_{j\rangle}E + 2\partial_{(i}\hat{E}_{j)} + 2\hat{E}_{ij}, \\ &= 2C\delta_{ij} + 2\underbrace{\left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)}_{\partial_{\langle i}\partial_{j\rangle}} E + \underbrace{\partial_i\hat{E}_j + \partial_j\hat{E}_i}_{\partial_{(i}\hat{E}_{j)}} + 2\hat{E}_{ij}, \end{aligned} \tag{4.10}$$

where we have explicitly included the expansion of the notational shorthand which we will adopt for $\partial_{\langle i}\partial_{j\rangle}$ and $\partial_{(i}\hat{E}_{j)}$. Note that all vector quantities are divergence free $\partial^i\hat{B}_i = \partial^i\hat{E}_i = 0$ and all tensor quantities are transverse and traceless $\partial^i E^T_{ij} = \delta^{ij}E^T_{ij} = 0$. Further note that when considering Einstein equations, as we often do in cosmology, there exists no linear order SVT mixing and so components are trivially separable.

Let us now adopt the unperturbed metric presented in equation 4.8 and perform an affine spatial transformation $x_i \mapsto \tilde{x}^i = x^i + \xi^i(\tau, x)$ for some $\xi \ll 1$. Under this affine transformation, the differential element dx^i is transformed, at linear order, such that

$$\begin{aligned} dx^i &= d\tilde{x}^i - \partial_\tau\xi^i d\tau - \partial_k\xi^i d\tilde{x}^k, \\ \Rightarrow ds^2 &= a^2(\tau)\left[d\tau^2 - 2\partial_\tau\xi_i d\tilde{x}^i d\tau - (\delta_{ij} + 2\partial_{(i}\xi_{j)})d\tilde{x}^i d\tilde{x}^j\right], \end{aligned} \tag{4.11}$$

Interestingly by simply applying a spacial transformation we have created spurious vector perturbations $B_i = \partial_\tau\xi_i$ and $\hat{E}_i = \xi_i$. These artificial perturbations are called *gauge modes* and can be introduced just as easily through equivalent temporal transformation *e.g.* $\tau \mapsto \tau + \xi^0(\tau, x)$. Physical quantities must necessarily be gauge-invariant, which is to say that physical quantities must be invariant under the group of Lorentz transformations, *i.e.* the concatenation of the non-Abelian Lie group of rotations and boosts. In fact, as the physics of a system must further be unchanged under affine translations, they must be Poincaré invariant (Hobson et al., 2006). In essence, this simply states that physical theories should be consistent between frames of reference.

To address such gauge-modes, begin by defining an arbitrary co-ordinate transformation $X^\mu \mapsto \tilde{X}^\mu + \xi^\mu(\tau, x)$ where $\xi^0 \equiv T$, and $\xi^i \equiv L^i = \partial^i L + \hat{L}^i$. As the space-time interval is by definition invariant, as one might reasonably expect, one may say that

$$ds^2 = g_{\mu\nu}dX^\mu dX^\nu = \tilde{g}_{\alpha\beta}d\tilde{X}^\alpha d\tilde{X}^\beta, \tag{4.12}$$

for dummy 4-vector indices α, β, and original or transformed metric $g_{\mu\nu}$ and $\tilde{g}_{\alpha\beta}$ respectively. Rearranging equation 4.12 one finds a relation between the original and transformed metrics such that

$$g_{\mu\nu} = \frac{\partial \tilde{X}^{\alpha}}{\partial X^{\mu}} \frac{\partial \tilde{X}^{\beta}}{\partial X^{\nu}} \tilde{g}_{\alpha\beta}. \tag{4.13}$$

Considering permutations of $\mu, \nu, \alpha, \beta \in [0,3]$ one may construct equations which govern the transformation of metric components (Challinor, 2009),

$$\begin{aligned} A &\mapsto \tilde{A} = A - T' - \mathcal{H}T, \\ B_i &\mapsto \tilde{B}_i = B_i + \partial_i T - \partial_\tau L_i, \\ h_{ij} &\mapsto \tilde{h}_{ij} = h_{ij} - 2\partial_{(i}L_{j)} - 2\mathcal{H}T\delta_{ij}, \end{aligned} \tag{4.14}$$

where $\mathcal{H}$ is the conformal Hubble parameter $\mathcal{H} \equiv \partial_\tau a/a$. With this in mind, one may now define the gauge-invariant *Bardeen potentials* (Bardeen, 1980):

$$\begin{aligned} \Psi &\equiv A + \mathcal{H}(B - \partial_\tau E) + \partial_\tau(B - \partial_\tau E), \\ \Phi &\equiv -C - \mathcal{H}(B - \partial_\tau E) + \frac{\nabla^2 E}{3}. \end{aligned} \tag{4.15}$$

Alternatively one can exploit a degeneracy of the gauge functions T and L to nullify two of the scalar metric perturbations, which may result in a simplification of the algebra relating to certain cosmological theories (Hobson et al., 2006). Perturbation theory permits the selection of gauge, though it is required that *physical observable quantities* must necessarily be invariant to this choice, or *gauge-invariant*. A complete description of cosmological perturbation theory, with an emphasis on the *synchronous* and *Newtonian* gauges (which is often adopted to simplify the mathematics of gravitational lensing) can be found in (*e.g.* Ma and Bertschinger, 1995).

4.1.3 The geodesic equation

The geodesic equation defines the path of particles over a manifold. It can be derived by first defining the Lagrangian:

$$\mathcal{L} = \frac{1}{2} g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = \frac{1}{2} g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu, \tag{4.16}$$

where $\dot{x} = dx/d\lambda$ with affine parameter λ. We can now use $\mathcal{L}$ as the argument for the standard Euler-Lagrange equations

$$\begin{aligned} \frac{d}{d\lambda} \frac{\partial \mathcal{L}}{\partial (dx^\mu/d\lambda)} = \frac{\partial \mathcal{L}}{\partial x^\mu} &\Rightarrow \frac{1}{2}(g_{\mu\lambda,\nu}\dot{x}^\mu \dot{x}^\nu + g_{\nu\lambda,\mu}\dot{x}^\mu \dot{x}^\nu) + g_{\mu\lambda}\ddot{x}^\mu = \frac{1}{2} g_{\mu\nu,\lambda}\dot{x}^\mu \dot{x}^\nu \\ &\Rightarrow g_{\mu\lambda}\ddot{x}^\mu = -\frac{1}{2} \underbrace{(g_{\mu\lambda,\nu} + g_{\nu\lambda,\mu} - g_{\mu\nu,\lambda})}_{\equiv 2\Gamma_{\lambda\mu\nu}} \dot{x}^\mu \dot{x}^\nu, \end{aligned} \tag{4.17}$$

where $\Gamma_{\lambda\mu\nu}$ denotes Christoffel symbols which appear throughout the mathematics of general relativity. Contracting the inverse metric simplifies the LHS by $g^{\alpha\mu} g_{\mu\lambda} =$

$\delta^{\alpha}_{\lambda} \Rightarrow g^{\alpha\mu}g_{\mu\lambda}\ddot{x}^{\mu} = \ddot{x}^{\lambda}$, for $\alpha = \lambda$. The RHS is correspondingly transformed, resulting in the geodesic equation

$$\ddot{x}^{\lambda} + \Gamma^{\lambda}_{\mu\nu}\dot{x}^{\mu}\dot{x}^{\nu} = 0 = \frac{d^2x^{\lambda}}{d\lambda^2} + \Gamma^{\lambda}_{\mu\nu}\frac{dx^{\mu}}{d\lambda}\frac{dx^{\nu}}{d\lambda}, \tag{4.18}$$

where we have explicitly unpacked the dot notation for clarity (see *e.g.* Hobson et al., 2006, for comprehensive details). Note that if, and only if, $\Gamma^{\lambda}_{\mu\nu} = 0$ particles will travel along straight paths, as the curvature is null. However, if $\Gamma^{\lambda}_{\mu\nu} \neq 0$ then particles travel along curved paths over a given space-time manifold.

4.1.4 Lensing in the conformal Newtonian gauge

The gauge typically adopted in the gravitational lensing setting is that of the *Newtonian gauge* (Bartelmann, 2010) which is defined by the choice of gauge $B = E = 0$ for the Bardeen potentials, defined in equation 4.15. This choice is primarily for simplicity, as the metric $g_{\mu\nu}$ is diagonal which leads to straightforward calculation of the geodesic equations (see *e.g.* Ma and Bertschinger, 1995). Perhaps more crucially, within the conformal Newtonian gauge Ψ is (in the Newtonian limit) the gravitational potential which provides a simple physical interpretation. Immediately this choice of gauge implies the weakly perturbed FRLW metric, see equation 4.9, in the Newtonian gauge is given by

$$ds^2 = a^2(\tau)\left[(1+2\Psi)d\tau^2 - (1-2\Phi)\delta_{ij}dx^i dx^j\right], \tag{4.19}$$

from which one can straightforwardly read off the metric. It is interesting to note that, in the absence of anisotropic stress, the Bardeen potentials $\Psi = \Phi$ are equivalent, hence this form of the metric is approximately the weak-field limit of Minkowski space, in which Ψ acts as the gravitational potential (Challinor, 2009).

Consider now the four-momentum of a photon with space-like magnitude p, given by $P^{\mu} = \partial_{\lambda}x^{\mu}$ and $p^2 = g_{ij}P^iP^j$ respectively. By definition, photons are null and so have vanishing four-momentum, *i.e.* $P^{\mu}P_{\mu} = 0$. Within the Newtonian gauge expressed in equation 4.19, this implies that the time-like component $\mu = 0$ of the photon momentum is given by

$$\Rightarrow P^0 = \frac{p}{a}(1+2\Psi)^{-1/2} = \frac{p}{a}\left[1 - \Psi + \frac{3}{2}\Psi^2 + \mathcal{O}(\Psi^3)\right] \approx \frac{p}{a}(1-\Psi), \tag{4.20}$$

at linear order for small perturbations in the potential Ψ. Considering the space-like components $\mu \in [1,3]$ of the geodesic equation, given by equation 4.18, under the transformation of variables $x^i = r\theta^i$ one finds

$$\begin{aligned}
\frac{d^2x^i}{d\lambda^2} &= -\Gamma^i_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}, \\
\frac{dt}{d\lambda}\frac{dr}{dt}\frac{d}{dr}\left(\frac{d(r\theta^i)}{dr}\frac{dr}{dt}\frac{dt}{d\lambda}\right) &= -\Gamma^i_{\mu\nu}\frac{dx^\mu}{dr}\frac{dx^\nu}{dr}\left(\frac{dt}{d\lambda}\right)^2\left(\frac{dr}{dt}\right)^2, \\
\frac{d}{dr}\left(\frac{d(r\theta^i)}{dr}\frac{dr}{dt}\frac{dt}{d\lambda}\right) &= -\left[\Gamma^i_{00}\left(\frac{dx^0}{dr}\right)^2 + 2\Gamma^i_{0j}\frac{dx^0}{dr}\frac{dx^j}{dr} + \Gamma^i_{jk}\frac{dx^j}{dr}\frac{dx^k}{dr}\right]\frac{dt}{d\lambda}\frac{dr}{dt},
\end{aligned} \tag{4.21}$$

where the second line is a straightforward application of the chain rule. In the Newtonian gauge, the Christoffel symbols are given at linear order by

$$\begin{aligned}
\Gamma^0_{00} &= \mathcal{H} + \dot{\Psi} & \Gamma^i_{00} &= \partial^i\Psi \\
\Gamma^i_{j0} &= \delta^i_j(\mathcal{H} - \dot{\Phi}) & \Gamma^i_{jk} &= \delta^{im}\delta_{jk}\partial_m\Phi - 2\delta^i_{(j}\partial_{k)}\Phi.
\end{aligned}$$

Substituting these expressions for the Christoffel symbols into equation 4.21 allows us to explicitly evaluate the right hand term in the final line of equation 4.21 giving us

$$\left[\partial^i\Psi\left(\frac{dt}{dr}\right)^2 + 2\delta^i_j(\mathcal{H} - \dot{\Phi})\frac{dt}{dr}\frac{dx^j}{dr} + \left(\delta^{im}\delta_{jk}\partial_m\Phi - 2\delta^i_{(j}\partial_{k)}\Phi\right)\frac{dx^j}{dr}\frac{dx^k}{dr}\right]\frac{dt}{d\lambda}\frac{dr}{dt}. \tag{4.22}$$

This expression is deceptively concise, due to the efficiency of Einstein notation, and in actuality is describing a substantial number of terms. However, the only component which is non-negligible is the component along the line of sight, *i.e.* $j = k = 3$, hence this expression is well approximated to linear order by

$$\frac{p}{a}(1 - \Psi)\left[a^2\partial^i\Psi + 2\mathcal{H}\frac{dx^i}{dr} + \delta^{im}\partial_m\Phi\right], \tag{4.23}$$

which, after again applying a transformation of variables $x^i = r\theta^i$ and noticing that the late Universe is matter dominated, *i.e.* $p \propto a^{-1}$, allows us to rewrite the final line of equation 4.21 as the set of differential equations

$$\frac{d}{dr}\left(\frac{1}{a^2}\frac{d(r\theta^i)}{dr}\right) \approx \left[-\partial^i(\Psi + \Phi) - \frac{2\mathcal{H}}{a^2}\frac{d(r\theta^i)}{dr}\right]. \tag{4.24}$$

This set of differential equations can be trivialized by expanding the bracketed term on the left hand side, doing which cancels terms on the right hand side, resulting in

$$\frac{d^2(r\theta^i)}{dr^2} \approx -\delta^{im}\partial_m(\Psi + \Phi). \tag{4.25}$$

Integrating equation 4.25, relabelling $m \mapsto j$, and inverting the order of integration gives us our desired equation

$$\theta^i(r) = \theta^i(0) - \delta^{ij} \int_0^r d\tilde{r}\, \partial_j [\Psi + \Phi] \, [\frac{r - \tilde{r}}{r}], \tag{4.26}$$

which describes the deviation in observed angle on the sky of a point source by the intervening gravitational potential as it travels to us here and now. However, structural information is encoded into collections of photons, *e.g.* images of distant galaxies. As such, it is necessary to consider collections of deviations in angular position of a source image defined on the 2-sphere. In effect we wish to differentiate under the integrand of equation 4.26. With substantial algebraic wrangling, assuming anisotropic stress is minimal $\Psi = \Phi$, and adopting Born's approximation (see *e.g.* Bernardeau et al., 2010; Kitching et al., 2017), one may convert equation 4.26 into the form

$$\beta^i = \theta^i \underbrace{- 2 \int_0^r d\tilde{r}\, \partial^i \Phi(f_k(\tilde{r})\theta^i, \tilde{r}) \frac{f_K(r - \tilde{r})}{f_K(r)}}_{\alpha^i(\theta_i)}, \tag{4.27}$$

which is written simply as $\beta^i = \theta^i + \alpha^i$, and is colloquially referred to as the *lens equation*. Further to this, introducing angular co-ordinate differentials $\partial_i = f_K^{-1}(r)\partial_\theta$ allows one to write down an expression for an effective *lensing potential*

$$\phi(\theta^i) = 2 \int_0^r d\tilde{r}\, \frac{f_K(r - \tilde{r})}{f_K(r) f_K(\tilde{r})} \Phi(f_k(\tilde{r})\theta^i, \tilde{r}). \tag{4.28}$$

Now, define the Jacobi matrix $\mathcal{J}^i_j = (\partial\beta^i / \partial\theta^j)$ permitting one to map linear order perturbations of β^i and α^i onto interpretable geometric quantities such that

$$\mathcal{J} \equiv \begin{bmatrix} 1 - \kappa - \gamma_1 & \gamma_2 \\ \gamma_2 & 1 - \kappa + \gamma_1 \end{bmatrix} = \begin{bmatrix} 1 - \kappa & 0 \\ 0 & 1 - \kappa \end{bmatrix} - \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{bmatrix}, \tag{4.29}$$

where the convergence κ is a spin-0 scalar field which represents an isotropic stretching (magnification) of the source image, and is formally defined to be

$$\kappa = \frac{1}{2}\partial^i \partial_i \phi = \frac{1}{2}\nabla^2 \phi, \tag{4.30}$$

and the shear quantities γ_1, γ_2 are the real and complex components of the spin-2 vector field respectively, representing a perturbation of the third-flattening (ellipticity) of a source galaxy, and is formally defined to be

$$\gamma_1 = \frac{1}{2}(\phi^1_1 - \phi^2_2), \qquad \gamma_2 = \phi^1_2 = \phi^2_1, \qquad \gamma = \gamma_1 + i\gamma_2. \tag{4.31}$$

It is important to note that the local Newtonian potential Φ in equation 4.28 must necessarily satisfy the Poisson equation, such that (Bartelmann and Schneider, 2001)

$$\begin{aligned}\nabla^2\Phi(r,\omega) &= \frac{3\Omega_M H_0^2}{2a(r)}\delta(r,\omega),\\ \Rightarrow \kappa = \frac{1}{2}\nabla^2\phi &= \partial^i\partial_j\int_0^r d\tilde{r}\,\frac{f_K(r-\tilde{r})}{f_K(r)f_K(\tilde{r})}\Phi(f_k(\tilde{r})\theta^i,\tilde{r}),\end{aligned} \tag{4.32}$$

$$\Rightarrow \kappa = \frac{3\Omega_M H_0^2}{2c^2}\int_0^r dr'\,\frac{f_K(r')f_K(r-r')}{f_K(r)}\frac{\delta(f_K(r')r',r')}{a(r)}. \tag{4.33}$$

The overwhelming majority of weak lensing literature only considers the scalar contribution, and it is assumed that the potential is highly non-relativistic $\phi \ll c^2$. Even in this restricted setting it can be shown (Grimm and Yoo, 2018; Yoo et al., 2018) that the forms of κ and γ presented in equations 4.33 and 4.31 are gauge-dependent and so necessarily not observable. Nonetheless, within the scope of this thesis we will consider expressions 4.30 and 4.31 to be adequately precise.

4.2 Dark matter inference

In actuality, from equation 4.29 one may notice that the physically observable quantity for gravitational lensing is a circular source of unit radius, referred to as the *reduced shear*, g which is defined by

$$g = \frac{\gamma}{1-\kappa}, \tag{4.34}$$

which raises an interesting issue. Consider the general transformation of the lensing potential, defined in equation 4.28, given by

$$\phi(\theta^i) \rightarrow \phi'(\theta^i) = \frac{1-\lambda}{2}\delta_{ij}\theta^i\theta^j + \lambda\phi(\theta^i), \tag{4.35}$$

under which the convergence κ is transformed by (Bartelmann, 2010) $\kappa(\theta^i) \rightarrow \kappa'(\theta^i) = \lambda\kappa(\theta^i) + (1-\lambda)$, and the shear trivially transformed as $\gamma(\theta^i) \rightarrow \lambda\gamma(\theta^i)$. Under such a generalized transform the reduced shear g is invariant, however both κ and γ are not invariant and therefore suffer from a degeneracy of the form of the transform (see *e.g.* Gorenstein et al., 1988; Bradač et al., 2004). For weakly lensed sources the shear $|\gamma| \approx |g| \ll 1$ and so the shear is conserved under the transformation. Thus, in the weak lensing setting, γ is, in effect, an observable quantity. In all cases the convergence κ is not an observable quantity as the underlying intrinsic magnification is an *a priori* unknowable of the lensing system. Conversely, the intrinsic shearing field can be modelled to sufficient accuracy (Troxel and Ishak, 2015). In practice, measurements of the shear γ are typically recovered and an inverse problem can be drawn to construct estimators for κ from γ.

Individual measurements of γ are overwhelmingly dominated by the intrinsic ellipticity $\gamma^{\text{Intrinsic}}$, hence observational data is often binned into pixels and the measurements within a given pixel are averaged (Bartelmann and Schneider, 2001). If the number of measurements within a pixel is large the residual zero-mean intrinsic value $\langle \gamma^{\text{Intrinsic}} \rangle \sim 0$. In contrast to this, the observable shear value has non-zero mean and thus is revealed, *i.e.*

$$\gamma = \langle \gamma^{\text{Lensing}} + \gamma^{\text{Intrinsic}} \rangle \approx \gamma^{\text{Lensing}} \tag{4.36}$$

However, for typical surveys only a limited number of galaxies can be observed within a given pixel. Upcoming stage IV survey Euclid forecasts ~ 30 gals/arcmin^2, a count which is also forecast to be a rough upper limit due to blending complications at high-redshift (see *e.g.* Chang et al., 2013).

4.2.1 Spin operators and the flat-sky

Often it is preferable to work directly on the sphere and so we require definitions of the physical quantities ${}_0\kappa$ and ${}_2\gamma$ in terms of spherical mapping operators, which will simplify the projection onto the spin-spherical harmonic basis functions. Note that in the following sections we will explicitly denote a spin-s field ${}_sx$: $s = 0$ denotes a scalar field which is SO(2)[1] invariant, and $s = 2$ denotes a field which under the group of SO(2) rotations transforms as $x \mapsto e^{-2i\psi}x$ for $\psi \in \mathbb{R}$. Consider the spin raising and lowering operators (Newman and Penrose, 1966; Goldberg et al., 1967):

$$\eth^{\pm} \equiv -\sin^{\pm s}\theta \Big(\partial_\theta \pm \frac{i}{\sin\theta}\partial_\psi\Big)\sin^{\mp s}\theta, \tag{4.37}$$

where $\omega = (\theta, \psi) \in \mathbb{S}^2$ are angular co-ordinates on the 2-sphere, and where $\pm$ denote raising and lowering respectively. Equations 4.30 and 4.31 can be recast to form (Bunn et al., 2003)

$${}_0\kappa(r,\omega) = \frac{1}{4}(\eth^+\eth^- + \eth^-\eth^+)\,\phi(r,\omega), \qquad {}_2\gamma(r,\omega) = \frac{1}{2}\eth^+\eth^+\,\phi(r,\omega), \tag{4.38}$$

where the $\eth^{\pm}$ differential operators do not necessarily commute. This recasting drastically simplifies extension to the spherical case in later sections, and provides a more intuitive understanding of the fundamental relationship between ${}_0\kappa$ and ${}_2\gamma$, which will be a core concept revisited often through this thesis. For small fields of view the sphere is, to a good approximation, flat. This allows computationally taxing spherical harmonic transforms to be replaced by Fourier transforms which have significantly lower complexity – as at high frequencies the harmonic and Fourier coefficients of a signal become coincident. This small-field simplification is colloquially referred to as the *flat-sky approximation*, and analysis within this approximation is often referred to as weak lensing on the plane.

On the plane, the spin raising and lowering operators reduce by first fixing the

---

[1] SO(2) denotes the special orthogonal group which has an irreducible representation on the sphere as rotations on the local tangent plane.

basis vector perpendicular to the tangent plane, varying θ and ψ, and exploiting the chain rule to find $\partial_\theta \propto -\partial_x$ and $\partial_\psi \propto \partial_y$. Transforming the differentials in equation 4.37 reduces the spin raising and lowering operators on the tangent plane to (Bunn et al., 2003; Wallis et al., 2021)

$$\eth^{\pm} \equiv -\sin^{\pm s}\theta \Big(\partial_\theta \pm \frac{i}{\sin\theta}\partial_\psi \Big) \sin^{\mp s}\theta \approx -\Big(\partial_x \pm i\partial_y \Big), \tag{4.39}$$

4.2.2 Inference over a Euclidean tangent plane

From equation 4.38, exploiting the approximation given in equation 4.39, one can derive the lensing forward model on the plane, which is a mapping relation between the unobservable convergence field ${}_0\kappa$ and the observable shear field ${}_2\gamma$ straightforwardly as follows

$$\begin{aligned}
{}_0\kappa(r,\omega) &= \frac{1}{4}(\eth^+\eth^- + \eth^-\eth^+)\,{}_0\phi(r,\omega) & {}_2\gamma(r,\omega) &= \frac{1}{2}\eth^+\eth^+\,{}_0\phi(r,\omega),\\
{}_0\kappa(r,\omega) &= \frac{1}{2}\left[\partial_{xx} + \partial_{yy}\right]{}_0\phi(r,\omega) & {}_2\gamma(r,\omega) &= \frac{1}{2}\left[\partial_{xx} - \partial_{yy} + 2i\partial xy\right]\,{}_0\phi(r,\omega),\\
\mathsf{F}\left[{}_0\kappa\right] \Rightarrow {}_0\tilde{\kappa}_{k_x,k_y} &= \frac{1}{2}\left[k_x^2 + k_y^2\right]\,\tilde{\phi}_{k_x,k_y} & \mathsf{F}\left[{}_2\gamma\right] \Rightarrow {}_2\tilde{\gamma}_{k_x,k_y} &= \frac{1}{2}\left[k_x^2 - k_y^2 + 2ik_xk_y\right]\,{}_0\tilde{\phi}_{k_x,k_y},
\end{aligned} \tag{4.40}$$

where F represents the usual Fourier transforms, tilde is used for Fourier coefficients and the shorthand $\partial_{xx} = \partial_x\partial_x$ has been used. Note that in line 3 the real-space arguments have been explicitly dropped for clarity. Clearly one can now substitute for ${}_0\tilde{\phi}_{k_x,k_y}$ to draw a Fourier space relation between the convergence field and the shear field such that,

$${}_2\gamma(r,\omega) = \mathsf{F}^{-1}\left[{}_2\tilde{\gamma}_{k_x,k_y}\right] = \mathsf{F}^{-1}\left[\underbrace{\left(\frac{k_x^2 - k_y^2 + 2ik_xk_y}{k_x^2 + k_y^2} \right)}_{\equiv \mathsf{D}_{k_x,k_y}} \mathsf{F}\left[{}_0\kappa(r,\omega)\right] \right] = \mathsf{F}^{-1}\,\mathsf{D}\,\mathsf{F}\,{}_0\kappa(r,\omega), \tag{4.41}$$

where this is our final expression, and is the planar forward model for weak lensing convergence reconstruction (Kaiser and Squires, 1993) (see chapter 2 for a discussion of forward models). This forward model is a good approximation of the full spherical forward model (derived below) when the sky-fraction considered is small, *i.e.* small-field surveys, but can be shown (Kitching et al., 2017; Wallis et al., 2021) to rapidly degrade for wide-fields. Fundamentally, spherical manifolds have non-zero intrinsic curvature, a result from *Gauss's Theorema Egregium* which is a fundamental theory of differential geometry.

4.2.3 Inference over the celestial sphere

To work with spin-s signals on the sphere, for $s \in \mathbb{Z}$, it is necessary to first define the action of the spin-s raising and lowering operators, given in equation 4.37, on the spherical harmonic coefficients (Newman and Penrose, 1966; Goldberg et al., 1967; Hu, 2000)

$$ð^{\pm}\,{}_sY_{\ell m}(\omega) = \pm\big[(\ell \mp s)(\ell \pm s + 1)\big]^{1/2}\,{}_{s\pm 1}Y_{\ell m}(\omega), \tag{4.42}$$

where $\ell \in \mathbb{Z}_{\geq 0}$ and $m \in \mathbb{Z}_{\leq \ell}$ are the standard spherical harmonic indices. It is then abundantly clear that any spin-s function can be written as s repeated applications of $ð^{\pm}$ such that we have the recursive relation (see *e.g.* McEwen et al., 2013a, 2015a; Wallis et al., 2017; Wallis et al., 2021)

$${}_sY_{\ell m}(\omega) = (-1)^{\min(s,0)}\left[\frac{(\ell \mp s)!}{(\ell \pm s)!}\right]^{\frac{1}{2}}[ð^{\pm}]^{\pm s}Y_{\ell m}(\omega), \tag{4.43}$$

With these tools we are now ready to derive the spherical relation. As with the planar derivation, let us derive the spherical relationship between the convergence field κ, the shear field γ and the lensing potential ϕ given in equation 4.38. First derive the harmonic relation between γ and ϕ,

$$\begin{aligned}
{}_2\hat{\gamma}_{\ell m} = \langle {}_2\gamma | {}_2Y_{\ell m}\rangle_{\mathbb{S}^2} &\equiv \int_{\mathbb{S}^2} d\Omega(\omega)\,{}_2\gamma\,{}_2Y^*_{\ell m} = \frac{1}{2}\int_{\mathbb{S}^2} d\Omega(\omega)\,ð^+ð^+\,{}_0\phi\,{}_2Y^*_{\ell m} \\
&= \frac{1}{2}\int_{\mathbb{S}^2} d\Omega(\omega)\,ð^+ð^+\,{}_2Y_{-\ell m}\,{}_0\phi = \frac{1}{2}\int_{\mathbb{S}^2} d\Omega(\omega)\left[\frac{(\ell+2)!}{(\ell-2)!}\right]^{\frac{1}{2}}{}_0\phi\,{}_0Y^*_{\ell m} \\
\Rightarrow {}_2\hat{\gamma}_{\ell m} &= \frac{1}{2}\left[\frac{(\ell+2)!}{(\ell-2)!}\right]^{\frac{1}{2}}{}_0\hat{\phi}_{\ell m},
\end{aligned} \tag{4.44}$$

where $d\Omega(\omega) = \sin\theta d\theta d\phi$ is the rotation invariant measure on the sphere, the Haar measure of $\mathbb{S}^2$, $\langle \cdot,\cdot \rangle_{\mathbb{S}^2}$ denotes the inner product over $\mathbb{S}^2$, and where we have dropped real-space arguments after line 1 for clarity. Notice that in line 2 we have used the conjugate relation for spherical harmonics, and exploited the recursive relations given by equation 4.43. Now let us similarly derive the harmonic space relation between the spin-0 convergence ${}_0\kappa$ and the lensing potential ${}_0\phi$,

$$\begin{aligned}
{}_0\hat{\kappa}_{\ell m} = \langle {}_0\kappa | {}_0Y_{\ell m}\rangle_{\mathbb{S}^2} &\equiv \int_{\mathbb{S}^2} d\Omega(\omega)\,{}_0\kappa\,{}_0Y^*_{\ell m} = \frac{1}{4}\int_{\mathbb{S}^2} d\Omega(\omega)\left[ð^+ð^- + ð^-ð^+\right]{}_0\phi\,{}_0Y^*_{\ell m} \\
&= \frac{1}{4}\int_{\mathbb{S}^2} d\Omega(\omega)\left[ð^+ð^- + ð^-ð^+\right]{}_0Y_{-\ell m}\,{}_0\phi \\
&= \frac{1}{4}\int_{\mathbb{S}^2} d\Omega(\omega)\left[ð^+ð^-\,{}_0Y_{-\ell m} + ð^-ð^+\,{}_0Y_{-\ell m}\right]{}_0\phi \\
&= -\frac{1}{2}\int_{\mathbb{S}^2} d\Omega(\omega)\left[\ell(\ell+1)\right]{}_0\phi\,{}_0Y^*_{\ell m}, \\
\Rightarrow {}_0\hat{\kappa}_{\ell m} &= -\frac{1}{2}\left[\ell(\ell+1)\right]{}_0\hat{\phi}_{\ell m},
\end{aligned} \tag{4.45}$$

where lines 1-3 are straightforward algebraic wrangling, and the penultimate line exploits the recursive relations provided in equation 4.43. With spherical harmonic space relations 4.44 and 4.45 in mind we can now explicitly state the spherical weak lensing forward model (Wallis et al., 2021)

$$ {}_2\hat{\gamma}_{\ell m} = \frac{-1}{\ell(\ell+1)}\left[\frac{(\ell+2)!}{(\ell-2)!}\right]^{\frac{1}{2}} {}_0\hat{\kappa}_{\ell m} = \mathsf{W}_{\ell 0}\hat{\kappa}_{\ell m}, \tag{4.46} $$

where W_ℓ is the harmonic space weighting kernel. When projected into real-space, this relation is given by

$$ {}_2\gamma(r,\omega) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} {}_2\hat{\gamma}_{\ell m}\, {}_2Y_{\ell m}(\omega) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} \frac{-1}{\ell(\ell+1)}\left[\frac{(\ell+2)!}{(\ell-2)!}\right]^{\frac{1}{2}} {}_0\hat{\kappa}_{\ell m}\, {}_0Y_{\ell m}(\omega), \tag{4.47} $$

where the first equality is the discretized inverse spin-2 spherical harmonic transform with Eigen-functions ${}_2\tilde{Y}_{\ell m}$ (Newman and Penrose, 1966; Goldberg et al., 1967), ${}_0Y_{\ell m}$ is the standard forward spherical harmonic transform. Introducing ${}_s\mathsf{Y}, {}_s\mathsf{Y}^{-1}$, defined to be operators which perform the forward/inverse spin-s spherical harmonic transforms respectively, one finds the spherical forward-model in operator form

$$ {}_2\gamma(r,\omega) = {}_2\mathsf{Y}^{-1}\,\mathsf{W}_0\mathsf{Y}\,{}_0\kappa(r,\omega). \tag{4.48} $$

Notice that both the planar relation 4.41 and the spherical relation 4.48 describe an *a priori* known mapping between a desired field of interest, in this case the convergence ${}_0\kappa$, and an albeit approximately observable field ${}_2\gamma$. As discussed in a general sense in chapter 2 and a probabilistic sense in chapter 3 such mappings represent the forward model, and can be used to construct inverse (inference) problems through which we can extract information pertaining to ${}_0\kappa$, in light of observations of ${}_2\gamma$. In the following subsection we will discuss how researchers leverage such methods to infer the *a priori* unobservable convergence, and by proxy the line of sight total matter distribution, from weak gravitational lensing observations. A hemispherical plot of ${}_0\kappa$ is displayed in Figure 4.2.

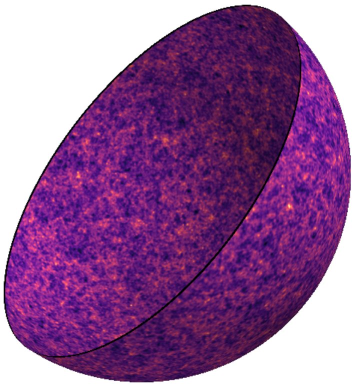


Figure 4.2: Hemispherical plot of the convergence at $z = 1$, extracted from large scale N-body simulations (see Takahashi et al., 2017).

4.2.4 Dark matter mass reconstruction techniques

Knowing now how both the convergence κ and shearing γ fields emerge from cosmological theory, and how one relates these quantities *via* the weak lensing forward model ϕ, we will now discuss how one leverages noisy, and often incomplete, observations of the shearing field to infer the convergence field. Such inferences, colloquially referred to as *dark matter mass-maps*, belong to the set of principle cosmological observables (Clowe et al., 2006), and will occupy much of the discussion within this thesis.

Inferences in the weak lensing domain are described by ill-posed, and often severely ill-conditioned, inverse problems (Hadamard, 1902) which take the general form of a noisy axisymmetric deconvolution problem with in-painting dependent on masking geometries. A substantial set of solutions to such problems exists throughout, *e.g.*, the image processing community, however only a comparatively small number have been applied in the weak gravitational setting. In this short subsection we will provide an overview of the main approaches currently adopted throughout the mass-mapping scientific community[2], although this brief summary is not intended to be comprehensive.

Uniform prior: Kaiser-Squires

Perhaps the most naive solution to the weak lensing inverse problem is simply to directly invert the forward model in Fourier space such that

$$\hat{\kappa} = \mathsf{D}^{-1}\hat{\gamma} \qquad\qquad \tilde{\kappa}_{\ell m} = \mathcal{W}_{\ell}^{-1}\tilde{\gamma}_{\ell m} \tag{4.49}$$

on the plane and celestial sphere respectively. These methods are referred to as Kaiser-Squires (KS) inversion (Kaiser and Squires, 1993) and Spherical Kaiser-Squires (SKS) inversion (Wallis et al., 2021) respectively. Note that $\hat{\cdot}$ represents Fourier coefficients, $\tilde{\cdot}$ represents spherical harmonic coefficients, D represents the planar Fourier relation given by equation 4.41 and $\mathcal{W}_\ell$ represents the spherical harmonic relation given by equation 4.48. Decomposition of spin-signals on bounded manifolds is known to be degenerate (Bunn et al., 2003) due to local loss of orthogonality between basis eigenfunctions. Therefore, in the presence of non-trivial masking geometries, which is more often than not the case, these estimators are patently sub-optimal.

Moreover, as these methods do not directly consider noise present in γ (a combination of instrumental noise and residual intrinsic ellipticity) a subsequent post-processing of the maps is undertaken; a large Gaussian smoothing kernel is convolved with the initial estimate of κ. For large spatial scales (at which the Universe is to a good approximation Gaussian) this convolution is somewhat well informed, though for the small scales a non-negligible fraction of the information content is inherently non-Gaussian (Taylor et al., 2018). Thus the information content at such scales is severely degraded by this convolution.

---

[2] The set of algorithms discussed in the following sections is far from exhaustive, however an example from each train of thought is included to broaden the discussion. Furthermore, the algorithms selected represent novel contributions upon which subsequently developed algorithms are largely derivative.

Gaussian family priors: Wiener Filtering

The Wiener filter is a linear filtering technique, often solved through Bayesian sampling methods or less commonly through optimisation methods (see chapters 2 and 3) which is explicitly restricted to Gaussian family posterior distributions (Teyssier et al., 2009; Jeffrey et al., 2018). Such posteriors are typically constructed from Gaussian distributed likelihoods and weakly informative conjugate priors. Consider noise covariance N, signal covariance S and forward model D, in such a case the likelihood and prior functions are given by

$$P(\gamma|\kappa,\mathsf{D}) = \frac{1}{\sqrt{2\pi \mathrm{det}(N)}} \exp\Big[-\frac{1}{2}(\gamma-\mathsf{D}\kappa)^\dagger N^{-1}(\gamma-\mathsf{D}\kappa)\Big] \sim \mathcal{N}(0,N) \tag{4.50}$$

$$P(\kappa) = \frac{1}{\sqrt{2\pi \mathrm{det}(S)}} \exp\Big[-\frac{\kappa^\dagger S^{-1}\kappa}{2}\Big] \sim \mathcal{N}(0,S), \tag{4.51}$$

from which the full posterior can be constructed as

$$P(\kappa|\gamma,\mathsf{D}) \propto \exp\Big[-\frac{(\kappa-\mathsf{W}\gamma)^\dagger(S^{-1}+\mathsf{D}N^{-1}\mathsf{D}^\dagger)(\kappa-\mathsf{W}\gamma)}{2}\Big], \tag{4.52}$$

where the operator W is the Wiener filter (Wiener et al., 1964) defined by $\mathsf{W} = S\mathsf{D}^\dagger[\mathsf{D}S\mathsf{D}^\dagger + N]^{-1}$. A more comprehensive discussion of the Wiener filter as a Bayesian method can be found earlier in this thesis, specifically in section 3.1.3.

This posterior is often approximated *via* Markov chain Monte Carlo (MCMC) sampling methods which naturally recovers the complete posterior distribution, from which one can compute 'optimal' solutions with associated credible regions. However, as the Wiener posterior is straightforwardly log-concave, one may also rapidly minimize the negative log-posterior *via* convex optimisation techniques, before leveraging *e.g.* the Laplace approximation (see *e.g.* Laplace, 1986) for posterior uncertainties. Bayesian sampling methods, in which the Wiener filter is most often computed, support principled uncertainty of κ (computed from the posterior samples), however there are two primary disadvantages of such approaches.

Currently, MCMC sampling algorithms are typically highly computationally demanding, in many cases prohibitively so, as they typically require very many posterior samples for convergence. Hence, sampling methods are typically restricted to low-dimensional settings and computationally tractable posterior distributions[3]. Additionally, though assumptions of Gaussianity are well-motivated in many cases, *e.g.*, for cosmic microwave background (CMB) research (see *e.g.* Horowitz et al., 2019; Kodi Ramanah et al., 2019)[4], they are inherently sub-optimal for weak lensing.

---

[3] To some small extend this motivates the choice of Gaussian forms as they lead to several computational speed-ups. In some cases such posteriors can be solved analytically, though this can involve the inversion of prohibitively large covariance matrices.

[4] The CMB is extremely close to Gaussian, thus validating Gaussian type likelihood and prior

Weak lensing is primarily set in the late universe, as such the matter (both visible and dark) distribution has undergone highly non-linear evolution under gravity, at least on the spatial scales of interest. As the convergence is inherently non-Gaussian, prior assumptions of Gaussianity are ill-informed and severely degrade the quality of weak lensing information.

Sparse regularisation: GLIMPSE

In contrast to the methods discussed so far GLIMPSE[5] is a more classical image-processing reconstruction algorithm which draws on proximal convex optimisation methods (see section 2.2), sparsity and wavelets to construct remarkably high-resolution point estimates of the convergence field (Leonard et al., 2014; Lanusse et al., 2016). However, GLIMPSE is not posed in a framework which supports principled statistical uncertainties on reconstructed convergence maps. Such uncertainty quantifications are a fundamental necessity when using these high-quality maps for cosmological inference, hence, despite producing excellent point estimate convergence reconstructions, cosmological inference is somewhat less statistically principled than *e.g.* sampling methods.

The GLIMPSE algorithm considers a general linear inverse problem of the form

$$\gamma = \Phi\kappa + n, \tag{4.53}$$

for shear observations $\gamma \in \mathbb{C}^N$, which are generated from an underlying convergence $\kappa \in \mathbb{C}^M$ by the forward model $\Phi \in \mathbb{C}^{M\times N} : \kappa \mapsto \gamma$, and under additive Gaussian noise $n \sim \mathcal{N}(0, \Sigma_\gamma \in \mathbb{C}^N)$. Note that in many weak lensing settings the noise n is at least an order of magnitude larger than the signal γ and $N \ll M$, hence such inverse problems are often seriously ill-posed and or ill-conditioned (Hadamard, 1902) (see section 2.1.2 for a more substantive discussion). In the GLIMPSE setting the measurement operator, the operator through which observations or measurements are modelled, is given by

$$\Phi = \mathsf{TDF}, \tag{4.54}$$

for standard Fourier transform F, the aforementioned planar forward relation in Fourier space D and the inverse NDFT (non-uniform Fourier transform) matrix T (Keiner et al., 2009). Two regularisation terms are introduced which encode sparsity in the starlet wavelet dictionary (Starck, J.-L. et al., 2006a,b) and reality of the convergence field, such that the Lagrangian dual type optimisation problem (see section 2.2) in the synthesis setting is given by

$$\kappa^{\text{optimal}} = \underset{\kappa}{\operatorname{argmin}} \left[\Sigma_\gamma^{-\frac{1}{2}} [\gamma - \Phi\kappa] + \lambda \|\omega \circ \Psi^* \kappa\|_1 + \mathbb{I}_{\Im(\cdot)=0}(\kappa) \right], \tag{4.55}$$

---

terms.

[5] http://www.cosmostat.org/software/glimpse

where λ is a regularisation parameter which controls the weighting between the two terms, Σ_γ is the shear covariance, $\mathbb{I}_{\Im(\cdot)=0}$ is the indicator function for vanishing imaginary convergence, ω is an adaptive re-weighting recalculated after each step in the algorithm and $\circ$ is the Hadamard product (element-wise product).

Refinements of the GLIMPSE algorithm have been introduced, *i.e.* the inclusion of higher-order lensing observables such as the spin-3 flexion, though for brevity we will truncate our discussion here. The reader is pointed to the original articles for a far more detailed overview of this algorithm (see *e.g.* Leonard et al., 2014; Lanusse et al., 2016). As GLIMPSE does not introduce, either implicitly or explicitly, an internal bias towards Gaussian signals[6] it is currently the algorithm of choice for high-resolution mass-mapping – in which signal information content is substantially non-Gaussian. In particular, the algorithm is an effective means by which to recover non-parametric reconstructions of dark matter haloes. The algorithm's efficacy on larger scales is less well constrained, and depends highly on the statistical metric with which one determines performance, *i.e.* GLIMPSE recovers far better estimates of the peak statistic (Jain and Van Waerbeke, 2000) than the power-spectrum (see *e.g.* Jeffrey et al., 2018).

Furthermore, the selection of the regularisation (hyper-) parameters λ in the objective is somewhat ill-motivated. Through an iterative grid-search over simulated κ maps, the value of λ which maximizes the correlation between the GLIMPSE estimate of κ and the simulated maps is located. This (pseudo-) optimal value of λ is then fixed and assumed to perform well on real data applications. In a sense this can be thought of as an implicit empirical Bayes approach, in which the regularisation (hyper-) parameter λ is assigned an infinitely concentrated hyper-prior distribution, *e.g.* a delta function. In a, somewhat approximate, Bayesian sense this handling of model (hyper-) parameters is predicated on the notion that simulated κ maps are indistinguishable from the true convergence of the Universe, a claim to which the authors seem to attribute infinite plausibility. One might, quite reasonably, be uncertain as to the value of λ, and instead marginalize over this nuisance parameter, as is done in later chapters of this thesis (see section 3.1.4 for a compact discussion). Last but not least, the current implementation of the GLIMPSE algorithm is structurally restricted to reconstructions of planar patches and does not facilitate the wide-field spherical reconstructions required (Wallis et al., 2021) by stage IV surveys, *e.g.* Euclid (Laureijs et al., 2011).

---

[6] Sources of such Gaussian biases include *e.g.* post-processing smoothing (as in Kaiser-Squires) or strong Gaussian priors (as in Weiner filtering).

Deep learning: Convolutional neural networks

Over the past year deep machine learning techniques have been applied (Jeffrey et al., 2020), in an end-to-end fashion, in an attempt to 'learn' a generalized inverse mapping function between the observed shearing and desired convergence cosmological fields. In much the same way as the GLIMPSE algorithm the inverse problem forward model is given by $\gamma = \Phi\kappa + n$, for forward model Φ and additive Gaussian noise n. Where GLIMPSE (Lanusse et al., 2016) adopts approximate, though highly effective, wavelet priors to stabilize the reconstruction, deep learning approaches[7] instead attempt to learn the parameters Θ of a mapping function $\mathcal{F}_\Theta$ such that

$$\kappa = \mathcal{F}_\Theta(\gamma). \tag{4.56}$$

Such a learnt function is recovered by minimisation of a well defined loss function, *e.g.* the mean squared error (MSE). In theory $\mathcal{F}_\Theta$ can approximate inversion of the forward model with arbitrary precision, whilst implicitly introducing regularisation *via* choices made during the training of the network parameters Θ. For example, if one samples from a known prior distribution $P(\kappa)$, generating simulated data vectors $P(\gamma|\kappa)$, which are adopted during training of model parameters Θ under an MSE loss function, then $\mathcal{F}_{\Theta_{\text{opt}}}$ for optimized model parameters Θ_{opt} is an estimate of the posterior mean (Jaynes, 2003).

Though it is undeniably difficult to overstate the potential of such approaches, there are a few interesting caveats worth noting. Though a learnt mapping function $\mathcal{F}_\Theta$, under certain conditions, has a principled Bayesian interpretation it is unclear how the benefits of Bayesian methodology, *e.g.* exploration of the posterior space, and principled uncertainty quantification, can be fully supported[8]. Furthermore, the performance of such end-to-end deep learning techniques often generalizes poorly under re-specification of the inverse problem at hand, *i.e.* when the noise distribution or magnitude varies. Hence, end-to-end deep learning approaches must often be retrained from scratch under problem variable re-specification. For many applications this is simply an inconvenience, however for cosmological applications, wherein training data is generated through computationally expensive simulations, this becomes a major blocker. Nevertheless, deep learning approaches, perhaps more specifically hybrid deep learning approaches (see *e.g.* Adler and Öktem, 2018; Lunz et al., 2018), are certainly strong contenders for upcoming, now imminent, stage IV weak lensing surveys (*e.g.* Laureijs et al., 2011).

---

[7] This specifically relates to end-to-end deep learning approaches. There are, of course, myriad machine learning techniques each with (often) nuanced interpretations.

[8] Certainly this problem has yet to be solved in a satisfactory manner, and is already forming a highly interesting avenue for future research.

Part II

Doctoral research

Chapter 5

Hypothesis testing of dark matter

This chapter is based on research presented in M. A. Price, J. D. McEwen, X. Cai, T. D. Kitching, and C. G. R. Wallis, "Sparse Bayesian mass-mapping with uncertainties: hypothesis testing of structure", Monthly Notices of the Royal Astronomical Society, vol. 506, no. 3, pp. 3678-3690, September 2021.

Price et al. (2021a)

Note: *This chapter assumes knowledge of the following. Mathematical context of weak gravitational lensing, presented in sections 4.1.4 and 4.2.2. Convex optimisation techniques, specifically sections 2.2.3 and 2.2.4. Probabilistic inference techniques, discussed at length in chapter 3.*

In this chapter we present a new mass-mapping formalism. We formulate the lensing inverse problem as a sparse hierarchical Bayesian inference problem from which we derive an unconstrained convex optimisation problem. We solve this optimisation problem in the analysis setting, with a wavelet-based, sparsity-promoting, ℓ_1-norm prior: similar priors have been shown to be effective in the weak lensing setting (Leonard et al., 2014; Lanusse et al., 2016; Peel et al., 2017a; Jeffrey et al., 2018). Formulating the problem in this way allows us, for the first time, to recover *maximum a posteriori* (MAP) estimators (see section 3.2.2), from which we can exploit analytic methods (Pereyra, 2017; Cai et al., 2018b) to recover approximate highest posterior density (HPD) credible regions, and perform hypothesis testing of structure in a variety of ways (see section 3.2.3). We apply our algorithm to a range of catalogs drawn from Bolshoi N-body simulation cluster catalogues (Klypin et al., 2011), and the hotly debated A520 cluster catalogs (see *e.g.* Clowe et al., 2012; Jee et al., 2014). We then demonstrate the aforementioned uncertainty quantification techniques on our MAP reconstructions from these catalogs.

5.1 Introduction

Gravitational lensing is an astrophysical phenomenon, that can be observed on galactic to cosmic spatial scales, through which distant images are distorted by the intervening mass density field. Due to its nature, lensing is sensitive to the total mass distribution (both visible and invisible) along a line of sight (Bartelmann and

Schneider, 2001; Munshi et al., 2008; Heavens, 2009; Bartelmann, 2010). Therefore, as the majority of massive structures in the universe predominantly consist of dark matter, lensing provides a novel way to probe the nature of dark matter itself. Weak gravitational lensing (WL) is a regime in which one makes the approximation that lensed sources have (at no time) come radially closer than an Einstein radius to the intervening mass concentrations — which ensures that sources are not multiply imaged. The effect of weak lensing on distant source galaxies is two-fold: the galaxy size is magnified by a convergence field κ; and the galaxy ellipticity (third-flattening) is perturbed from an underlying intrinsic value by a shearing field γ (see section 4.1.4 for an extended discussion).

Due to the mass-sheet degeneracy the weak lensing convergence field is not directly observable (see section 4.2.4). In the weak lensing regime, the shearing field does not suffer such degeneracies and can accurately be modelled from observed ellipticities. Therefore, observations of γ are typically inverted to recover estimators of κ. Such estimators are colloquially named *dark matter mass-maps*, and constitute one of the principle observables for cosmology (Clowe et al., 2006). Standard cosmological protocol is to extract weak lensing information in the form of second order statistics (Kilbinger, 2015; Alsing et al., 2016; Taylor et al., 2018) which are then compared to theory. In this approach mass-maps are not required. However, as two-point global statistics are by definition sensitive only to Gaussian contributions, and weak lensing is inherently non-Gaussian, it is informative to consider higher-order statistics (Coles and Chiang, 2000; Munshi and Coles, 2017). Many higher-order statistical techniques can be performed directly on mass-maps (κ-fields), which motivates investigation into alternate mass-map reconstruction methodologies.

Reconstructing mass-maps from shear observations requires solving an (often seriously) ill-posed inverse problem (see section 4.2.4). Many approaches to solving this lensing inverse problem have been developed (*e.g.* Kaiser and Squires, 1993; VanderPlas et al., 2011; Lanusse et al., 2016; Chang et al., 2018; Jeffrey et al., 2018; Wallis et al., 2021), with the industry standard being Kaiser-Squires (KS; Kaiser and Squires, 1993). Although these approaches often produce reliable convergence estimators, they lack principled statistical approaches to uncertainty quantification and often assume Gaussianity during the reconstruction process, or post-process by Gaussian smoothing, which is sub-optimal when one wishes to analyze small-scale non-Gaussian structure. Most methods refrain from quantifying uncertainties in reconstructions, but those that do often do so by assuming Gaussian priors and adopting Markov-chain Monte-Carlo (MCMC) techniques (see *e.g.* Corless et al., 2009; Schneider et al., 2015; Alsing et al., 2016) – see section 3.2.1 for further discussion. The computational cost of MCMC approaches is typically excessive. Recent developments in probability concentration theory have led to advancements in fast approximate uncertainty quantification techniques (Pereyra, 2017; Cai et al., 2018a,b), which will be leveraged throughout this chapter (see section 3.2.3).

The structure of this chapter is as follows. In section 5.2 we provide the details of our algorithm, as well as some updates to super-resolution image recovery. In section 5.3 we present the uncertainty quantification techniques, both mathematically and mechanistically. In sections 5.4 and 5.5 we apply both our reconstruction algorithm and the uncertainty quantification techniques to the aforementioned datasets and analyze the results. Finally, in section 5.6 we draw conclusions from this work and propose future avenues of research. Section 5.2 relies on a moderate level of understanding in the fields of proximal calculus and compressed sensing. For the reader solely interested in practical application of these techniques we recommend sections 5.4 onwards.

5.2 Sparse convergence estimators

Several alternate approaches for solving the inverse problem between convergence κ and shear γ which do not assume or impose Gaussianity have been proposed, some of which are based on the concept of wavelets and sparsity (Pires et al., 2009; Jullo et al., 2014; Lanusse et al., 2016; Peel et al., 2017a). We propose a mass-mapping algorithm that relies on sparsity in a given wavelet dictionary (see sections 2.1.3 and 3.1.3). Moreover, we formulate the problem such that we can exploit recent developments in the theory of probability concentration, which have been developed further to produce novel uncertainty quantification techniques (Pereyra, 2017). Crucially, this allows us to recover principled statistical uncertainties on our MAP reconstructions (as in Cai et al., 2018a,b) as will be discussed in detail in the following section.

As discussed in section 4.2.4, galaxies have an intrinsic ellipticity. To mitigate the effect of intrinsic ellipticity we choose to project the ellipticity measurements onto a grid and average. If we assume that galaxies have no preferential orientation in the absence of lensing effects, then the average intrinsic ellipticity tends to zero. This is a good approximation for the purposes of this discussion, but weak correlation between the intrinsic alignments of galaxies has been observed (Troxel and Ishak, 2015; Piras et al., 2018).

5.2.1 Hierarchical Bayesian framework

Hierarchical Bayesian inference provides a rigorous mathematical framework through which theoretically optimal solutions can be recovered. Moreover it allows one to construct measures of the uncertainty on recovered point estimates. Look to chapter 3 for a detailed discussion, much of which will be assumed in the following analysis.

As is common for hierarchical Bayesian models, we begin from Bayes' theorem for the posterior distribution,

$$p(\kappa|\gamma) = \frac{p(\gamma|\kappa)p(\kappa)}{\int_{\mathbb{C}^N} p(\gamma|\kappa)p(\kappa)d\kappa}, \tag{5.1}$$

where $p(\gamma|\kappa)$ is the likelihood function representing data fidelity, N is the dimensionality of κ and $p(\kappa)$ is a prior on the statistical nature of κ. The denominator is called the *Bayesian evidence* which is constant and so can be dropped for our purposes. Typically the Bayesian evidence is used for model comparison (see *e.g.* Robert, 2001; Sivia and Skilling, 2006), which we will not be considering within the context of this discussion. Given Bayes' theorem, and the monotonicity of the logarithm function, we can easily show that the maximum posterior solution is defined by,

$$\underset{\kappa \in \mathbb{C}^N}{\operatorname{argmax}} \Big[p(\kappa|\gamma) \Big] = \underset{\kappa \in \mathbb{C}^N}{\operatorname{argmin}} \Big[-\log(\, p(\kappa|\gamma)\,) \Big], \tag{5.2}$$

as discussed in section 3.2.2. This step is crucial, as it allows us to solve the more straightforward problem of minimizing the log-posterior rather than maximizing the full posterior. Conveniently, in most physical situations the operators associated with the log-posterior are convex. Drawing from the field of convex optimisation, the optimal solution for the posterior can be recovered extremely quickly — even in high dimensional settings.

5.2.2 Sparsity and inverse problems

Let $\gamma \in \mathbb{C}^M$ be the discretized complex shear field extracted from an underlying discretized convergence field $\kappa \in \mathbb{C}^N$ by a *measurement operator* $\mathsf{\Phi} \in \mathbb{C}^{M \times N} : \kappa \mapsto \gamma$. In the planar setting $\mathsf{\Phi}$ can be modeled by,

$$\mathsf{\Phi} = \mathsf{M}\mathsf{F}^{-1}\mathsf{D}\mathsf{F}. \tag{5.3}$$

Here F is the discrete fast Fourier transform (FFT), F^{-1} is the inverse discrete fast Fourier transform (IFFT), M is a standard masking operator, and D is a diagonal matrix applying the scaling of the forward model in Fourier space as defined in equation (4.41). In the case of independent and identically distributed *i.i.d.* Gaussian noise, measurement of γ will be contaminated such that:

$$\gamma = \mathsf{\Phi}\kappa + \mathcal{N}(0, \sigma_i^2), \tag{5.4}$$

where $\mathcal{N}(0, \sigma_i^2) \in \mathbb{C}^M$ is additive i.i.d. Gaussian noise of variance σ_i^2 for pixel i. Often in weak gravitational lensing experiments the total number of binned measurements is less than the number of pixels to be recovered, $M < N$, and the inverse problem becomes ill-posed (see section 2.1.2).

Likelihood function

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 $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2) \in \mathbb{R}^{M \times M}$. Let $\Phi_i \kappa$ be the value of $\boldsymbol{\Phi}\kappa$ at pixel i, then the overall likelihood is then defined as,

$$p(\gamma|\kappa) \propto \prod_{i=0}^{M} \exp\left(\frac{-(\Phi_i\kappa - \gamma_i)^2}{2\sigma_i^2}\right) = \prod_{i=0}^{M} \exp\left(\frac{-1}{2}\left(\bar{\Phi}_i\kappa - \bar{\gamma}_i\right)^2\right) = \exp\left(\frac{-\|\bar{\boldsymbol{\Phi}}\kappa - \bar{\gamma}\|_2^2}{2}\right), \tag{5.5}$$

where $\|\cdot\|_2$ is the ℓ_2-norm and $\bar{\boldsymbol{\Phi}} = \Sigma^{-\frac{1}{2}}\boldsymbol{\Phi}$ is a composition of the measurement operator and an inverse covariance weighting. 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 γ. This likelihood function allows one to map from the number count of observations per pixel to a corresponding noise variance (assuming an intrinsic ellipticity dispersion of ~ 0.37), from which the noise (under a central limit theory argument of Gaussianity) may be correctly incorporated into the reconstruction. In practice this requires only the number density of observations per pixel, which is trivially inferred from raw observational data catalogues.

Prior function

To regularise this inverse problem, we then define a sparsity promoting Laplace-type prior:

$$p(\kappa) \propto \exp\Big(-\mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1\Big), \tag{5.6}$$

where $\boldsymbol{\Psi}$ is an appropriately selected wavelet dictionary, and $\mu \in \mathbb{R}_{>0}$ is a regularisation parameter — effectively a weighting between likelihood and prior, as discussed in section 2.2.3. Note that one may choose any convex log-prior within this formalism *e.g.* an ℓ_2-norm prior from which one essentially recovers Wiener filtering (see Padmanabhan et al., 2003; Horowitz et al., 2019, for alternate iterative Wiener filtering approaches). From equations (9.1) and (5.2) the unconstrained optimisation problem which minimises the log-posterior is,

$$\kappa^{\text{map}} = \underset{\kappa \in \mathbb{C}^N}{\text{argmin}} \Bigg[\underbrace{\mu\|\boldsymbol{\Psi}^\dagger\kappa\|_1 + \frac{\|\bar{\boldsymbol{\Phi}}\kappa - \gamma\|_2^2}{2}}_{\text{Objective function}}\Bigg], \tag{5.7}$$

where the bracketed term is called the *objective function*. To solve this convex optimisation problem we adopt a forward-backward splitting algorithm (*e.g.* Combettes and Pesquet, 2011). A full description of this algorithm applied in the current context is outlined in Cai et al. (2018b).

Proximal iterations

Let $f(\kappa) = \mu\|\Psi^\dagger\kappa\|_1$ denote our prior term, and $g(\kappa) = \|\bar{\Phi}\kappa - \gamma\|_2^2/2$ denote our data fidelity term. Then our optimisation problem can be re-written compactly as,

$$\kappa^{\text{map}} = \underset{\kappa\in\mathbb{C}^N}{\operatorname{argmin}} \Big[f(\kappa) + \mu g(\kappa) \Big]. \tag{5.8}$$

The forward-backward iteration step is then defined to be,

$$\kappa^{(i+1)} = \operatorname{prox}_{\mu^{(i)} f}\Big(\kappa^{(i)} - \mu^{(i)} \nabla g(\kappa^{(i)}) \Big), \tag{5.9}$$

for iteration i, with gradient $\nabla g(\kappa) = \bar{\Phi}^\dagger(\Phi\kappa - \gamma)$. If the wavelet dictionary Ψ is a tight frame (i.e. $\Psi^\dagger\Psi = \mathbb{I}$) the proximity operator is given by,

$$\operatorname{prox}_{\mu f}(z) = z + \Psi\Big(\operatorname{soft}_{\mu\lambda}(\Psi^\dagger z) - \Psi^\dagger z \Big), \tag{5.10}$$

where $\operatorname{soft}_\mu(z)$ is the point-wise soft-thresholding operator derived in section 2.2.4, and λ is a parameter related to the step-size (which is in turn related to the Lipschitz differentiability of the log-prior) which should be set according to Cai et al. (2018b). The iterative algorithm is given explicitly in the primary iterations of algorithm 2. Adaptations for frames which are not tight can be found in Cai et al. (2018b) and are readily available within our framework.

Our algorithm has distinct similarities to the GLIMPSE algorithm presented by Lanusse et al. (2016), but crucially differs in several aspects. Most importantly we formulate the problem in a hierarchical Bayesian framework which allows us to recover principled statistical uncertainties. In addition to this we include Bayesian inference of the regularisation parameter, a robust estimate of the noise-level (which can be folded into the hierarchical model), and we use super-resolution operators instead of non-discrete fast Fourier transforms.

5.2.3 Reduced shear

Due to a degeneracy between γ and κ the true observable quantity is in fact the *reduced shear* $g = \gamma[1-\kappa]^{-1}$ (Bartelmann and Schneider, 2001). Deep in the weak lensing regime one can safely approximate $\gamma \approx g \ll 1$ which ensures that the optimisation problem remains linear. However, when reconstructing regions close to massive structures (galaxy clusters) this approximation is no longer strictly valid and we must unravel this additional factor. We adopt the procedure outlined in Wallis et al. (2021), which is more comprehensively discussed in Mediavilla et al., pg 153. We find that these corrections typically converge after $\sim$ 5-10 iterations.

5.2.4 Regularisation parameter selection

One key issue of sparsity-based reconstruction methods is the selection of the regularisation parameter μ. Several methods have arisen (Paykari et al., 2014; Lanusse et al., 2016; Peel et al., 2017a; Jeffrey et al., 2018) for selecting μ, though often the regularisation parameter is chosen somewhat arbitrarily — as the integrity of the MAP solution is assumed to be weakly dependent on the choice of μ. However, to extract principled statistical uncertainties on the recovered images, one must select this parameter in a principled statistical manner.

We apply the hierarchical Bayesian formalism developed by Pereyra et al. (2015) — the details of which are elegantly presented by the authors, though we will outline roughly the underlying argument here. First define a sufficient statistic f to be k-homogeneous if $\exists\, k \in \mathbb{R}_{>0}$ such that $f(\eta x) = \eta^k f(x), \forall x \in \mathbb{R}^N, \forall \eta > 0$. All norms, composite norms and composition of norms with linear operators are 1-homogeneous — and so our ℓ_1-norm is 1-homogeneous. If a sufficient statistic f is k-homogeneous, then the normalization factor $C(\mu)$ of $p(\kappa|\mu)$ is given by (proposition 3.1 of Pereyra et al., 2015),

$$C(\mu) = A\mu^{-\frac{N}{k}}, \tag{5.11}$$

where A is a constant independent of μ. The proposed Bayesian inference model then implements a gamma-type hyper-prior, which is a typical hyper-prior for scale-parameters

$$p(\mu) = \frac{\beta^\alpha}{\Gamma(\alpha)}\mu^{\alpha-1}e^{-\beta\mu}\mathbb{I}_{\mathbb{R}_{>0}}(\mu), \tag{5.12}$$

where without loss of generality $\alpha = \beta = 1$. The result is effectively insensitive to their value, *e.g.* in numerical experiments values of $\alpha, \beta \in [10^{-2}, 10^5]$ produced essentially no difference in μ.

Now, let us extend the inference problem of the log-posterior to the case where μ is an additionally unknown parameter. In this context we compute the joint MAP estimator $(\kappa^{\text{map}}, \mu^{\text{map}}) \in \mathbb{C}^N \times \mathbb{R}_{>0}$ which maximizes $p(\kappa, \mu|\gamma)$ such that,

$$\mathbf{0}_{N+1} \in \partial_{\kappa,\mu} \log p(\kappa^{\text{map}}, \mu^{\text{map}}|\gamma), \tag{5.13}$$

where $\mathbf{0}_i$ is the i-dimensional null vector and $\partial_s h(s')$ is the set of sub-gradients of function $h(s)$ at s'. This in turn implies both that,

$$\Rightarrow \mathbf{0}_N \in \partial_\kappa \log p(\kappa^{\text{map}}, \mu^{\text{map}}|\gamma) \qquad \text{and} \qquad \Rightarrow \mathbf{0} \in \partial_\mu \log p(\kappa^{\text{map}}, \mu^{\text{map}}|\gamma). \tag{5.14}$$

From equation (5.14 - LHS) we recover the optimisation problem with known regularisation parameter μ given in equation (5.7). However, from equations (5.11, 5.12, 5.14 - RHS) it follows that the MAP regularisation parameter μ is given by (Pereyra et al., 2015),

$$\mu^{\text{map}} = \frac{\frac{N}{k} + \alpha - 1}{f(\kappa^{\text{map}}) + \beta}, \tag{5.15}$$

where we recall that N is the total dimension of our convergence space.

It is precisely this optimal μ value which we wish to use in our hierarchical Bayesian model. Hereafter we drop the map superscript from μ for clarity. To calculate μ we perform preliminary iterations found in algorithm 2. Typically we find that these preliminary iterations take $\sim$ 5-10 iterations to converge, and recover close to optimal parameter selection for a range of test cases. Note that, in this case the optimal selection of μ is that which maximizes the SNR of a recovered image. Another factor which can influence the quality of reconstructions is the selection of wavelet dictionary. In this chapter we consider Daubechies (8 levels) and SARA dictionaries (Carrillo et al., 2012, 2013), though a wide variety of wavelet dictionaries exists, see *e.g.* starlets (Starck et al., 2010). The 8-level SARA dictionary is a combination of the Dirac and Daubechies 1 to 8 wavelet dictionaries. It is important to note that we use the SARA dictionary, not the complete SARA scheme (Carrillo et al., 2012, 2013), which involves an iterative re-weighting scheme that is not considered here.

Algorithm 2 Forward-backward analysis

```
Input: $\gamma \in \mathbb{C}^M$, $\kappa^{(0)} \in \mathbb{C}^N$, $\lambda$, $\mu^{(0)} = i = t = 0$, $T_{1,2} \in \mathbb{R}_{>0}$
Output: $\kappa^{\text{map}} \in \mathbb{C}^N$, $\mu \in \mathbb{R}_{>0}$
  Precomputation:
    Do:
      $\kappa^{(t)} \leftarrow \text{argmin}_{\kappa} \left[ f(\kappa) + \mu^{(t)} g(\kappa) \right]$,
      $\mu^{(t+1)} \leftarrow [(N/k) + \alpha - 1]/[f(\kappa^{(t)}) + \beta]$,
      $t \leftarrow t + 1$,
      On convergence, $\mu$ becomes fixed.
    Until: Iteration limit reached.
  Primary Iterations:
    Do:
      $\nu^{(i+1)} \leftarrow \kappa^{(i)} - \lambda \bar{\Phi}^{\dagger}(\bar{\Phi}\kappa^{(i)} - \gamma)$,
      $\eta \leftarrow \Psi^{\dagger}\nu^{(i+1)}$,
      $\kappa^{(i+1)} \leftarrow \nu^{(i+1)} + \Psi(\text{soft}_{\lambda\mu}(\eta) - \eta)$,
      $i \leftarrow i + 1$.
    Until: Stopping criterion satisfied.
      i.e. $\frac{\|\kappa^{(i)} - \kappa^{(i+1)}\|_2}{\|\kappa^{(i)}\|_2} < T_1$ and $\frac{\text{obj}(\kappa^{(i)}) - \text{obj}(\kappa^{(i+1)})}{\text{obj}(\kappa^{(i)})} < T_2$.
```

5.2.5 Super-resolution image recovery

Gridding of weak lensing data is advantageous in that it can provide a good understanding of the noise properties — a necessary feature for principled uncertainty quantification. However, an inherent drawback of projecting data into a grid is the possibility of creating an incomplete space due to low sampling density – often referred to as masking. Decomposition of spin signals on bounded manifolds is inherently degenerate (Bunn et al., 2003); specifically the orthogonality of eigenfunctions is locally lost at the manifold boundaries, leading to signal leakage between Fourier (or, on the sphere, harmonic) modes. One approach to mitigate this problem is to avoid the necessity of gridding by substituting a *non-uniform discrete Fourier transform* (NFFT) into the RHS of equation (9.4) as presented by Lanusse et al. (2016). A downside of this NFFT approach is that the noise is more difficult to handle, leading to complications when considering uncertainty quantification. Another approach is to perform super-resolution image recovery.

Suppose the dimension of our gridded measurement space is M, as before, and the desired dimension of our solution space is N', where $N' \geq N$. In this setting we have shear measurements $\gamma \in \mathbb{C}^M$ and recovered convergence $\kappa \in \mathbb{C}^{N'}$. Let us now define a *super-resolution* (subscript SR) measurement operator to be,

$$\mathsf{\Phi}_{\mathrm{SR}} = \mathsf{M}\,\mathsf{F}_{\mathrm{lr}}^{-1}\,\mathsf{D}\,\mathsf{Z}\,\mathsf{F}_{\mathrm{hr}}, \tag{5.16}$$

where F_{hr} is a high resolution (dimension N') fast Fourier transform, $\mathsf{Z} \in \mathbb{C}^{N\times N'}$ is a Fourier space down-sampling which maps $\tilde{\kappa}' \in \mathbb{C}^{N'}$ on to $\tilde{\kappa} \in \mathbb{C}^N$, where tilde represents Fourier coefficients, D is the planar forward model given by equation (4.41), and M is a standard masking operator. Finally, $\mathsf{F}_{\mathrm{lr}}^{-1}$ is a low resolution (dimension M) inverse fast Fourier transform. For completeness the super-resolution adjoint measurement operator is given by,

$$\mathsf{\Phi}_{\mathrm{SR}}^{\dagger} = \mathsf{F}_{\mathrm{hr}}^{-1}\,\mathsf{Z}^{\dagger}\,\mathsf{D}^{\dagger}\,\mathsf{F}_{\mathrm{lr}}\,\mathsf{M}^{\dagger}, \tag{5.17}$$

where $\mathsf{M}^{\dagger}$ is adjoint masking (gridding), $\mathsf{D}^{\dagger}$ is the adjoint of D (which is self-adjoint hence $\mathsf{D}^{\dagger} = \mathsf{D}$), and $\mathsf{Z}^{\dagger} \in \mathbb{C}^{M'\times M}$ is zero padding in Fourier space which acts by mapping $\tilde{\gamma} \in \mathbb{C}^M$ to $\tilde{\gamma}' \in \mathbb{C}^{M'}$. Note that when considering the KS estimate in the super-resolution setting a rescaling function to account for the different Fourier normalization factors must be introduced (which we absorb into the Fourier operators). As before, this super-resolution measurement operator is concatenated with the inverse covariance weighting to form an analogous composite operator $\bar{\mathsf{\Phi}}_{\mathrm{SR}}$ which is used throughout the following analysis.

Conceptually super-resolution allows partial inpainting of higher resolution Fourier modes. In this way one is able to recover high resolution structure for images from comparatively low resolution datasets. Such high resolution structure is of course dependent on the prior information injected when solving the inverse problem. Interestingly this raises another consideration: in scenarios where the pixel-level observation count is very low the noise level dilutes high frequency components and can limit the efficacy of reconstruction algorithms. In such a setting gridding observational data onto a lower resolution map, with inherently lower pixel-level noise, and performing a super-resolution reconstruction can recover far better estimates of the high frequency modes, and thus often recovers greater reconstruction fidelity.

5.3 Bayesian hypothesis testing

Extending the concept of HPD credible regions outlined in chapter 3, one can perform *knock-out* hypothesis testing of the posterior to determine the physicality of recovered structure (Cai et al., 2018b). To perform such tests one first creates a surrogate image κ^{sgt} by masking a feature of interest $\Omega_D \subset \Omega$ in the MAP estimator κ^{map}. It is then sufficient to check if

$$f(\kappa^{\text{sgt}}) + g(\kappa^{\text{sgt}}) \leq \epsilon'_\alpha. \tag{5.18}$$

If this inequality holds, we interpret that the physicality of Ω_D is undetermined and so no strong statistical statement can be made. Should the objective function evaluated at κ^{sgt} be larger than ϵ'_α then it no longer belongs to the approximate credible set C'_α and therefore (as ϵ'_α is conservative) it **cannot** belong to the HPD credible set C_α. Therefore, for κ^{sgt} which do not satisfy the above inequality we determine the structure Ω_D to be strictly physical at $100(1-\alpha)\%$ confidence level. A schematic of hypothesis testing is provided in Figure 5.1.

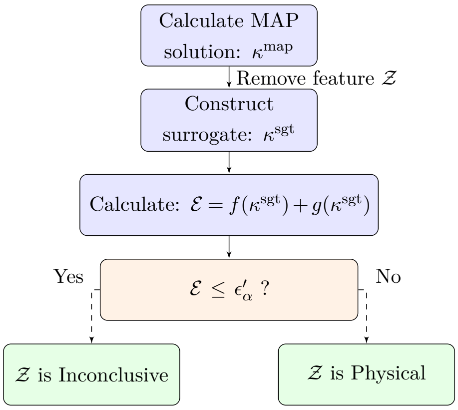


Figure 5.1: Schematic of hypothesis testing. The feature $\mathcal{Z}$ is entirely general and can be constructed by any well defined operator on the MAP solution κ^{map}.

In pixel-space we begin by masking out a feature of interest, creating a rough surrogate image — setting the pixels associated with a selected structure to 0 — this rough surrogate is then passed through an appropriate wavelet filter Λ as part of *segmentation-inpainting* to replace generic background structure into the masked region. Mathematically, this amounts to the iterations,

$$\kappa^{(i+1),\text{sgt}} = \kappa^{\text{map}} \mathbb{I}_{\Omega/\Omega_D} + \Lambda^\dagger \text{soft}_{\lambda_t}(\Lambda \kappa^{(i),\text{sgt}}) \mathbb{I}_{\Omega_D}, \tag{5.19}$$

where $\Omega_D \subset \Omega$ is the sub-set of masked pixels, $\mathbb{I}_{\Omega-\Omega_D}$ is the set indicator function and λ_t is a thresholding parameter which should be chosen appropriately for the image.

A second straightforward method for generating surrogate images is to blur local pixel substructure into one collective structure — in a process called *segmentation-smoothing*. This approach provides a simple way to determine if the substructure in a given region is physical or likely to be an artifact of the reconstruction process. For example, if several massive peaks are located near one another, one can blur these structures into a single cohesive peak. This would be useful when considering peak statistics on convergence maps — which is often used to constrain the cosmological parameters associated with dark matter. One can conduct such blurring of structure by: specifying a a subset of the reconstructed pixels $\Omega_D \subset \Omega$; convolving κ^{map} with a Gaussian smoothing kernel; and replacing pixels that belong to Ω_D with their smoothed counterparts. This can be displayed algorithmically as,

$$\kappa^{\text{sgt}} = \kappa^{\text{map}} \mathbb{I}_{\Omega / \Omega_D} + (\kappa^{\text{map}} * \mathcal{G}(0, \chi)) \mathbb{I}_{\Omega_D}, \tag{5.20}$$

where $\mathcal{G}(0, \chi)$ is a chosen Gaussian smoothing kernel and $*$ is a trivially extended 2D version of the usual 1D Fourier convolution operator,

In the scope of this chapter we focus primarily on pixel-space features, but it is important to stress that the *knock-out* approach is entirely general and can be applied to any well defined feature of a MAP estimator — *i.e.* masking certain Fourier space features, removal of global small scale structure *etc.*

5.4 Illustration on simulations

We now consider a selection of realistic simulations to illustrate our sparse reconstruction method on cluster scales which are particularly challenging for a variety of factors. Further to this, we showcase the aforementioned uncertainty quantification methods in a variety of idealised cluster scale MAP reconstructions. We place emphasis on uncertainty quantification rather that the reconstruction fidelity.

Datasets

In this section we focus primarily on 4 large clusters (those with significant friends-of-friends, *i.e.* significant substructure) extracted from the Bolshoi N-body simulation (Klypin et al., 2011). On the cluster scale we showcase our formalism on a variety of Bolshoi N-body simulation data sets. The Bolshoi N-body cluster simulation catalogs we work with in this section are those used in Lanusse et al. (2016), which were extracted using the CosmoSim web tool[1]. Construction of these weak lensing realizations assumed a redshift of 0.3, with a 10×10 arcmin^2 field of view, and have convergence normalized with respect to lensing sources at infinity. Explicitly this results in pixel-dimensions of ~ 2.5 arcseconds. Due to the relatively low particle density, these images were subsequently de-noised by a multi-scale Poisson de-noising algorithm.

Method

Typically, we begin by creating an artificial shear field $\hat{\gamma} \in \mathbb{C}^M$ from a known *ground-truth* convergence field κ, that is extracted from a given dataset. This is a common approach in the imaging community and presents a closed scenario in which the true input is known. These $\hat{\gamma}$ fields are created by $\hat{\gamma} = \Phi\kappa + \mathcal{N}(0, \sigma_i^2)$ where σ_i, *i.e.* the noise covariance, is determined entirely from a pre-defined number density of observations n_{gal} per arcminute^2, an assumed intrinsic ellipticity dispersion of 0.37, and the resolution of the images (in this case 10×10 arcminutes). In this way the

---

[1] https://www.cosmosim.org

noise can be tuned to directly mimic that present in practical settings. Using the simulated noise covariance (which in practice would be provided by the observation team) we then utilise the SOPT[2] framework to perform our reconstruction algorithm on $\hat{\gamma}$ such that we recover a MAP estimator of the convergence κ^{map}. From this reconstructed convergence field a recovered SNR is computed and a selection of hypothesis tests are conducted to showcase the power of this formalism. In the case where the underlying clean γ are unavailable (*i.e.* application to A520 data) we conduct the same analysis as before but instead of creating artificial noisy $\hat{\gamma}$ maps we used the real noisy observational data.

Throughout our analysis the recovered SNR (dB) is defined to be,

$$\text{SNR} = 20 \times \log_{10}\left(\frac{\|\kappa\|_2}{\|\kappa - \kappa^{\text{map}}\|_2}\right), \tag{5.21}$$

when the ground-truth convergence is known. Furthermore we quantify the topological similarity between the true convergence and the estimator *via* the Pearson correlation coefficient which is defined to be

$$r = \frac{\sum_{i=1}^{N}[\kappa^{\text{map}}(i) - \bar{\kappa}^{\text{map}}][\kappa(i) - \bar{\kappa}]}{\sqrt{\sum_{i=1}^{N}[\kappa^{\text{map}}(i) - \bar{\kappa}^{\text{map}}]^2}\sqrt{\sum_{i=1}^{N}[\kappa(i) - \bar{\kappa}]^2}}, \tag{5.22}$$

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.

5.4.1 Bolshoi cluster catalogs

The Bolshoi cluster data used consists of 4 large clusters extracted from the Bolshoi N-body simulation (Klypin et al., 2011; Lanusse et al., 2016). These images were then multi-scale Poisson de-noised to create suitable ground truth simulations. We choose to analyze the same clusters considered in Lanusse et al. (2016), as they showcase a wide variety of structure on all scales. Hereafter, we restrict ourselves to the SARA dictionary (Carrillo et al., 2012) truncated at the 4th Daubechies wavelet (DB4) for simplicity — *i.e.* the combination of the Dirac, and DB1 to DB4 wavelet dictionaries only.

To investigate the SNR gain of our formalism over KS in the cluster scale setting, we created realizations of noisy pseudo-shear maps for assumed number density of galaxy observations $n_{\text{gal}} \in [500, 100, 30, 10]$ from one Bolshoi cluster map, upon which we applied our reconstruction algorithm pipeline, the results of which are presented in Table 5.1. It should be noted that for comparison's sake the KS estimate without convolution with a Gaussian smoothing kernel is provided in addition to an optimally smoothed KS estimator. This has been done to highlight the difference

---

[2] A highly optimised sparse optimisation solver, `https://github.com/astro-informatics/SOPT`

in reconstruction fidelity between the raw KS estimator and the KS estimator after post-processing (Gaussian smoothing), a discrepancy often not addressed by the community. As this post-processing convolution is known to degrade the quality of non-Gaussian information (which cosmologists are becoming increasingly interested in) such plots demonstrate the trade-off between non-Gaussian information and reconstruction fidelity. As can be seen in Figure 5.2 and Table 5.1, sparse approaches significantly outperform the smoothed (and non-smoothed) KS approach in all cases, over all metrics tracked. Importantly sparse approaches are able to recover reasonable results even when the noise level entirely dilutes the true signal, as in the $n_{\text{gal}} = 10$ setting, making such approaches on (at least) cluster data very attractive for future studies.

Input n_{gal}	**KS**	**KS Smooth**	**Sparse**	**Difference**
SNR (dB)				
500	2.917	6.276	27.506	+ 21.230
100	-4.497	5.774	21.955	+ 16.181
30	-10.400	5.340	21.462	+ 16.122
10	-15.970	5.041	14.409	+ 9.368
Pearson Correlation				
500	0.166	0.902	0.977	+ 0.075
100	0.076	0.796	0.970	+ 0.174
30	0.039	0.689	0.955	+ 0.266
10	0.029	0.716	0.949	+ 0.233

Table 5.1: Contains both reconstruction SNR and Pearson correlation coefficient (topological correlation) metrics for the raw KS (no smoothing), an optimally smoothed KS (grid search for smoothing kernel which maximizes the recovered SNR), and our sparse reconstructions of the Bolshoi-3 cluster simulated with realistic noise derived from the presented number density of galaxy observations n_{gal}. The difference column is calculated as the difference between the Sparse and smoothed KS recovered SNR, given in log-space.

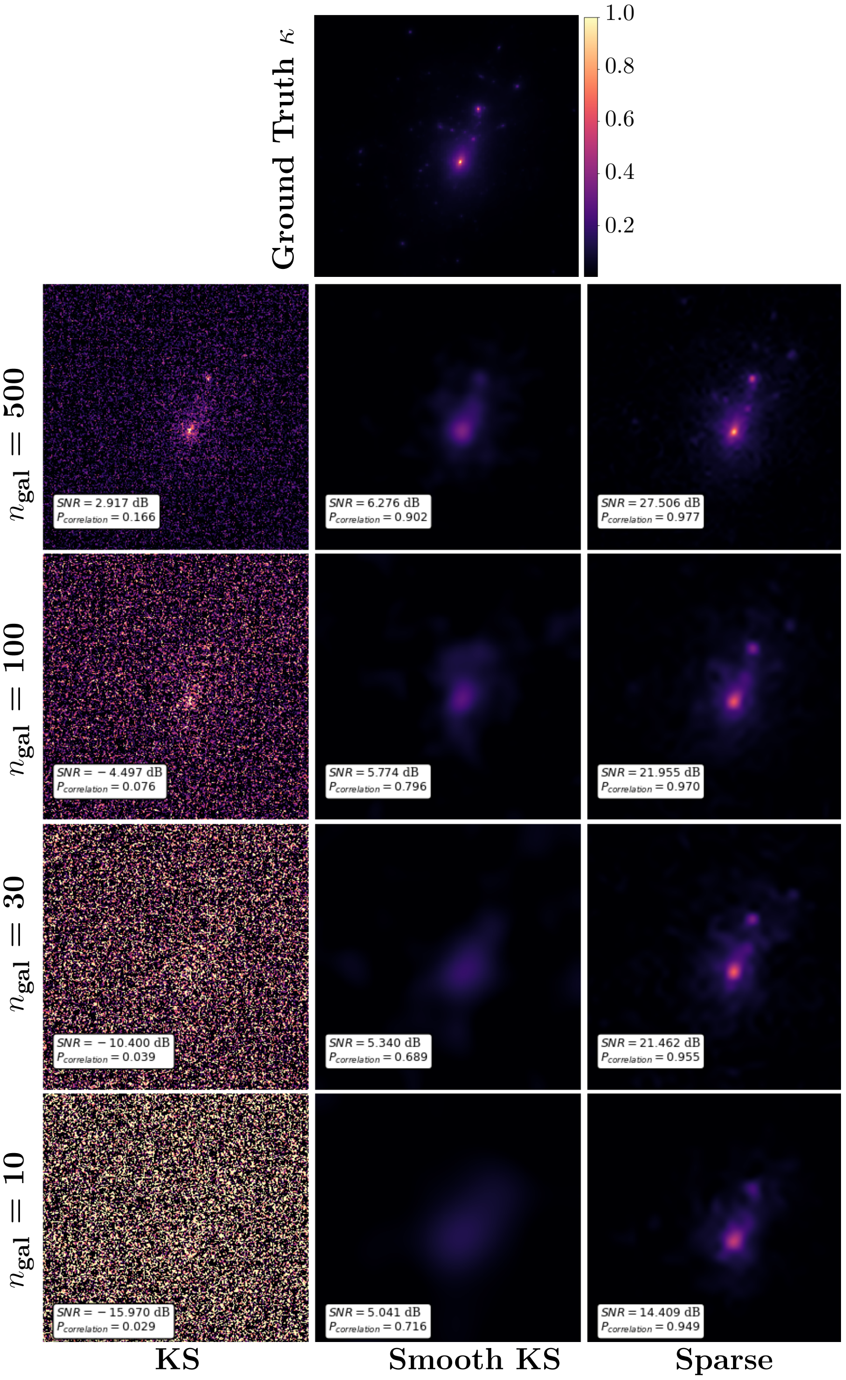


Figure 5.2: Top to bottom: Ground truth convergence map, simulations with noise levels corresponding to $n_{\rm gal} \in [500, 100, 30, 10]$ respectively. Notice the clear effectiveness of sparse reconstruction over the standard KS method for a range of input SNR values. The numerical details can be found in Table 5.1. The vertical labels indicate the input $n_{\rm gal}$ used to simulate realistic noise for a given row, whereas horizontal labels indicate the reconstruction type. An optimal (grid searched to maximise the recovered SNR) Gaussian smoothing kernel was applied to the KS recovery to yield the KS (smooth) recovery in an attempt to remove noise from the KS estimator (obviously this is not possible in practice, where the ground truth is unknown: results shown therefore present the best possible performance for the smoothed KS estimator). Clearly, in all cases, the super-resolution sparse approach produces convergence maps which are far more representative of the ground truth across the aforementioned metrics.

Hypothesis Testing: Bolshoi Clusters

Perhaps more interestingly, we now perform a series of hypothesis tests as discussed in Section 5.3. For each of the remaining 3 Bolshoi cluster we construct three possible example hypothesis tests which one may wish to perform. In this case these hypotheses were either: structure removal followed by segmentation-inpainting; or Gaussian smoothing of certain structures (*i.e.* smoothing multiple peaks into a single larger peak which may be of interest when conducting peak-count analysis). Though these are both extremely useful considerations, it is important to stress the generality of our approach such that any well defined operation on the reconstructed image, with a clear understandable hypothesis, is applicable. To ensure the mindset behind hypothesis testing is clear, we will walk through a typical application. The top row of Figure 5.3 displays the hypothesis tests applied to the first Bolshoi cluster. Conceptually, the correct way to interpret Hypothesis 1 (H1, red) is:

> *"The central dark core is likely just an artifact of the reconstruction."*

This structure is then removed from the image by segmentation-inpainting (lower left image), and the objective function is then recalculated. It is found that the objective function is now larger than the approximate level-set threshold $\epsilon'_{99\%}$, the surrogate segmentation-inpainted image falls outside of the 99% HPD credible region, and so the hypothesis is **rejected**. This implies that the structure is not simply an artifact, but is necessary to the integrity of the reconstruction, *i.e.* this structure is now determined to be physical at 99% confidence. However, had removing this region **not** raised the objective function above $\epsilon'_{99\%}$, then the conclusion is that there is insufficient evidence to reject the hypothesis (which is **not** equivalent to saying that the region is strictly not physical).

An identical thought process can be applied to H2 and H3 of the top row in Figure 5.3, H4 in the second row of Figure 5.3, and all three hypothesis tests H7-9 presented in the final row. In each case a substructure of the κ^{map} is removed *via* segmentation-inpainting and it is queried whether the resulting surrogate solution $\kappa^{\text{sgt}} \in C'_{\alpha}$. Each of the large substructures H7-9 on the final row, and H4 of the second row, are determined to be physical at 99% confidence. Conversely, the comparatively smaller substructures considered in H2 and H3 of the top row do not saturate the level-set threshold, and are therefore undetermined. All numerical data related to hypothesis testing of the Bolshoi cluster reconstructions can be found in Table 5.2.

H5 and H6 of the middle row of Figure 5.3 have a slightly different interpretation. In these cases the central region has been blurred by segmentation-smoothing (convolution with a Gaussian smoothing kernel) – the difference between these two cases being simply the degree of smoothing. Here the hypothesis is: *'The central region is likely to be just a single peak, rather than two'*. As in the previous example, the

Test	**Initial** $f(\kappa)+g(\kappa)$	**Threshold** $\epsilon'_{99\%}$	**Surrogate** $f(\kappa^{\mathrm{sgt}})+g(\kappa^{\mathrm{sgt}})$	**Reject** H_0 **?**
Bolshoi-1				
H1	95426	163408	805513	✓
H2	95426	163408	134080	×
H3	95426	163408	100582	×
Bolshoi-2				
H4	97121	165103	824260	✓
H5	97121	165103	221492	✓
H6	97121	165103	366981	✓
Bolshoi-3				
H7	83419	151401	369939	✓
H8	83419	151401	234305	✓
H9	83419	151401	314089	✓

Table 5.2: Displays the MAP objective function, level-set threshold at 99% confidence, surrogate objective function and whether the removed region was successfully identified as being physical. This data-set corresponds to Figure 5.3, the caption of which provides detailed descriptions of each hypothesis test.

objective function is recalculated and is now greater than $\epsilon'_{99\%}$ and so the hypothesis is rejected. The natural conclusion is thus that the data **is** sufficient to determine that at least two peaks are physically present at 99% confidence.

5.5 Application to Abel-520 observational catalogs

We perform an application of our entire reconstruction pipeline to real observational datasets. We select two observational datasets of the A520 cluster (Clowe et al., 2012; Jee et al., 2014) — hereafter for brevity we refer to them as C12 and J14 (as in Peel et al., 2017a)[3]. For a full description of the datasets, how they were constructed, and how they account for different systematics we recommend the reader look to the respective papers. These initial investigations claim to have detected several over-dense regions within the merging A520 system, the most peculiar of which was a so called 'dark core' (location 2 in Figure 5.4) for which multi-wavelength observations could not determine an optical counterpart. Such a dark core would provide a contradiction to the currently understood model of collisionless dark matter – the idea being that during the collision of two massive clusters, dark matter was stripped from each cluster through self-interactions, forming an over-dense residual between the two clusters, which would naturally not exhibit an optical counterpart.

The J14 catalog contains approximately twice the number of galaxies than that of C12, though both are derived from the same ACS (four pointings) and Magellan images. In addition, J14 combines these images with the CFHT catalog used in the author's previous work (Jee et al., 2012). The C12 observing area extends over a larger angular surface than the J14 so for this analysis we limit both datasets to the region spanned by both sets. Due to the number density of measurements being

---

[3] `http://www.cosmostat.org/software/glimpse`

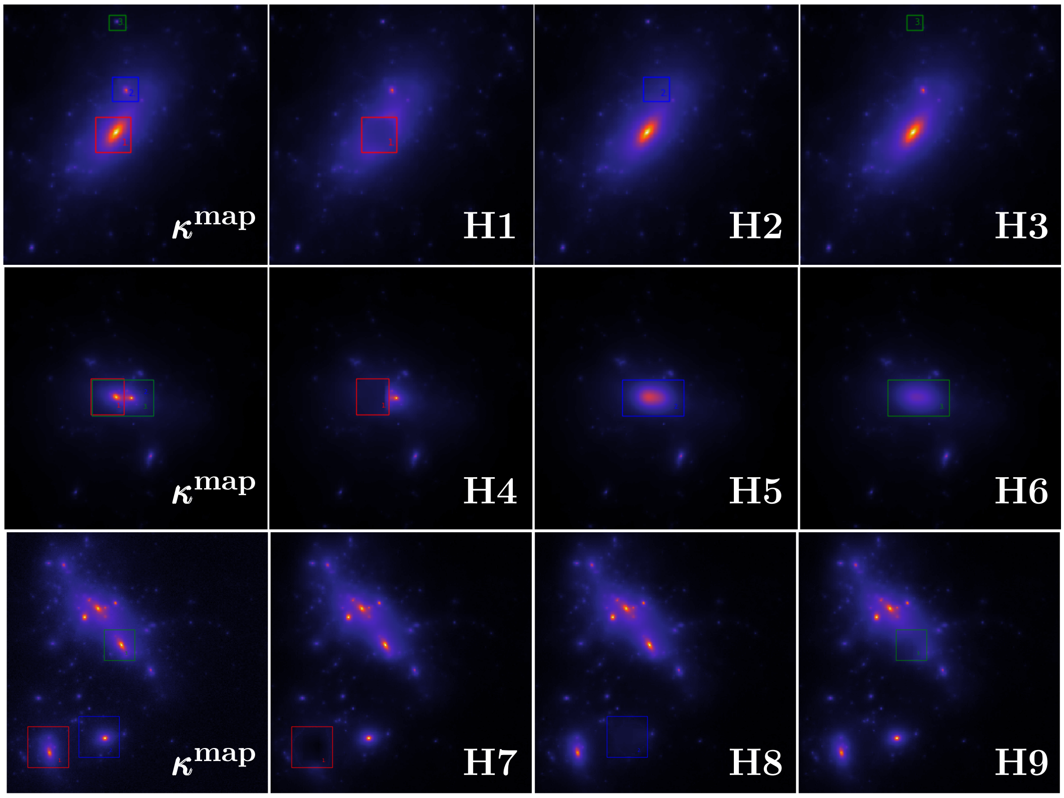


Figure 5.3: General: Hypothesis testing of three selected structures in the Bolshoi-1 cluster convergence field. The number density of galaxy observations n_{gal} was set idealistically to 500 arcminute^{-2} simply for demonstration purposes. Additionally super-resolution was not active and the masking was trivially set to the identity, again to simplify the example for demonstration purposes. All numerical details can be found in Table 5.2. **Top row:** We correctly determine that region 1 (*red*) is physical with 99% confidence. Regions 2 (*blue*) and 3 (*green*) remain within the HPD region and are therefore inconclusive, given the data and noise level. **Middle row:** We correctly determine that all three null hypotheses (*red, blue* and *green*) are rejected at 99% confidence. In H4 the conclusion is that the left hand peak was statistically significant. In H5 and H6 the conclusions is that an image with the two peaks merged is unacceptable, and therefore the peaks may be considered distinct at 99% confidence. **Bottom row:** We correctly determine that all three hypothesis regions H7-9 (*red, blue* and *green*) Ω_D are physical with 99% confidence.

very low we are forced to project the measurements into a 32×32 grid — to ensure that the average number of galaxies in each grid pixel is at least above 1, though ideally we want many galaxies in each pixel to minimise the noise contribution from intrinsic ellipticity. In fact, even at this resolution, the space is incomplete in several pixels, but we draw a compromise between the completeness of the space and the resolution of the data. The data covariance was constructed directly from the number density of observations per pixel (directly inferred during catalogue gridding), with an assumed intrinsic ellipticity dispersion of 0.37. Combining this data covariance, the associated gridded datasets, and the associated mask, MAP reconstructions of the C12 and J14 convergence maps were recovered at a super-resolution magnification of 8. Reconstructions are presented in Figure 5.4.

5.5.1 Hypothesis testing of local structure: A520 datasets

We conducted hypothesis tests on the three primary over-dense regions, in addition to the contested dark core, in both the C12 and J14 datasets. In the absence of an optical counterpart, detection at high confidence of the dark core (location 2 in Figure 5.4) would provide a contradiction to the collisionless model of dark matter – indicating potential self-interaction of dark matter. Due to the high estimated noise-level present in the data, and the limited data resolution, only the two largest peaks in both datasets (peaks 1 and 3 of Figure 5.4) sufficiently raised the objective function to reject the hypothesis at any meaningful confidence. This is to say that; given the limited, noisy data and using the measurement operator and prior (ℓ_1-term) presented in this chapter we can say that the data is insufficient to statistically determine the physicality of local small scale substructure (such as the dark core) in both the C12 and J14 datasets. The initial conflict between C12 and J14 was over the existence and position of a dark core (location 2 in Figure 5.4), with a notably large mass-to-light ratio, which indicated the possibility of self-interacting dark matter. A subsequent inquiry was conducted (Peel et al., 2017a) using the GLIMPSE reconstruction algorithm (Lanusse et al., 2016) and concluded that this peculiar peak existed in the J14 dataset but not in the C12 dataset. As such, our conclusions agree well with Peel et al. (and generally with those drawn in both C12 and J14). However, within our Bayesian hierarchical formalism (which constitutes a principled statistical framework) we push this conclusion further to say that the data are insufficient to determine the physicality of these peaks.

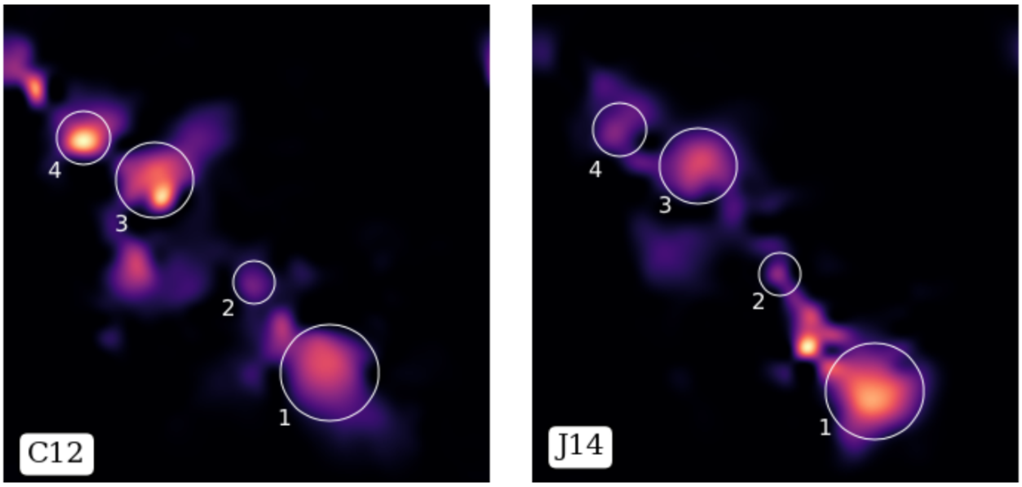


Figure 5.4: Left to Right: Super-resolution sparse Bayesian reconstruction of [C12] and [J14] respectively. In a Bayesian manner it is found that the two datasets do not globally disagree at 99% confidence. However, given the data resolution and noise-levels, only peaks 1 and 3 (in both datasets) could be determined to be statistically significant. This is **not** to say they do not exist, but implies that the data quantity and quality is insufficient to make a robust, principled statistical statement which could be used as evidence of their existence. The contested peak 2 is not detected at any reasonable confidence in either dataset.

5.5.2 Hypothesis testing of global structure: A520 datasets

Interestingly we can perform a final novel hypothesis test of global structure. This hypothesis is as follows: *'The two MAP estimates are consistent with both sets of data', i.e.* the MAP convergence estimate recovered from the J14 (C12) data is within the credible-set (at 99% confidence) of the C12 (J14) objective function. We find that the J14 (C12) MAP reconstruction is an acceptable solution to the C12 (J14) inverse problem and so the MAP solutions do not disagree — numerically this is shown in Table 5.3. Given the inherent limitations of the data we are forced to conclude:

> *"The data are insufficient to determine the existence of individual substructures at high confidence — though the two largest over-dense regions are found to be globally physical at* 99% *confidence. The two maximum a posteriori estimates are also found to be consistent at* 99% *confidence."*

Hypothesis Test	**Initial** $f(\kappa) + g(\kappa)$	**Threshold** $\epsilon'_{99\%}$	**Surrogate** $f(\kappa^{\text{sgt}}) + g(\kappa^{\text{sgt}})$	**Reject** $\boldsymbol{H_0}$**?**
C12 ⇔ J14	99231	168044	125601	×
J14 ⇔ C12	98943	167243	134391	×

Table 5.3: Displays the MAP objective function, level-set threshold at 99% confidence, surrogate objective function and whether the null hypothesis H_0 is rejected. As can be seen, both MAP solutions fail to reject the null hypothesis in the other's objective function. This leads us to conclude that the two datasets do not disagree at 99% confidence. Further discussion akin to the Kullback-Leibler divergence (see *e.g.* Kullback, 1997) of the two posteriors is beyond the scope of the current discussion, but perhaps of interest in future work.

5.6 Summary

We have presented a sparse hierarchical Bayesian mass-mapping algorithm which provides a principled statistical framework through which, for the first time, we can conduct uncertainty quantification on recovered convergence maps without relying on any assumptions of Gaussianity. Moreover, the presented formalism draws on ideas from convex optimisation (rather than MCMC techniques) which makes it notably fast and allows it to scale well to big data, *i.e.* high resolution and wide-field convergence reconstructions (which will be essential for future stage IV surveys, such as LSST and Euclid). Additionally, we demonstrate a hierarchical Bayesian inference approach to automatically approximate the regularisation parameter, and show that it produces near optimal results in a variety of cases. We however note that this approach does not work generally, and can be unstable in extreme settings.

We showcase our Bayesian inference approach (with emphasis on the application of the uncertainty quantification techniques) on both simulation datasets and observational data (the A520 merging cluster dataset). Our mass-mapping formalism is shown to produce significantly more accurate convergence reconstruction than the Kaiser-Squires estimator on all simulations considered. Hypothesis tests of substructure are demonstrated. It is found that neither of the two A520 datasets considered could provide sufficient evidence to determine the physicality of any contested substructure (*i.e.* the existence of so called ‘dark cores’) at significant confidence. It is informative to note that our methods were, in fact, sufficiently sensitive to detect the largest peaks in both datasets at 99% confidence. Nonetheless, global hypothesis tests indicate a good agreement between the two sets of data. These conclusions are roughly in agreement with those drawn previously but go further to demonstrate just how uncertain these types of cluster-scale weak lensing reconstruction inherently are (typically as a limitation of the relative information content of low-resolution, noisy datasets).

Chapter 6

Local credible intervals

This chapter is based on research presented in M. A. Price, X. Cai, J. D. McEwen, M. Pereyra, T. D. Kitching, and LSST Dark Energy Science Collaboration, "Sparse Bayesian mass mapping with uncertainties: local credible intervals", Monthly Notices of the Royal Astronomical Society, vol. 492, no. 1, pp. 394-404, Dec. 2019.

Price et al. (2019a)

Note: *This chapter assumes knowledge of the following. Sparse Bayesian convergence estimators, as presented in chapter 5 — in particular we will adopt the MAP estimators of section 5.2. Proximal analysis, discussed at length in sections 2.2.4 and 3.2.1.*

In this chapter we introduce a further uncertainty quantification technique called *local credible intervals* (*cf.* pixel-level error bars). Both hypothesis testing and local credible intervals were previously developed and applied to the radio interferometric setting (Cai et al., 2018a,b). We also remark that there are alternative methods through which one may test structural content of images (see *e.g.* Repetti et al., 2019). This chapter serves as a benchmark comparison of the recently developed sparse hierarchical Bayesian formalism (see Price et al., 2021a)[1] to a bespoke MCMC algorithm, Px-MALA (Pereyra, 2016; Cai et al., 2018a,b; Durmus et al., 2018). This comparison is grounded in the context of cosmic shearing signals, the primary topic of this thesis. Px-MALA utilizes Moreau-Yoshida envelopes and proximity operators (tools from convex analysis) to efficiently support non-differentiable terms in the prior or likelihood, thus permitting non-smooth sparsity-promoting priors (on which our sparse Bayesian mass-mapping framework is based) in high dimensional settings. Full details pertaining to proximal sampling algorithms can be found in section 3.2.1.

6.1 Introduction

As photons from distant sources (galaxies) travel through space-time to us here their trajectories are perturbed by local mass over and under-densities, causing the observed shapes of structures to be warped, or *gravitationally lensed*. This cosmological effect

---

[1] See chapter 5

is sensitive to all matter (both visible and invisible), and so provides a natural cosmological probe of dark matter. On large scales the lensing information is primarily Gaussian in nature, though on smaller scales (at higher resolutions) there becomes a non-negligible non-Gaussian contribution which encodes information about baryonic interactions and clustering amongst other non-linear effects. Analysis of such effects is expected (Munshi et al., 2008) to provide competitive and more importantly complementary constraints on cosmological parameters – in particular parameters related closely to dark matter such as σ_8 and Ω_M. Consequently, mapping techniques which preserve the non-Gaussian information content are a crucial step forward for dark matter analysis *via* weak gravitational lensing.

In chapter 5 we discussed a new sparse hierarchical Bayesian formalism for reconstructing the convergence field (Price et al., 2021a). This not only regularizes the ill-posed inverse problem (see section 2.1.2) but allows us to explore the Bayesian posterior in order to recover principled uncertainties on our reconstruction. It is important to note here that this mathematical framework is entirely general and can be applied for any posterior which belongs to the set of log-concave functions – of which both sparsity enforcing Laplace type priors and standard Gaussian priors are members. Often hierarchical Bayesian inference problems are solved by *Markov Chain Monte Carlo* (MCMC) techniques (see *e.g.* Trotta, 2017), which explicitly return a large number of samples from the full posterior distribution — from which one can construct true Bayesian uncertainties. Samples of the posterior *via* MCMC algorithms construct theoretically optimal estimates of the posterior (in the limit of a large number of samples), but in practice can be extremely computationally taxing to recover fully. In fact, when the dimensionality becomes large these methods become unfeasible — often referred to as *the curse of dimensionality*. In the context of lensing inverse problems each pixel constitutes a dimension, and so for a pixelisation of 1024×1024 (which is typical) the dimension of the problem is $\mathcal{O}(10^6)$.

Recent advancements in probability density theory (Robert, 2001) allow conservative approximations of Bayesian credible regions of the posterior from knowledge of the MAP solution alone (Pereyra, 2017). The sparse Bayesian method presented in previous work (see Price et al., 2021a) recasts the maximisation of the posterior distribution as a convex optimisation problem from which the *maximum a posteriori* (MAP) solution can be rapidly computed. Uncertainty quantification is then conducted utilizing the aforementioned approximate credible regions of the posterior. In Price et al. (2021a) hypothesis testing (determining the statistical significance of a feature of the recovered convergence map) was introduced to the weak lensing setting as a form of uncertainty quantification.

The remainder of this chapter is structured as follows. In section 6.2 we introduce the concept of *local credible intervals*, an additional form of uncertainty quantification, to the weak lensing setting. In section 6.3, we conduct a series of mock scenarios to compare the uncertainties recovered by our *maximum a posteriori* (MAP) approach, and the full MCMC (Px-MALA) treatment. Finally we draw conclusions in section 6.4. For the reader interested only in the application and benchmarking section 6.3 onwards is relevant content.

6.2 Local credible intervals

Local credible intervals can be interpreted as error bars on individual pixels or super-pixel regions (collection of pixels) of a reconstructed κ map. This concept can be applied to any method for which the HPD credible-region (and thus the approximate HPD credible-region) can be computed. Mathematically local credible intervals can be computed as follows (Cai et al., 2018b).

Select a partition of the κ domain $\Omega = \cup_i \Omega_i$ such that super-pixels Ω_i (*e.g.* an 8×8 block of pixels) are independent sub-sets of the κ domain $\Omega_i \cap \Omega_j = \varnothing, \forall \{i \neq j\}$. Clearly, provided the super-pixels Ω_i completely tessellate Ω they can be of arbitrary dimension. We define indexing notation on the super-pixels Ω_i via the index operator ζ_{Ω_i} which satisfy analogous relations to the standard set indicator function, *i.e.* $\zeta_{\Omega_i} = 1$ if the pixel of the convergence map κ belongs to Ω_i and 0 otherwise. For a given super-pixel region Ω_i we quantify the uncertainty by finding the upper and lower bounds $\xi^+_{\Omega_i}$, $\xi^-_{\Omega_i}$ respectively, which raise the objective function above the approximate level-set threshold ϵ'_α (or colloquially, 'saturate the HPD credible region C'_α'). In a mathematical sense these bounds are defined by

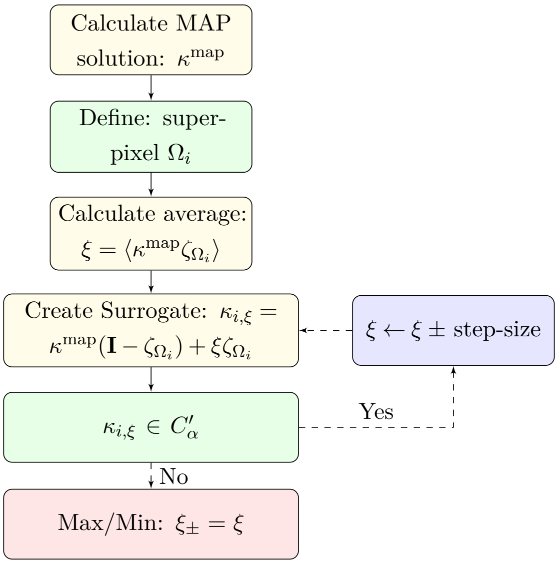


Figure 6.1: Schematic of the process to construct local credible intervals. At each iterative step the super-pixel region is uniformly increased (decreased) by a step-size. Once the level-set threshold ϵ'_α is saturated the iteration is terminated.

$$\xi^{\pm}_{\Omega_i} = \max_{\xi \in \mathbb{R}_{>0}} \left\{ \pm \xi | f(\kappa_{i,\xi}) + g(\kappa_{i,\xi}) \leq \epsilon'_\alpha \right\}, \tag{6.1}$$

where $\kappa_{i,\xi} = \kappa^{\text{map}}(\mathbb{I} - \zeta_{\Omega_i}) + \xi \zeta_{\Omega_i}$ is a surrogate solution where the super-pixel region has been replaced by a uniform intensity ξ. We then construct the difference image $\sum_i (\xi^+_{\Omega_i} - \xi^-_{\Omega_i})$ which represents the length of the local credible intervals (*cf.* error bars) on given super-pixel regions at a confidence of $100(1-\alpha)\%$. In this section we locate $\xi^{\pm}$ iteratively *via* bisection, though faster converging algorithms could be used to further increase computational efficiency. A schematic diagram for constructing local credible intervals is found in Figure 6.1. Conceptually, this is finding the

maximum and **minimum** constant values which a super-pixel region could take, at $100(1-\alpha)\%$ confidence — which is effectively Bayesian error bars on the convergence map.

In plain english, starting from the MAP convergence solution κ^{map} — at which all pixels are in positions which minimise the objective function — we then select a sub-set of the pixels, *e.g.* an 8×8 block of pixels. We start by averaging the pixels in the block which is selected. We then set the pixels within this block to the average value. Following this we iteratively raise/lower the now uniform value of the pixels within this block whilst keeping the rest of the image fixed. After each iteration we check if the surrogate solution (κ^{map} with the block of interest replaced by some constant value) is an acceptable solution, *i.e.* the objective function is below the threshold ϵ'_α. We find the values (upper and lower bounds) at which the objective function is equal to the threshold ϵ'_α. We then take the difference between these bounds, which is the *local credible interval* for a given 'block of interest' (super-pixel region).

6.3 Evaluation on simulations

For computing Bayesian inference problems one might ideally adopt MCMC sampling approaches as they are (assuming convergence) guaranteed to produce optimal results, however these approaches are computationally demanding and can often be computationally unfeasible. Therefore it is beneficial to adopt approximate but significantly computationally cheaper methods, such as the MAP estimation approach presented in chapter 5 — first presented in Price et al. (2021a).

However, the approximation error introduced through these approximate methods must be ascertained. Therefore, in this section we benchmark the uncertainties reconstructed *via* our MAP algorithm to those recovered by the *state-of-the-art* proximal MCMC algorithm, Px-MALA (Pereyra, 2016; Durmus et al., 2018). Additionally we compare the computational efficiencies of both approaches, highlighting the computational advantages provided by approximate methods. For simplicity and brevity throughout we will refer to any uncertainties recovered *via* our aforementioned *maximum a posteriori* reconstruction method as 'MAP uncertainties'. Additionally we will refer to the *maximum a posteriori* reconstruction method discussed throughout this chapter as the 'MAP algorithm'.

6.3.1 Datasets

We select four test convergence fields: two large scale Buzzard N-body simulation (DeRose et al., 2019) planar patches selected at random; and two of the largest dark matter halos from the Bolshoi N-body simulation (Klypin et al., 2011). This selection is chosen so as to provide illustrative examples of the uncertainty quantification techniques in both cluster and wider-field weak lensing settings.

Bolshoi N-body

The Bolshoi cluster convergence maps used were produced from 2 of the largest halos in the Bolshoi N-body simulation. These clusters were selected for their large total mass and the complexity of their substructure, as can be seen in Figure 6.3. Raw particle data was extracted from the Bolshoi simulation using CosmoSim[2], and was then gridded into 1024×1024 images. These images inherently contain shot-noise and so were passed through a multi-scale Poisson de-noising algorithm before being re-gridded to 256×256. The de-noising algorithm consisted of a forward Anscombe transform (to Gaussianise the noise), several TV-norm (total-variation) de-noising optimisations of different scale, before finally applying an inverse Anscombe transform. Finally, the images were re-scaled onto $[0, 1]$ — a similar de-noising approach for Bolshoi N-body simulations was adopted in related articles (see Lanusse et al., 2016).

Buzzard N-body

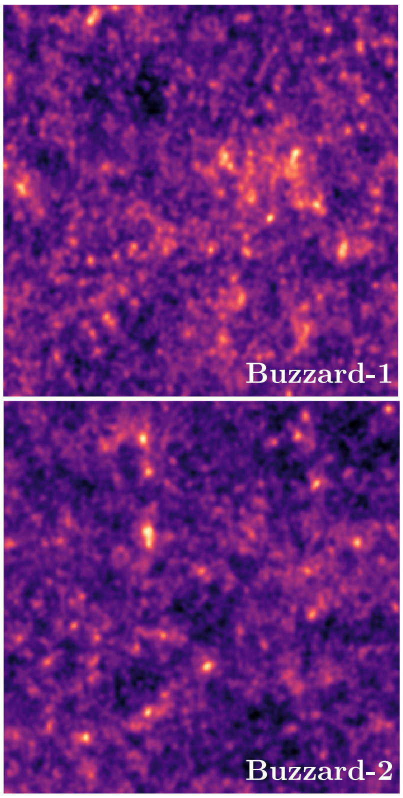


Figure 6.2: Two $\sim 1.2\,\text{deg}^2$ planar random extractions from the Buzzard V-1.6 N-body simulation catalog, each containing $\mathcal{O}(10^6)$ galaxies.

The Buzzard v-1.6 shear catalogs are extracted by ray-tracing from a full end-to-end N-body simulation. The origin for tracing is positioned in the corner of the simulation box. Access to the Buzzard simulation catalogs was provided by the LSST-DESC collaboration[3].

In the context of this chapter we restrict ourselves to working on the plane, and as such we extracted smaller planar patches. To do so we first project the shear catalog into a coarse HEALPix[4] (Gorski et al., 2005) gridding (with N_{side} of 16). Inside each HEALPix pixel we tessellate the largest possible square region, onto which we rotate and project the shear catalog. Here HEALPix pixelisation is solely used for its equal area pixel properties. After following the above procedure, the Buzzard v-1.6 shear catalog reduces to $\sim 3 \times 10^3$ planar patches of angular size $\sim 1.2\,\text{deg}^2$, with $\sim 4 \times 10^6$ galaxies per patch. In previous work (Price et al., 2021a) we utilized 60 of these realisations, but for the purpose of this analysis we select at random two planar re-

---

[2] `https://www.cosmosim.org`

[3] `http://lsst-desc.org`

[4] `http://healpix.sourceforge.net/documentation.php`

gions to study, which we project onto a 256×256 grid. These plots can be seen in Figure 6.2.

6.3.2 Method

To draw comparisons between our MAP uncertainties and those recovered *via* Px-MALA we conduct the following set of tests on the aforementioned datasets (see section 6.3.1). Initially we transform the ground truth convergence κ^{in} into a clean shear field γ^{in} by $\gamma^{\text{in}} = \Phi\kappa^{\text{in}}$. This clean set of shear measurements is then contaminated with a noise term n to produce mock noisy observations γ such that $\gamma = \gamma^{\text{in}} + n$. For simplicity we choose the noise to be zero mean i.i.d. Gaussian noise of variance σ_n^2, *i.e.* $n \sim \mathcal{N}(0, \sigma_n^2)$. In this setting σ_n is calculated such that the signal to noise ratio (SNR) is 20 dB (decibels) where

$$\sigma_n = \sqrt{\frac{\|\Phi\kappa\|_2^2}{N}} \times 10^{-\frac{\text{SNR}}{20}}. \tag{6.2}$$

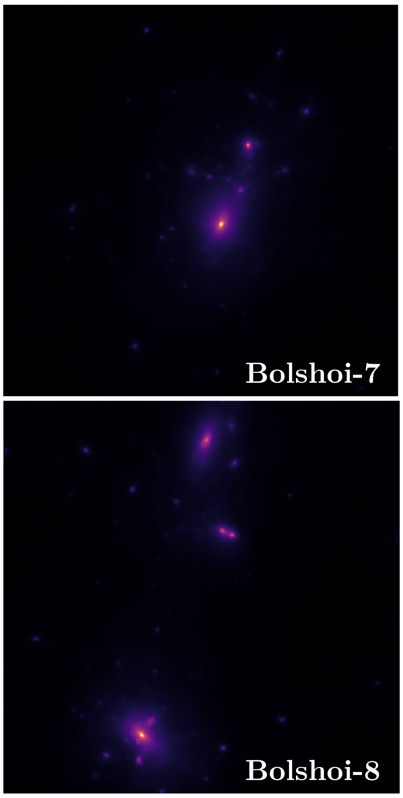


Figure 6.3: Two of the largest clusters extracted from the Bolshoi simulation database, labeled as Bolshoi 7 and 8 somewhat arbitrarily. In both cases at least one massive sub-halo is located within the FoF (friends of friends) sub-catalog.

Throughout this uncertainty benchmarking we use a fiducial noise level of 20 dB. For further details on how a noise level in dB maps to quantities such as galaxy number density and pixel size see Price et al. (2021a). The noise level of 20 dB considered here is somewhat optimistic (corresponding to between 30 and 100 galaxies per square arcmin for a band-limit of ∼400), which is appropriate for the purposes of benchmarking against MCMC simulations, which is the focus of the current discussion. Less optimistic simulations would simply increase the absolute level of the quantified uncertainties but not their relative level.

We then apply the reconstruction algorithm discussed in chapter 5, to recover *maximum a posteriori* convergence estimate κ^{map} (see section 3.2.2), along with the objective function — with regularisation parameter μ and noise variance σ_n^2. Adopting these quantities, and leveraging the statistical tools outlined in section 6.2, we conduct uncertainty quantifications on κ^{map}. To benchmark the MAP reconstructed uncertainties we first construct an array of *local credible interval* maps for super-pixel regions of sizes $[4, 8, 16]$ at 99%

confidence. These local credible interval maps are then compared to those recovered from the full MCMC analysis of the posterior (see section 3.2.1). We adopt two basic statistical measures to compare each set of recovered local credible interval maps: the Pearson correlation coefficient r; and the recovered SNR. The Pearson correlation coefficient between our MAP local credible interval map $\xi^{\text{map}} \in \mathbb{R}^{N'}$ and the Px-MALA local credible interval map $\xi^{\text{px}} \in \mathbb{R}^{N'}$, where N' is the dimension of the super-pixel space, is defined to be

$$r = \frac{\sum(\xi_i^{\text{map}} - \bar{\xi}^{\text{map}})(\xi_i^{\text{px}} - \bar{\xi}^{\text{px}})}{\sqrt{\sum(\xi_i^{\text{map}} - \bar{\xi}^{\text{map}})^2}\sqrt{\sum(\xi_i^{\text{px}} - \bar{\xi}^{\text{px}})^2}}, \tag{6.3}$$

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

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

$$\text{SNR} = 20 \times \log_{10}\left(\frac{\|\xi^{\text{px}}\|_2}{\|\xi^{\text{px}} - \xi^{\text{map}}\|_2}\right) \quad \text{and} \quad \text{RMSE} = 100 \times \left(\frac{\|\xi^{\text{px}} - \xi^{\text{map}}\|_2}{\|\xi^{\text{px}}\|_2}\right)\%, \tag{6.4}$$

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

Conceptually, the SNR roughly compares the absolute magnitudes of recovered local credible intervals and the Pearson correlation coefficient gives a rough measure of how geometrically similar the local credible intervals are. In this sense the closer r is to 1 the more similar the recovered local credible intervals are, and the higher the SNR the smaller the approximation error given by equation (10.12). Thus, a positive result is quantified by both large correlation and large SNR.

6.3.3 Numerical benchmarking results

As can be seen in Figures 6.4 and 6.5 the local credible intervals recovered through our sparse hierarchical Bayesian formalism are at all times larger than those recovered via Px-MALA — confirming that the uncertainties are conservative, as proposed in chapter 3, specifically section 3.2.3. Moreover, a strong correlation between the reconstructions can be seen.

The largest correlation coefficients r are observed for super-pixel regions of dimension 16×16 in all cases ($\langle r \rangle \approx 0.9$), peaking as high as 0.98 for the Buzzard 1 extraction — which constitutes a near maximal correlation, and thus an outstanding topological match between the two recovered local credible intervals. Additionally, in the majority of cases the recovered SNR is ≥ 10 dB — in some situations rising as high as ≈ 13 dB (corresponding to $\approx 20\%$ RMSE percent error) — which indicates that the recovered MAP uncertainties are close in magnitude to those recovered *via* Px-MALA. However, for super-pixels with dimension 4×4 the structural correlation between ξ^{map} and ξ^{px} becomes small — in one case becoming marginally negatively correlated. This is likely to be a direct result of the error given by equation (10.12) inherited from the definition of the approximate HPD credible region — as this approximation has the side-effect of smoothing the posterior hyper-volume, and for small super-pixels the hyper-volume is typically not smooth, thus the correlation coefficient r decreases.

We conducted additional tests for large 32×32 dimension super-pixels, which revealed a second feature of note. For particularly large super-pixel regions (32×32 or larger) the SNR becomes small for both Buzzard maps. This is a result of the assumption that within a super-pixel there exists a stable mean which is roughly uniform across the super-pixel. Clearly, for buzzard type data, on large scales this breaks down and so the recovered local credible intervals deviate from those recovered *via* Px-MALA. It is important to stress this is a breakdown of the assumptions made when constructing local credible intervals and not an error of the approximate HPD credible region. The numerical results are summarised in Table 6.1. Typically, structures of interest in recovered convergence maps cover super-pixel regions of roughly 8×8 to 16×16, and so for most realistic applications our MAP uncertainties match very well with those recovered through Px-MALA. In most situations weak lensing data is gridded such that it best represents the features of interest, and so structures of interest (by construction) typically fall within 8×8 to 16×16 dimension super-pixel regions for 256×256 gridded images — for higher resolution images the

Super Pixel	**Pearson Correlation**	**SNR (dB)**	**RMSE Error**
Bolshoi-7			
4x4	0.463	11.737	25.892 %
8x8	0.848	11.994	25.137 %
16x16	0.945	12.509	23.690 %
32x32	0.937917	11.6928	26.0233 %
Bolshoi-8			
4x4	-0.168	11.467	26.710 %
8x8	0.929	11.490	26.637 %
16x16	0.941	11.350	27.070 %
32x32	0.921363	10.2934	30.5724 %
Buzzard-1			
4x4	0.164	10.666	29.289 %
8x8	0.916	10.473	29.948 %
16x16	0.984	9.262	34.427 %
32x32	0.633823	2.34941	76.3009 %
Buzzard-2			
4x4	0.140	10.653	29.333 %
8x8	0.904	10.465	29.973 %
16x16	0.926	9.217	34.605 %
32x32	0.643482	2.60863	74.0574 %

Table 6.1: Comparisons between the local credible interval maps recovered *via* MAP and those recovered *via* Px-MALA. Note that larger super-pixels corresponds to coarser resolutions whereas smaller super-pixels leads to higher resolution. This is because the super-pixel size is the size of the groups of pixels used to tile the original image — therefore larger tiling components leads to fewer tiles, and therefore lower resolution.

structures of interest, and corresponding optimal super-pixels will follow a similar ratio.

Overall, we find a very close relation between the local credible intervals recovered through our MAP algorithm with those recovered *via* Px-MALA — a state-of-the-art MCMC algorithm. We find that MAP and Px-MALA local credible intervals are typically strongly topologically correlated (Pearson correlation coefficient ≈ 0.9) in addition to being physically tight (RMSE error of $\approx 20 - 30\%$). Moreover, we find that the MAP local credible intervals are, everywhere, larger than the Px-MALA local credible intervals, corroborating the assertion that the approximate HPD level-set threshold ϵ'_α is in fact conservative.

Px-MALA Time (s)	**MAP Time (s)**	**Ratio**
Buzzard-1		
133761	0.182	0.734×10^6
Buzzard-2		
141857	0.175	0.811×10^6
Bolshoi-7		
95339	0.153	0.623×10^6
Bolshoi-8		
92929	0.143	0.650×10^6

Table 6.2: Numerical comparison of computational time of Px-MALA and MAP. The MAP approach typically takes $\mathcal{O}(10^{-1})$ seconds, compared to Px-MALA's $\mathcal{O}(10^5)$ seconds. Therefore for linear reconstructions MAP is close to $\mathcal{O}(10^6)$ times faster.

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

6.4 Summary

In this chapter we introduce the concept of local credible intervals (*cf.* pixel-level error bars) — developed in previous work and applied in the radio-interferometric setting — to the weak lensing setting as an additional form of uncertainty quantification. Utilizing local credible intervals we validate the sparse hierarchical Bayesian mass-mapping formalism presented in Price et al. (2021a). Specifically we compare the local credible intervals recovered *via* the MAP formalism and those recovered *via* a complete MCMC analysis — from which the true posterior is effectively recovered.

To compute the asymptotically exact posterior we utilise Px-MALA — a *state-of-the-art* proximal MCMC algorithm. Using the local credible intervals; we benchmark the MAP uncertainty reconstructions against Px-MALA. Quantitatively, we compute the Pearson correlation coefficient (r, as a measure of the correlation between hyper-volume topologies), recovered signal to noise ratio and the root mean squared percentage error (SNR and RMSE, both as measures of how tightly constrained is the absolute error). We find that for a range of super-pixel dimensions the MAP and Px-MALA uncertainties are strongly topologically correlated ($r \geq 0.9$). Moreover, we find the RMSE to typically be $\sim 20-30\%$ which is tightly constrained when one considers this is a conservative approximation along each of at least $\mathcal{O}(10^3)$ dimensions. Additionally we compare the computational efficiency of Px-MALA and our MAP approach. In a 256×256 setting, the computation time of the MAP approach was $\mathcal{O}$(seconds) whereas the computation time for Px-MALA was $\mathcal{O}$(days). Overall, the MAP approach is shown to be $\mathcal{O}(10^6)$ times faster than the *state-of-the-art* Px-MALA algorithm.

A natural progression is to extend the planar sparse Bayesian algorithm to the sphere, which will be the aim of upcoming work — a necessity when dealing with wide-field stage IV surveys such as LSST[5] and EUCLID[6]. Additionally, we will expand the set of uncertainty quantification techniques to help propagate principled Bayesian uncertainties into the set of higher-order statistics typically computed on the convergence field.

---

[5] `https://www.lsst.org`

[6] `http://euclid-ec.org`

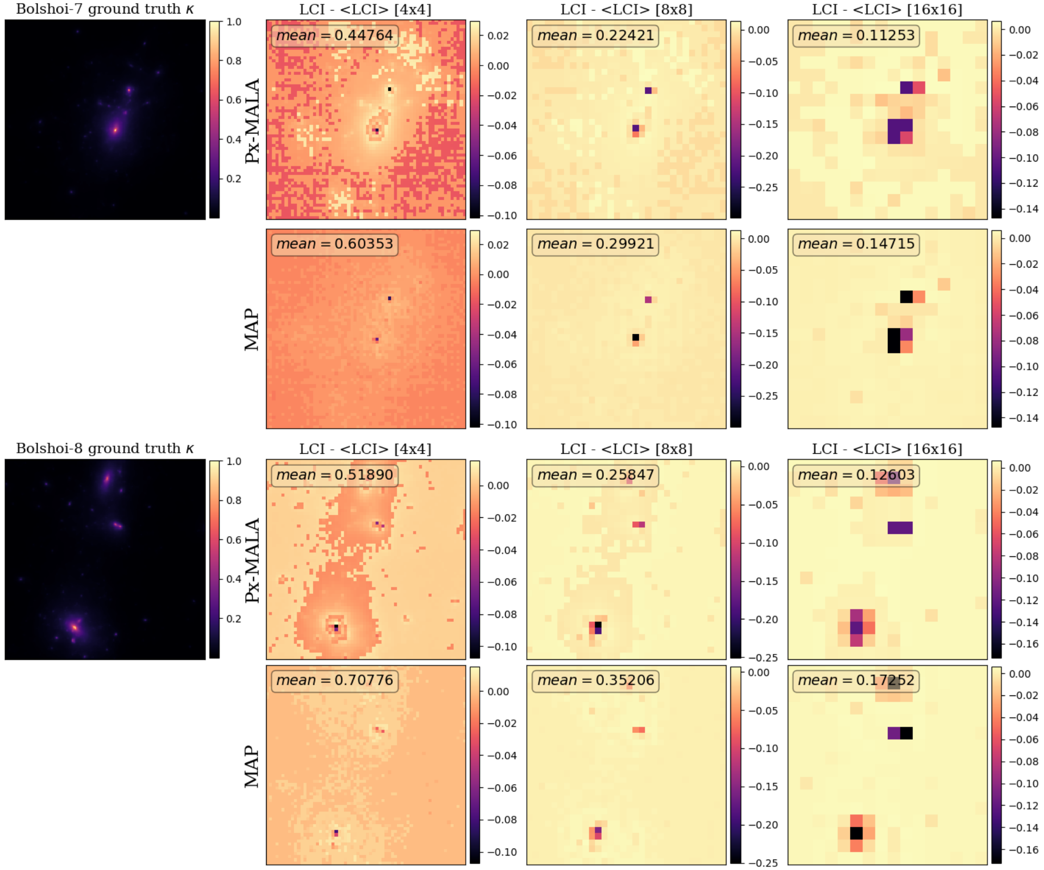


Figure 6.4: Local Credible Intervals (*cf. Bayesian error bars*) at 99% confidence for the Bolshoi-7 (**top**) and Bolshoi-8 (**bottom**) cluster sparse reconstruction in both the Px-MALA setting (*top*) and MAP (*bottom*) for super-pixel regions of dimension $(4 \times 4), (8 \times 8)$, and (16×16) — left to right respectively. Note that these plots display the variation about the mean of each set of LCI's, with the mean being given numerically in the sub-figure legends — this is done to best display the topological similarity whilst also conveying the absolute difference in size between the methods. Notice that the mean of the MAP LCIs is in all cases larger than that of the corresponding Px-MALA LCIs. Further note that the smaller the dimension of the super-pixel the larger the local credible interval which is because adjusting fewer pixels raises the objective function by less, and so the smaller super-pixels can be raised/lowered by more before saturating the level-set threshold. All numerical results are displayed in Table 6.1.

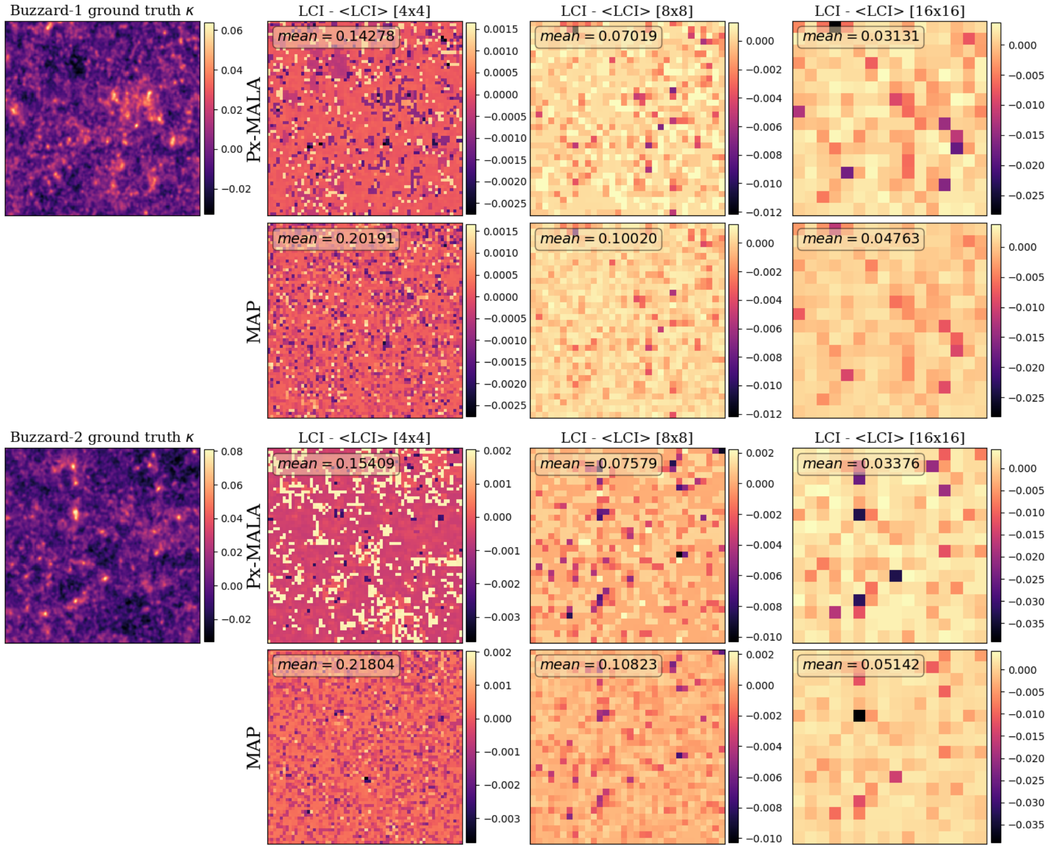


Figure 6.5: Local Credible Intervals (*cf. Bayesian error bars*) at 99% confidence for the Buzzard-1 (**top**) and Buzzard-2 (**bottom**) cluster sparse reconstruction in both the Px-MALA setting (*top*) and MAP (*bottom*) for super-pixel regions of dimension $(4 \times 4), (8 \times 8)$, and (16×16) — left to right respectively. Note that these plots display the variation about the mean of each set of LCI's, with the mean being given numerically in the sub-figure legends — this is done to best display the topological similarity whilst also conveying the absolute difference in size between the methods. Notice that the mean of the MAP LCIs is in all cases larger than that of the corresponding Px-MALA LCIs. Further note that the smaller the dimension of the super-pixel the larger the local credible interval which is because adjusting fewer pixels raises the objective function by less, and so the smaller super-pixels can be raised/lowered by more before saturating the level-set threshold. All numerical results are displayed in Table 6.1.

Chapter 7

Peak count and location uncertainties

This chapter is based on research presented in M. A. Price, J. D. McEwen, X. Cai, T. D. Kitching, and LSST Dark Energy Science Collaboration, "Sparse Bayesian mass mapping with uncertainties: peak statistics and feature locations", Monthly Notices of the Royal Astronomical Society, vol. 489, no. 3, pp. 3236-3250, Dec. 2019.

Price et al. (2019b)

Note: *This chapter assumes knowledge of the following. Mathematical context of weak gravitational lensing, presented in sections 4.1.4 and 4.2.2. A general understanding of higher-order convergence statistics (see* e.g. *Munshi et al., 2008; Lin, 2016; Munshi and Coles, 2017). This section adopts the dark matter reconstruction algorithm presented in chapter 5, with the approximate HPD-credible regions derived in section 3.2.3.*

In this chapter we propose two novel uncertainty quantification techniques, designed to answer two questions frequently asked of the recovered convergence map. The first of these questions asks where a feature of interest in the reconstructed convergence map could have been observed — typically this has been addressed by bootstrapping; however we can now infer it directly in a Bayesian manner. Drawing on the stone and pond analogy of chapter 2, this locational uncertainty quantification technique addresses the question of: what is the likely region over which the stone was initially dropped? The second question pertains to the weak lensing peak statistic (Lin, 2016), asking: given a magnitude threshold what is the maximum and minimum number of peaks which could have been observed, within some well defined confidence?

7.1 Introduction

In an empty universe the *null geodesics* along which photons travel correspond directly to straight lines. However, in the presence of a non-uniform distribution of matter the null geodesics are perturbed *via* gravitational interaction with the local matter over or under density, *i.e.* the photons are *gravitationally lensed* (Schneider, 2005; Munshi et al., 2008; Heavens, 2009; Grimm and Yoo, 2018). As this gravitational interaction is sensitive only to the total matter distribution, and the overwhelming majority of matter is typically dark, gravitational lensing provides a natural probe of dark matter itself (Clowe et al., 2006).

A wealth of information may be calculated directly from the shear field (often in the form of second order statistics (Kilbinger, 2015) — such as the power spectrum as in Alsing et al., 2016; Taylor et al., 2018) though recently there is increasing interest in extracting non-Gaussian information from the convergence field, *e.g.* peak statistics, Minkowski functionals, extreme value statistics (Coles and Chiang, 2000; Munshi and Coles, 2017; Peel et al., 2017b; Fluri et al., 2018; Peel et al., 2018). Primarily, the interest has arisen as higher-order statistics of the convergence field have been shown to provide complementary constraints on dark matter cosmological parameters which are typically poorly constrained by second-order statistics (Pires et al., 2010). However, to make principled statistical inferences from the convergence field, the inversion from γ to κ must be treated in a principled statistical manner — something which until recently was missing from convergence reconstruction algorithms which were either not framed in a statistical framework (see *e.g.* Kaiser and Squires, 1993; VanderPlas et al., 2011; Lanusse et al., 2016; Jeffrey et al., 2018; Wallis et al., 2021) or made assumptions of Gaussianity (see *e.g.* Corless et al., 2009; Schneider et al., 2015; Alsing et al., 2016). As the information of interest in higher-order convergence statistics is non-Gaussian, assumptions of Gaussianity in the reconstruction process severely degrade the quality of the cosmological information.

A mass-mapping framework was developed in chapter 5 which addressed precisely this issue. This new sparse hierarchical Bayesian mass-mapping formalism can be rapidly computed, can be extended to big data, and provides a principled statistical framework for quantifying uncertainties on reconstructed convergence maps (see Price et al., 2021a). Notably, it has been shown to accurately reconstruct very high dimensional Bayesian estimators, many orders of magnitude faster than *state-of-the-art* proximal MCMC algorithms. In chapter 6 we discuss a recent article (Price et al., 2019a), in which this algorithm was quantitatively benchmarked against Px-MALA (Pereyra, 2016; Durmus et al., 2018).

The structure of this article is as follows. In section 7.2 we develop a novel Bayesian inference approach to quantifying the uncertainty in reconstructed feature location, which we then showcase on illustrative N-body cluster simulation data in section 7.3. We then introduce a novel Bayesian inference approach for recovery of principled uncertainties on the aggregate peak count statistic in section 7.4. Following this we showcase this Bayesian inference approach to quantify uncertainty in the aggregate peak statistic in section 7.5 on N-body large scale structure (LSS) illustrative simulation data. Finally we draw conclusions in section 7.6.

7.2 Bayesian peak locations

Often one wishes to know the location of a feature of interest within the reconstructed convergence κ^{map}. Typically, this uncertainty is assessed *via* bootstrapping of the recovered image for a large number of simulated noise fields (as in *e.g.* Peel et al., 2017a). With the concept of approximate HPD credible regions in mind (see section 3.2.3), we propose a novel Bayesian approach to quantifying uncertainty in the peak location which we will refer to as the *'Bayesian location'*.

In essence the Bayesian location is computed as follows: A feature of interest is removed from the recovered convergence map, this feature is then inserted back into the convergence map at a new position to create a surrogate convergence map, if this surrogate map is within the approximate credible set then the position at which the feature was inserted cannot be rejected, if the surrogate is not in the approximate credible set then the position can be rejected. This process is computed for a sample of the total possible insertion positions, eventually providing an iso-contour of 'acceptable' positions. This iso-contour, at a well-defined confidence level, is the Bayesian location.

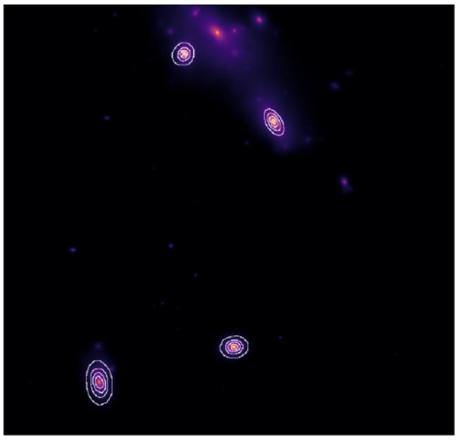


Figure 7.1: Combined plot of the 99% confidence Bayesian locations at SNR $= 12, 15, 17, 20$ dB. The outer rings represent the noisier position iso-contours whereas as the data becomes cleaner the iso-contour ring becomes smaller (therefore the rings represent iso-contours at SNR $= 12, 15, 17, 20$ dB, from the outer rings inwards respectively). N-splitting Circular Bisection (see section 7.2.2) was used to efficiently compute each iso-contour. For input SNR's below ~ 10 the smaller local features cannot be determined to be physical *via* the initial hypothesis test, hence the truncation at SNR $= 12$.

7.2.1 Bayesian locational uncertainties

Suppose we recover a (MAP) convergence field κ^{map} *via* optimisation of the objective function defined in equation (9.7) which contains a feature of interest (*e.g.* a large peak). Let us define the sub-set of pixels which contain this feature to be $\Omega_{\mathcal{Z}} \subset \Omega$, where Ω is the entire image domain. To begin, extract the feature $\mathcal{Z} = \kappa^{\text{map}} \mathbb{I}_{\Omega_{\mathcal{Z}}}$, *i.e.* a convergence field which contains only the feature of interest. Now we adopt the process of *segmentation inpainting* (Cai et al., 2018a,b; Price et al., 2021a) to create a convergence field realization without the feature of interest $\mathcal{Z}$, but instead with background signal replaced in $\Omega_{\mathcal{Z}}$. Mathematically segmentation inpainting is represented by the iterations

$$\kappa^{(t+1),\text{sgt}} = \kappa^{\text{map}} \mathbb{I}_{\Omega \setminus \Omega_{\mathcal{Z}}} + \Lambda \text{soft}_{\lambda}(\Lambda^{\dagger} \kappa^{(t),\text{sgt}}) \mathbb{I}_{\Omega_{\mathcal{Z}}}, \tag{7.1}$$

where Λ is an appropriately selected dictionary. To this end, we simply use the Daubechies 8 (DB8) wavelet dictionary with 8-levels and λ is the soft-thresholding parameter.

Following the wavelet inpainting, in order to separate the true feature from the background residual convergence the signal which was inpainted into the region $\Omega_{\mathcal{Z}}$ is subtracted from the extracted feature $\mathcal{Z}$, thus effectively accounting for the residual background signal which would likely have been present even in the absence of the feature $\mathcal{Z}$. At this junction the surrogate convergence κ^{sgt} is hypothesis tested for physicality (Cai et al., 2018b; Price et al., 2021a). If a feature is not found to be physical, the algorithm terminates at this point as, fundamentally, it is illogical to evaluate the uncertainty in position of an object for which existence cannot be statistically determined.

Now that we have successfully isolated $\mathcal{Z}$ we can insert it back into the surrogate field κ^{sgt} at a perturbed position. It is then sufficient to check if

$$f(\kappa^{\text{sgt}\prime}) + g(\kappa^{\text{sgt}\prime}) = h(\kappa^{\text{sgt}\prime}) \leq \epsilon^{\prime}_{\alpha}, \tag{7.2}$$

where $\kappa^{\text{sgt}\prime}$ represents the surrogate with the feature $\mathcal{Z}$ inserted at a perturbed location, and $h(\kappa)$ is the objective function discussed in chapter 5. Notice that this is simply a Bayesian hypothesis test, as outlined in chapter 5. In fact, all MAP uncertainty quantification techniques discussed in this thesis are hypothesis tests of varying complexity. If the inequality does hold, then the conclusion is that at $100(1-\alpha)\%$ confidence we cannot say that the feature could not be found at this location. If the equality does not hold then $\mathcal{Z}$ in its observed form could not have been found at the new location at $100(1-\alpha)\%$ confidence. The question then becomes, which perturbed positions are acceptable and which are not.

With the above in mind, we propose a typical inverse nested iterative scheme to determine the pixel-space iso-contour for a given feature in the reconstructed convergence field. Schematically this iterative process is outlined in Figure 7.2. Essentially, inverse nesting is: start in a ring 1 pixel from the MAP peak location in the first iteration, moving the ring one pixel outwards after each iteration.

7.2.2 N-splitting circular bisection

Inverse nested iterations are sufficiently fast for low-dimensional reconstructions (256×256), however as the dimensionality of the reconstructed domain grows it becomes increasingly beneficial to adopt more advanced algorithms to compute the Bayesian location in an efficient manner. We propose a novel iterative algorithm for computing the pixel-space position iso-contour which we call *N-splitting Circular Bisection* (N-splitting), the full details of which can be found in appendix A.1. A brief outline of N-splitting is given below. Suppose we wish to compute positions on the Bayesian location iso-contour at *equiangular intervals* $\Delta\Theta$ (defined clockwise

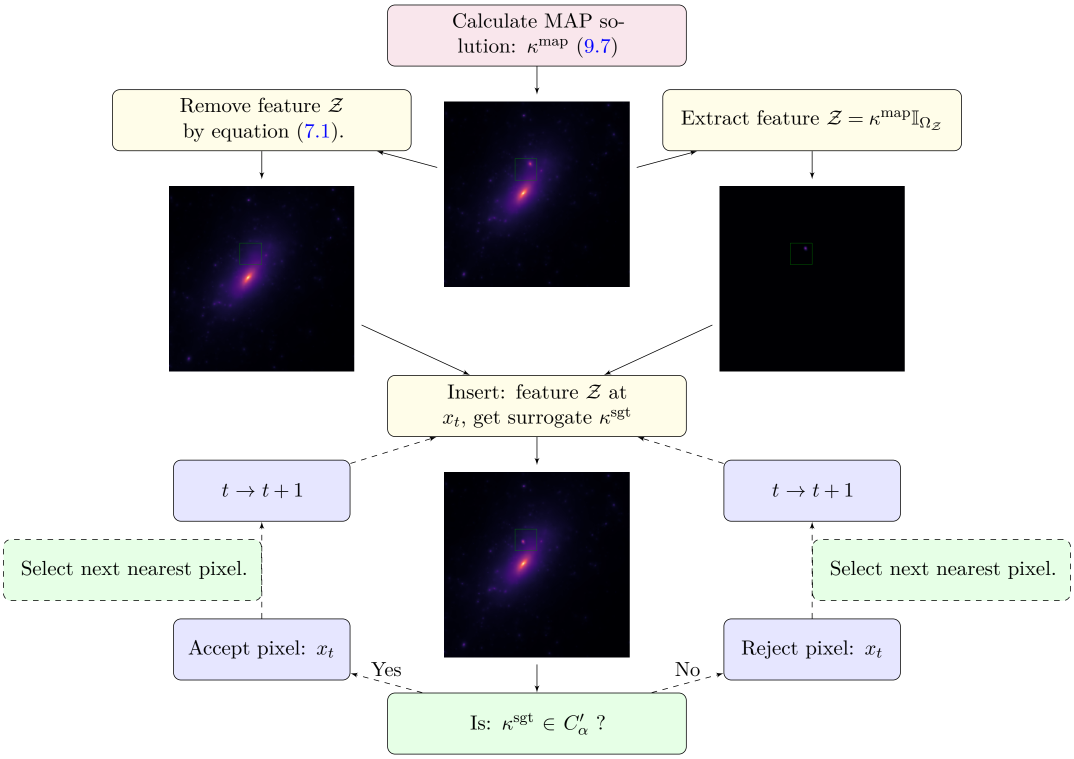


Figure 7.2: Schematic representation of the inverse nested iterations to determine the *Bayesian location* (see section 7.2). The Bayesian location is a positional uncertainty on a feature of interest $\mathcal{Z}$ within a recovered convergence field. Once a complete ring of pixels has been rejected the algorithm returns a binary map of accepted pixels which we call the Bayesian location. Any pixel outside of this location is rejected at $100(1-\alpha)\%$ confidence. Alternately the probability iso-contour bounding the set of acceptable pixels can be located by N-splitting circular bisection as described in section 7.2.2 and Appendix A.1.

about the peak location) relative to the y-axis. Then we require $n = 2\pi/\Delta\Theta$ sampling angles which are trivially given by $\Theta_i = i\Delta\Theta$, where i is an iterative factor which sets the angle for a given direction Θ_i.

Once Θ_i is defined for a single direction, the distance d'_α along direction Θ_i such that the objective function saturates the level-set threshold ϵ'_α is found by bisection. Mathematically, this is formally defined to be,

$$d^i_\alpha = \min_{d \in \Gamma_i} \Big[\, d \mid h(\kappa^{\text{sgt}}_d) > \epsilon'_\alpha \, \Big] \quad \text{for} \quad \Gamma_i = \Big[q_1 \sin(\Theta_i), q_2 \cos(\Theta_i) \mid q_{1,2} \in \mathbb{R}_{>0} \Big], \qquad (7.3)$$

where $d^2 = q_1^2 + q_2^2$, the sub-space $\Gamma_i \subset \mathbb{R}^2$ lies on the line of infinite extent along a given direction Θ_i sourced at the peak location, and κ^{sgt}_d is the surrogate convergence map constructed by inserting the feature of interest $\mathcal{Z}$ into a perturbed location $[q_1 \sin(\Theta_i),\, q_2 \cos(\Theta_i)]$. Once a representative set of positions on the location iso-contour are computed, the *convex hull* is taken – the convex hull is simply the smallest convex set which contains all samples of the location iso-contour (see section 2.2.1). The envelope of this closed convex set of acceptable pixels is taken as the Bayesian location.

7.3 Illustrative example of the Bayesian location

In this section we perform sparse Bayesian reconstructions of a large cluster extracted from the Bolshoi N-body simulation (Klypin et al., 2011; Lanusse et al., 2016), upon which we construct and assess the performance of Bayesian locations for each of the four largest sub-halos. In line with previous work of Price et al. (2021a) and in the related article of Lanusse et al. (2016) we refer to this extracted cluster as Bolshoi-3. We grid the Bolshoi convergence field onto a discretized complex map of dimension (1024×1024) so as to demonstrate that the sparse Bayesian approach can construct Bayesian estimators efficiently even when the dimension of the problem is in excess of $\mathcal{O}(10^6)$, *i.e.* dimensions at which MCMC techniques can become highly computationally challenging.

7.3.1 Method

First, we construct a complex discretized set of artificial shear measurements $\tilde{\gamma} \in \mathbb{C}^M$ by $\tilde{\gamma} = \Phi\kappa$, where $\kappa \in \mathbb{C}^M$ is the input Bolshoi-3 convergence map. Notice that for simplicity we restrict ourselves to the situation in which $\dim(\tilde{\gamma}) = \dim(\kappa)$, *i.e.* the measurements process is lossless. We then contaminate these mock measurements with noise n, which for simplicity we select to be *i.i.d.* Gaussian noise $n \sim \mathcal{N}(0, \sigma_n^2)$ of zero mean and variance σ_n^2. The variance is selected such that the SNR of the noisy artificial shear maps can be varied, and is therefore set to be,

$$\sigma_n = \sqrt{\frac{\|\Phi\kappa\|_2^2}{N}} \times 10^{-\frac{\text{SNR}}{20}}. \tag{7.4}$$

The MAP convergence field κ^{map} is recovered *via* the sparse Bayesian mass-mapping algorithm using DB10 wavelets (10-levels), and the Bayesian location for the set of 4 peaks is constructed.

7.3.2 Analysis and computational efficiency

To demonstrate this uncertainty quantification technique we construct 99% confidence Bayesian locations for the 4 largest sub-halos in the Bolshoi-3 cluster, for input SNR in decibels (dB) of $\in [12, 15, 17, 20]$. In Figures 7.1 and 7.3 it is apparent that, as expected, the positional uncertainty iso-contour at 99% confidence is smaller for less noisy data, growing in proportion to the noise. In our analysis 32 N-splitting directions (pointings) were used, though as can be seen in Figures 7.1 and 7.3 as few as 16 directions would easily have been sufficient given the smoothness of the iso-contour. Computationally, reconstruction of the MAP convergence field and computation of the Bayesian location for the complete set of peaks took ~ 5 minutes on a standard 2016 MacBook Air. Notably, this is a conservative (Pereyra, 2017) and tight (Price et al., 2019a) approximate Bayesian inference in an over 10^6-dimensional space on a personal laptop in minutes. Further to this, the sparse Bayesian algorithmic structure is easily parallelisable and so this computational efficiency can be optimized further.

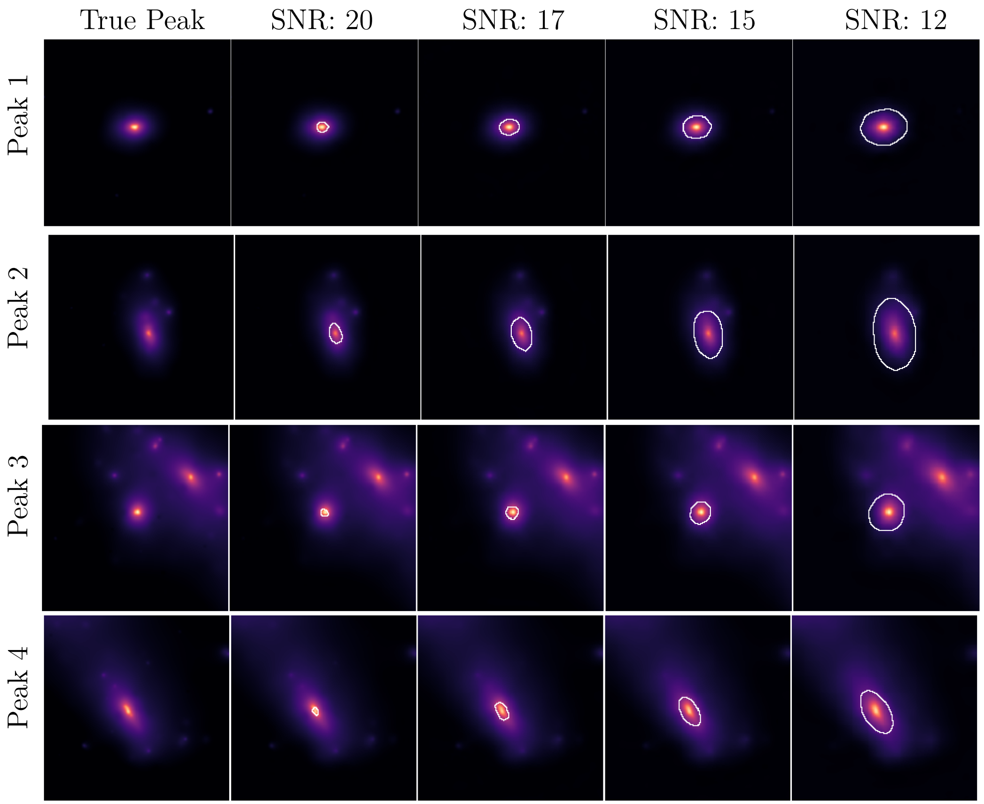


Figure 7.3: Left to right: Sparse Bayesian reconstructions of Bolshoi-3 peaks 1 to 4 (*top to bottom* respectively) followed by Bayesian locations (see section 7.2) at 99% confidence for input SNR of 20.0 to 12.0 dB— which are overlaid on the sparse Bayesian MAP recovered convergence maps κ^{map} at the corresponding SNR level. As the input artificial shear becomes more contaminated with noise, the relative information content decreases, and so inferred uncertainty of the reconstructed peak positions increases, as one would logically expect. Note the asymmetry in the 99% iso-contour, which motivates the N-splitting searching algorithm (see section 7.2.2 and Appendix A) rather than a naive circular inference (*e.g.* finding the maximal x and y displacements and assuming a circular iso-contour). Finally, observe that the 99% iso-contour for Peaks 3 and 4 are proportionally more tightly constrained than Peaks 1 and 2. This is due to the local information density typically being higher in more signal dense regions — perturbations to pixels in more information dense regions are more tightly constrained and can therefore move less distance before saturating the approximate level-set threshold ϵ'_{α}. This effect has been observed in the context of *local credible intervals* as presented in Cai et al. (2018b) and introduced to the weak lensing setting in Price et al. (2019a).

7.4 Aggregate uncertainty in peak counts

Building on the notion of an approximate HPD credible region discussed in section 3.2.3 we now ask the question: given a reconstructed convergence field κ^{map}, and at a given SNR threshold K, what is the maximum and minimum peak count at $100(1-\alpha)\%$ confidence. In this analysis, we choose to define a *peak* in κ^{map} by a pixel $\kappa^{\text{map}}(x)$, for $x \in \mathbb{R}^2$ which is larger than the 8 pixels which surround it (as in Lin, 2016). A point of the peak statistic is computed as follows: A threshold $K \in \mathbb{R}_{>0}$ is taken on κ^{map}, and the *peak count* (number of peaks which have intensity larger than K) is taken on the sub-set of pixels larger than the threshold.

Formally we define the *excursion set* $\Omega^+ \subset \Omega$ as $\Omega^+ = \{\, x \mid \kappa^{\text{map}}(x) > K \,\}$, where Ω is the complete set of recovered pixels. Define a further sub-set $\Pi \subset \Omega^+$ as the set of peaks in Ω^+:

$$\Pi(\kappa^{\text{map}}) = \Big\{\, x \mid \kappa^{\text{map}}(x) > \kappa^{\text{map}}(x'),\, \forall\, x' \in \mathcal{N}(x) \,\Big\}, \tag{7.5}$$

where $\mathcal{N}(x)$ represents the set of immediately surrounding pixels. Note that this definition is not valid for pixels on the boundary of the field, and so these pixels are necessarily not considered. This not only excludes the outer edge of κ^{map} but also any pixels surrounding masked regions (of which there are typically many). Conceptually, we would then like to know at a given threshold K what is the maximum and minimum number of peaks which could exist such that the *surrogate solution* κ^{sgt} still belongs to the approximate HPD credible set C'_α. Let $\eta^{\max}_\alpha$ be the *upper bound* on the number of peaks, and $\eta^{\min}_\alpha$ be the *lower bound* on the number of peaks, for a given threshold K, at $100(1-\alpha)\%$ confidence. Further let $\eta \equiv |\Pi(\kappa^{\text{map}})| \in \mathbb{R}_{>0}$ be the number of peaks calculated from the MAP solution κ^{map} at threshold K. Formally these quantities are given by,

$$\eta^{\min/\max}_\alpha \equiv \min_{\kappa^{\text{sgt}} \in \mathbb{C}^M}/\max \Big\{\, |\Pi(\kappa^{\text{sgt}})| \in \mathbb{R}_{>0} \,|\, h(\kappa^{\text{sgt}}) \leq \epsilon'_\alpha \,\Big\}, \tag{7.6}$$

where $|\Pi(\kappa)|$ is the *cardinality of the peak set* of a convergence field κ. It is not at all obvious how one should locate the extremum of such optimisation problems as they are inherently non-linear, non-convex problems. We can, however, propose a logical iterative approach to calculate well motivated approximations to the upper and lower peak count limits $\eta^{\max}_\alpha$ and $\eta^{\min}_\alpha$.

7.4.1 Approximate Bayesian lower bound on peak counts

It is perhaps conceptually more straightforward to minimise the cardinality of the peak set and so we will first describe this process. To calculate an approximate bound on $\eta^{\min}_\alpha$ we begin by creating the initial peak set Π from the recovered convergence κ^{map}. The peak in $\Pi(\kappa^{\text{map}})$ with lowest magnitude is located. The shortest distance $r_{\min}$ from the pixel location x to a pixel x' such that $\kappa^{\text{map}}(x') = y$ (where y is some magnitude at which it is assumed the peaks influence is sufficiently small) is computed in Euclidean space as $r_{\text{min}} = |x - x'|$. For the current discussion we simply set $y = 0$.

Let us define the region of interest $\Omega_{\mathcal{A}} \subset \Omega$ to be a circular aperture centred on the peak pixel location x with radius r_{min}. Conceptually, this acts as a proxy for the area effected by a large over-density sourced at the location of the peak. Now, define a *down-scaling kernel* $\mathcal{S}_{K,\Omega_{\mathcal{A}}} \in \mathbb{C}^{N \times N}$ which has the action of scaling the magnitude of the sub-set $\kappa^{\text{map}}\mathbb{I}_{\Omega_{\mathcal{A}}}$ of pixels belonging to the region of interest $\Omega_{\mathcal{A}}$ onto $[0, K]$. Application of the down-scaling operator returns a surrogate solution κ^{sgt}. Mathematically this is the isometric transformation

$$\kappa^{\text{sgt}} = \mathcal{S}_{K,\Omega_{\mathcal{A}}}(\kappa^{\text{map}}) = \kappa^{\text{map}}\mathbb{I}_{\Omega \backslash \Omega_{\mathcal{A}}} + \frac{K}{\max\left(\kappa^{\text{map}}\mathbb{I}_{\Omega_{\mathcal{A}}}\right)}(\kappa^{\text{map}}\mathbb{I}_{\Omega_{\mathcal{A}}}). \tag{7.7}$$

Application of $\mathcal{S}_{K,\Omega_{\mathcal{A}}}$ to an isolated region $\Omega_{\mathcal{A}}$ conserves the local topology of the field, *i.e.* is an isometric transformation over $\Omega_{\mathcal{A}}$, which is desirable as it means we are making no assumptions about the halo profile around a given peak.

Removal of a peak through the action of $\mathcal{S}_{K,\Omega_{\mathcal{A}}}$ creates a surrogate solution κ^{sgt} which is likely to minimise the increase in the objective function, at least in a decision theoretic sense. As such $\mathcal{S}_{K,\Omega_{\mathcal{A}}}$ is a good stratagem for excluding peaks from $\Pi(\kappa^{\text{map}})$ as it will likely maximise the number of peaks which can be removed from $\Pi(\kappa^{\text{map}})$ before the level-set threshold ϵ'_{α} is saturated. Thus, it will likely be near decision-theoretically optimal at minimizing equation (7.6), which is precisely what we want.

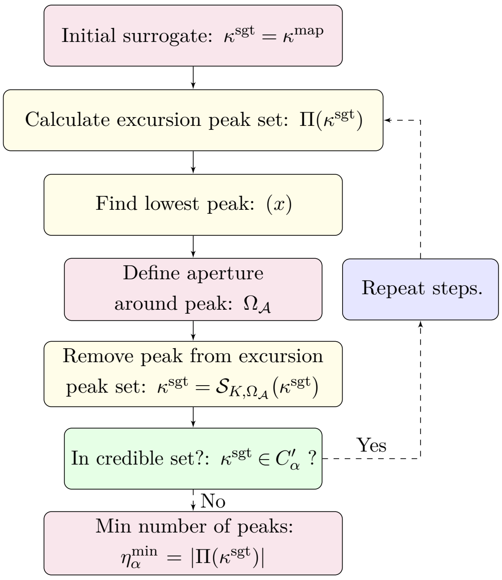


Figure 7.4: Schematic representation of the iteration steps in finding the Bayesian lower bound $\eta^{\text{min}}_{\alpha}$ at confidence $100(1-\alpha)\%$ of the peak count $|\Pi|$ for a given MAP reconstruction κ^{map}.

A schematic of the iterative process proposed to find the Bayesian lower bound on the peak statistic can be seen in Figure 7.4. In words, the process is as follows. Within each iteration, the lowest intensity peak within the peak set is removed forming a new surrogate convergence field κ^{sgt}, the objective function is recalculated and if the objective function is below the approximate level-set threshold ϵ'_{α} then the lowest peak within κ^{sgt} is now removed, so on and so forth until the objective function rises above ϵ'_{α}, at which the iterations are terminated and the minimum number of peaks is recovered.

7.4.2 Approximate Bayesian upper bound on peak counts

Now we invert our perspective in order to approximate the maximum number of peaks which could be observed at a given threshold K at $100(1-\alpha)\%$ confidence.

Here we will be considering the non-linear maximisation problem constructed in equation (7.6).

First, we introduce the notion of the *inclusion set* Ω^-, conjugate to Ω^+ such that $\Omega^- \cup \Omega^+ \equiv \Omega$ and $\Omega^- \cap \Omega^+ = \varnothing$, defined by $\Omega^- = \{x | \kappa^{\text{map}}(x) \leq K\}$. With this in mind, we can now cast the maximisation problem into a minimisation problem analogous to that used before. We now wish to minimise the number of peaks that belong to the *inclusion set* Ω^- which is by definition equivalent to maximizing the number of peaks which belong to the *excursion set* Ω^+, which is precisely what we want. Analogously to section 7.4.1 to construct our approximate bound we calculate the further sub-set $\Pi^- \subset \Omega^-$ which is defined similarly to the relation in equation (7.5) such that,

$$\Pi^-(\kappa^{\text{map}}) = \Big\{ x \mid \kappa^{\text{map}}(x) > \kappa^{\text{map}}(x'), \, \forall \, x' \in \mathcal{N}(x) \Big\}, \tag{7.8}$$

i.e. the sub-set of peaks below a threshold K. In contrast to section 7.4.1 we now locate the largest peak in Π^-. Suppose that this peak is found at $\Pi^-(x)$, we now construct a circular aperture about x with radius r_{min} as defined before. Let this circular aperture set of pixels be $\Omega_{\mathcal{A}} \subset \Omega$.

Now we define an *up-scaling kernel* $\mathcal{S}^{\dagger}_{K,\Omega_{\mathcal{A}}} \in \mathbb{C}^{N \times N}$ which has action,

$$\mathcal{S}^{\dagger}_{K,\Omega_{\mathcal{A}}}(\kappa^{\text{map}}) = \kappa^{\text{map}} \mathbb{I}_{\Omega \setminus \Omega_{\mathcal{A}}} + \frac{K + \Delta}{\max\left(\kappa^{\text{map}} \mathbb{I}_{\Omega_{\mathcal{A}}}\right)} (\kappa^{\text{map}} \mathbb{I}_{\Omega_{\mathcal{A}}}) \tag{7.9}$$

which deviates from the the down-scaling operator in the numerator of the second term. Here Δ is an infinitesimal quantity added such that the re-scaled peak within $\Omega_{\mathcal{A}}$ falls infinitesimally above the threshold K and is therefore counted as a peak. In practice we set Δ to be $\sim 10^{-5}$ and find that adjusting this quantity by $\mathcal{O}(10^2)$ has negligible effect on the recovered uncertainties.

With these conceptual adjustments we then follow the same iterations discussed in section 7.4.1 to find the approximate Bayesian upper bound on the peak count $\eta^{\text{max}}_{\alpha}$. Schematically this is given in Figure 7.5. Finally we return the tuple $(\eta^{\text{min}}_{\alpha}, \eta, \eta^{\text{max}}_{\alpha})$ which is in the form (minimum, most likely, maximum) at $100(1-\alpha)\%$ confidence.

7.4.3 Limitations of re-scaling

Suppose the SNR threshold K is large enough such that during iterations in schematic of Figure 7.4 the cardinality of the excursion peak set $|\Pi(\kappa^{\text{sgt}})| \rightarrow 0$. In this situation even though the approximate level-set threshold ϵ'_{α} is not saturated, the algorithm is forced to stop as there are simply no more peaks to exclude (push down). At this point the strategy for removing peaks becomes locally ill-defined. Effectively this is a clipping artifact. To avoid this effect entirely, if $|\Pi(\kappa^{\text{sgt}})| = 0$ at any point within the iterations at a given threshold, the lower bound $\eta^{\text{min}}_{\alpha}$ at threshold K is set to 0, *i.e.* we are infinitely uncertain by construction.

Analogously, consider the case when K is small enough that during the iterations in Figure 7.5 the cardinality of the inclusion peak set $|\Pi^{-}(\kappa^{\text{sgt}})| \rightarrow 0$. In this situation there are simply no more peaks to include (pull up). Again we remove this clipping effect by setting η_α^{max} at threshold K to $|\Pi(\kappa^{\text{sgt}})|$. Typically these clipping effects only occur for very small $K \leq 2$ or very large $K \geq 8$ thresholds, and so a wealth of information can be extracted from the intervening scales. Low thresholds clip the upper limit η_α^{max} as the cardinality of the peak set drops to 0 quickly, but the objective function rises comparatively slowly, as this SNR range is statistically dominated by noise. High thresholds clip the lower limit η_α^{min} simply due to the inherently low count of peaks at high SNR thresholds.

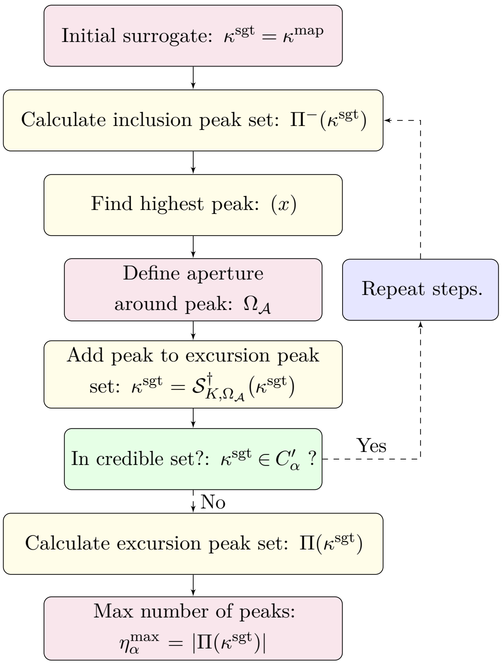


Figure 7.5: Schematic representation of the iteration steps in finding the Bayesian upper bound η_α^{max} at confidence $100(1-\alpha)\%$ of the peak count $|\Pi|$ for a given MAP reconstruction κ^{map}.

Further to this, the decision-theory approach adopted here for locating the maximal and minimal values of the cumulative peak statistic is based on several assumptions: removing lower peaks increases the objective function by less than larger peaks; the extent of a peak (dark matter over-density) is approximated by a circular aperture; and removal of a peak has little to no effect on locations in the image domain which are outside of this aperture. All three of these assumptions are very reasonable.

Although further computational optimisations are not an immediate concern since our approach is already highly computationally efficient, we acknowledge that this iterative approach for removing peaks can easily be formulated as a bisection style problem which is likely to drastically reduce the computation time further — particularly for low thresholds, as it mitigates the number of trivial noise peak removal recalculations which are done in the brute force approach presented above. In future, should computational efficiency become of primary interest this speed up will be considered.

7.5 Illustrative example of peak uncertainties

In this section we apply the sparse Bayesian mass-mapping pipe-line to high resolution (2048×2048) convergence maps extracted from the Buzzard V-1.6 N-body[1] simulation, upon which we construct the cumulative peak statistic (number of peaks above a threshold as a function of the threshold). Additionally, we recover the 99%

---

[1] Obtained due to our affiliation with the LSST-DESC collaboration.

approximate Bayesian constraints on the peak count at each threshold, from which we infer the 68% constraint so as to aid the reader in comparison to typical 1σ error-bars quoted in related literature.

7.5.1 Simulated data-sets

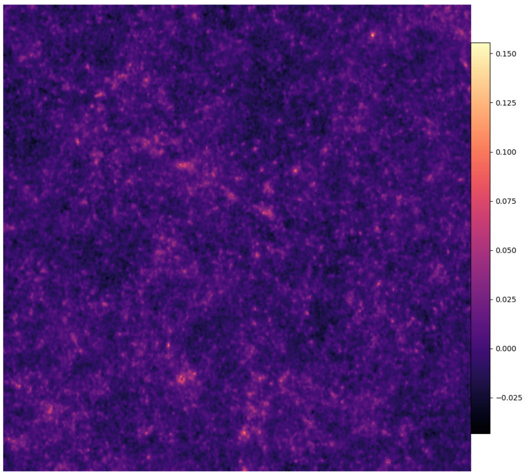


Figure 7.6: Input 2048×2048 convergence map extracted from the Buzzard N-body simulation.

The Buzzard V-1.6 N-body simulation convergence catalog (DeRose et al., 2019) has a quarter sky coverage and is extracted by full ray-tracing. For wide-fields the *flat sky approximation* breaks down (Wallis et al., 2021) and so this quarter sky coverage was reduced to smaller planar patches. The complete quarter sky convergence catalog was projected into a coarse HEALPix[2] (Gorski et al., 2005) pixelisation ($N_{\text{side}} = 4$). Inside of each pixel, we further tessellated the largest square region which we then project into a 2048×2048 grid. These gridded convergence maps formed our ground truth, discretized convergence fields. As HEALPix samples in such a way as to provide equal area pixels, and the Buzzard simulation galaxy density is fairly uniform, each extracted square region contained $\sim 2 \times 10^7$ galaxies leading to ~ 5 galaxies per pixel.

Due to a comparatively low density of samples, Poisson noise is prevalent, and thus extracted planar regions were passed through a multi-scale Poisson de-noising algorithm. This consisted of a forward Anscombe transform (in order to Gaussianise the Poisson noise), several TV-norm (total-variation) de-noising optimisations of differing scale, followed by an inverse Anscombe transform (as in Lanusse et al., 2016; Price et al., 2019a). A more involved treatment could be applied, but this approach is sufficient to demonstrate our peak reconstructions.

7.5.2 Application to Buzzard V-1.6

We select at random one of many planar patches produced for the following application. Following the methodology presented in section 9.3.2 we generate an artificial shear catalog which we then contaminate with independent and identically distributed (i.i.d.) Gaussian noise such that the SNR of mock shear measurements is 30 dB – *i.e.* an idealized noise-level simply for illustrative purposes.

The MAP convergence estimator κ^{map} is recovered from these noisy mock shear measurements *via* our sparse Bayesian mass-mapping framework. From κ^{map} we then calculate $\sigma^2 = \langle (\kappa^{\text{map}})^2 \rangle$ which we then use as a measure of the noise-level in

---

[2] `http://healpix.sourceforge.net/documentation.php`

the reconstructed convergence field. Implementing the uncertainty quantification technique presented in section 7.4 we then construct the cumulative peak statistic for SNR thresholds $K \in [2\sigma, 8\sigma)$ at increments of 0.25σ with upper and lower 99% approximate Bayesian confidence limits. Figure 7.7 displays the recovered cumulative peak statistic in both a linear and logarithmic scale. Typically, similar figures in the literature will quote 1σ error-bars, and so for comparisons sake we convert the Bayesian 99% confidence limits into the 68% confidence limits which are comparable to 1σ constraints (in Figure 7.7 we provide both confidence limits for clarity).

Complete reconstruction of the peak statistics for 24 threshold bins, each with approximate Bayesian upper and lower bounds, for a 2048×2048 resolution convergence map, with DB11 (11-level) wavelets, took ~ 2 hours on a 2016 MacBook Air. This is a non-trivial Bayesian inference in over 4×10^6 dimensions, and so 2 hours is a very reasonable computation time — though further speedups are possible, *e.g.* we can trivially parallelize the calculations for each threshold leading to an increase in computational efficiency by a factor of the number of thresholds taken (in our case 24). Additionally, the computational bottleneck is for lower thresholds as many low-intensity peaks must be removed, and thus an adaptive scheme could be implemented as discussed previously to avoid unnecessary sampling of these thresholds. With the aforementioned speed-ups, computation of the complete peak statistic is likely to take $\mathcal{O}$(minutes) on a personal laptop.

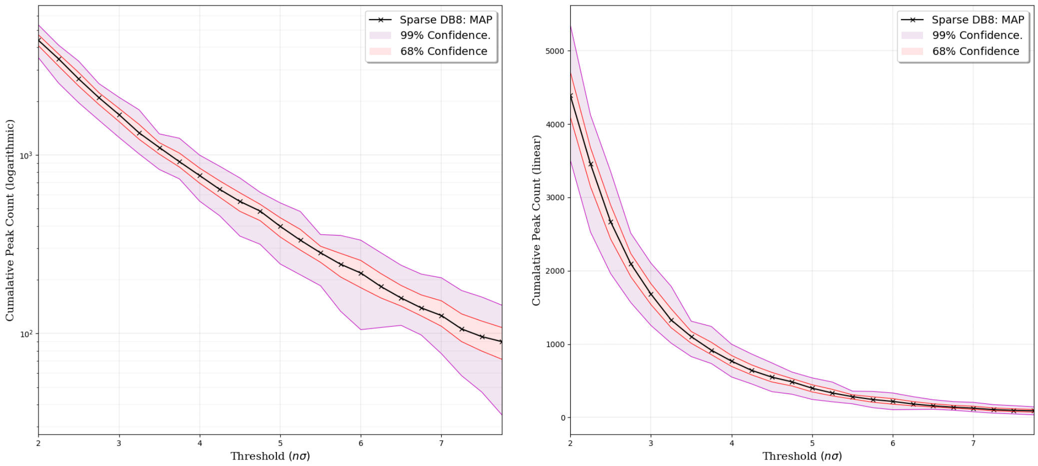
Following this initial analysis we reduce the SNR to investigate the effect of increased noise on shear measurements to the cumulative peak statistics within our Bayesian framework. We first decrease the SNR to 25 dB, seen in Figure 7.8, following which, we then reduce the input SNR further to 20 dB, the corresponding results being plotted in Figure 7.9. This higher noise level of 20 dB is still a very optimistic (somewhat unrealistic) estimate of what upcoming surveys may reach; however in this section we are primarily focused on demonstrating the methodology and leave detailed realistic simulations and forecasting for future work. A detailed description of how these noise levels in dB translate into observation constraints (*e.g.* galaxy number density *e.t.c.*) can be found in Price et al. (2021a).

7.5.3 Analysis of peak statistic

Figures 7.7, 7.8 and 7.9 clearly show that as the noise level in the discretized complex shear field increases the iso-contours of the cumulative peak statistic at 99% and 68% loosen noticeably. Therefore this, unsurprisingly, indicates that cleaner measurements are likely to give tighter constraints on cosmological parameters – though it should be noted that increasing the number of data-points (*i.e.* pixels) would have a similar effect to reducing the noise level per pixel.

For an input SNR of 20 dB (Figure 7.9) the first feature of note is the shaded blue region which indicates that for high thresholds the lower bound on the number of peaks at 99% confidence is consistent (and clipped) at 0 — this is saying that at 99%

confidence the true number of peaks at a threshold in the blue shaded region could be 0. Note that in the blue region the Bayesian upper bound is still entirely valid, it is only the Bayesian lower bound which, in our novel approach, is no longer well defined. Clearly the upper and lower bounds on the peak count statistic are dependent on the threshold one is considering and the total area over which observations are made – for wide-field surveys, more data is collected which is likely to reduce the variance of the statistic. In a general sense we summarize the mean (over all considered thresholds K) order of magnitude percentage spread on the peak statistic for the considered SNR thresholds below.



Bayesian Uncertainty in 2048×2048 Buzzard Peak statistic: SNR = 30 dB

Figure 7.7: Cumulative peak statistic for a 2048×2048 planar convergence map extracted from the Buzzard V-1.6 simulation (see section 7.5.1) contaminated with i.i.d. Gaussian noise such that the discretized simulated shearing fields are of SNR 30 dB. The *purple* outer contours are the computed upper and lower bounds at 99% confidence, with the inner *red* contours representing the 68% ($\sim 1\sigma$) bounds, included to aid comparison to similar literature which typically quote 1σ errors. Note that the information content drops for higher σ thresholds as fewer peaks are present, leading to larger relative uncertainty as fewer samples are recovered. Further note that this example is computed in a highly idealized low-noise setting.

At input SNR of 20 dB, for thresholds $\in [2\sigma, 6\sigma)$ on a single 2048×2048 planar patch the upper and lower bounds exist and are of $\mathcal{O}(48\%)$ at 99% confidence and of $\mathcal{O}(13\%)$ at 68%. At input SNR of 25 dB, for thresholds $\in [2\sigma, 8\sigma)$ on a single 2048×2048 planar patch the upper and lower bounds exist and are of $\mathcal{O}(25\%)$ at 99% confidence and of $\mathcal{O}(7\%)$ at 68%. At input SNR of 30 dB, for thresholds $\in [2\sigma, 8\sigma)$ on a single 2048×2048 planar patch the upper and lower bounds exist and are of $\mathcal{O}(15\%)$ at 99% confidence and of $\mathcal{O}(3\%)$ at 68%. These illustrative examples imply that for the Bayesian peak statistic to tightly constrain the cumulative peak statistic, comparatively larger and or cleaner data-sets may be required – or, of course, a more informative prior (though this must be well justified). However, to reduce the shot noise introduced *via* intrinsic ellipticities more galaxies must be observed within a given pixel.

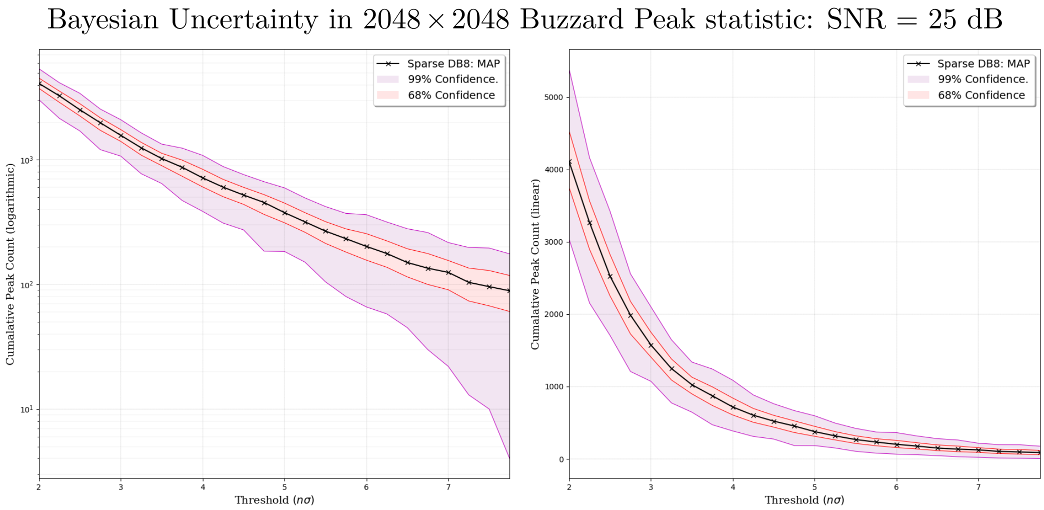


Figure 7.8: Cumulative peak statistic for a 2048 × 2048 planar convergence map extracted from the Buzzard V-1.6 simulation (see section 7.5.1) contaminated with *i.i.d.* Gaussian noise such that the discretized simulated shear (see section 9.3.2) are of SNR 25 dB. The *red* inner contours represent the upper and lower bounds at 68% ($\sim 1\sigma$) confidence, with the outer *purple* contours representing the computed bounds at 99% confidence.

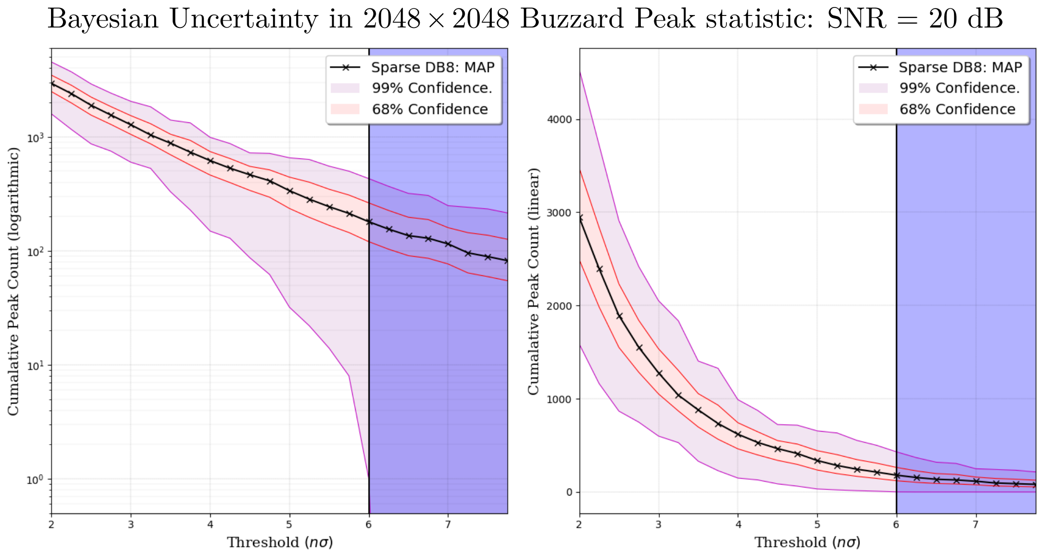


Figure 7.9: Cumulative peak statistic for a 2048 × 2048 planar convergence map extracted from the Buzzard V-1.6 simulation (see section 7.5.1) contaminated with i.i.d. Gaussian noise such that the discretized simulated shear (see section 9.3.2) are of SNR 20 dB. The *red* inner contours represent the upper and lower bounds at 68% ($\sim 1\sigma$) confidence, with the outer *purple* contours representing the computed bounds at 99% confidence. The shaded *blue* region indicates threshold values for which at 99% confidence the data cannot rule out the possibility that no peaks exist above this threshold (note that in these regions the lower bound is technically 0 and there still exists a well defined upper bound which is given). Comparing this plot to Figure 7.7 we see that as the noise level increases the 68% and 99% confidence iso-contours expand (as one would expect) and that in all cases the MAP peak statistics do not disagree at 99% confidence.

One way to increase this is to simply increase the observed number density of galaxy observations, however to do so one must observe galaxies at lower magnitude (for a fixed redshift), which inherently leads to more bright distant galaxies being detected which results in galaxy blending. Hence, increasing the number density significantly above ~ 30 gals/arcmin^2 is typically quite difficult in practice. Alternatively, the pixelisation can be adjusted to ensure that the mean galaxy count per pixel is above a given threshold — though for weak lensing the majority of non-Gaussian information is stored at fine-scales, which require small pixels, and so using larger pixels to reduce the noise level is sub-optimal for information extraction.

Within the definition of the up- and down-scaling kernels (see sections 7.4.1 and 7.4.2) we define a circular aperture around a selected peak which we define to be the extent of the peak. These regions are roughly equivalent to super-pixel regions as described in Cai et al. (2018b). In previous work it was shown (Price et al., 2019a) that for local credible intervals (*c.f.* pixel level error bars) the typical error in the approximate HPD credible region is of $\mathcal{O}(12.5\%)$, and is conservative — note that the quoted 25% mean RMSE error is split approximately equally between the upper and lower bounds, therefore this roughly corresponds to a mean error of 12.5% on both. Therefore the bounds drawn on the peak statistic here are likely to be $\sim 12.5\%$ less tight than the true Bayesian bounds — which could be formed if one were to reconstruct the 4×10^6 dimensional posterior *via* MCMC.

In this chapter (particularly the second half) we are primarily concerned with demonstrating how one may recover principled uncertainties on aggregate statistics of the convergence map – such as, but not limited to, the peak statistics. Hence we do not provide detailed analysis of how these Bayesian uncertainties may effect cosmological constraints derived from such statistics – this is saved for future work. However it is worth mentioning that one could either leverage these uncertainties to define the data covariance in a Bayesian manner (as opposed to MC which is fast but may not necessarily be fully principled, or MCMC which is $\mathcal{O}(10^6)$ times slower than our MAP approach) before then running a standard likelihood analysis, or perform a grid search in parameter space using these uncertainties again as the data covariance. Correctly accounting for the uncertainties introduced during mass-mapping has been shown to be an important consideration for the future prospects of statistics such as this (Lin, 2016).

7.6 Summary

Leveraging the sparse Bayesian mass-mapping framework previously developed (Price et al., 2019a, 2021a) we have presented two novel Bayesian uncertainty quantification techniques which can be performed directly on weak lensing convergence maps.

The first of these techniques recovers the uncertainty in the location of a feature of interest within a reconstructed convergence map — *e.g.* a large peak — at some well defined confidence. We call this locational uncertainty the *'Bayesian location'*. Additionally, for computational efficiency we develop a novel sampling scheme of the position iso-contour of a given feature which we call *'N-splitting circular bisection'*. We find that sampling the position iso-contour in this way could be many orders of magnitude faster in high dimensions than typical inverse nesting approaches. To evaluate this technique, we perform sparse Bayesian reconstructions of 1024×1024 convergence maps extracted from Bolshoi N-body simulation datasets upon which we compute the Bayesian location of the four largest sub-halos for a range of noise-levels.

The second of theses techniques quantifies the uncertainty in the cumulative peak statistic of a recovered convergence map. With this technique we can for the first time provide principled Bayesian lower and upper bounds on the number of observed peaks at a given signal to noise threshold, for a single observation, at well defined confidence. We extract 2048×2048 convergence maps from the Buzzard V-1.6 N-body simulation, upon which we calculate the cumulative peak statistic with Bayesian upper and lower bounds at 99% for a range of input noise-levels. We also provide the 68% confidence bounds which we infer from the 99% bounds to aid comparison to typical bootstrapping (MC) approaches.

For upcoming wide-field surveys convergence reconstruction will likely be done natively on the sphere (a single collective sample) to avoid projection effects, making bootstrapping approaches difficult and at worst infeasible due to the fact that they are only asymptotically exact. Bayesian approaches require only a single set of observations to make exact inferences, and so extend trivially to the more complex spherical setting. Moreover the novel uncertainty quantification techniques presented in this section and those presented previously in Cai et al. (2018b) and Price et al. (2019a, 2021a) can be rapidly computed and support algorithmic structure which can be highly parallelized, making them the ideal tools for principled analysis of convergence maps.

Chapter 8

General spherical inverse problems

This chapter is based on research presented in M. A. Price, L. Pratley, J. D. McEwen, "Sparse image reconstruction on the sphere: a general approach with uncertainty quantification", submitted to IEEE Transactions on Image Processing, 2021.

Price et al. (2021c)

Note: *This chapter assumes knowledge of the following. Convex optimisation and proximal analysis, as outlined in section 2.2. Bayesian methods, particularly those discussed in section 3.2.2 and 3.2.3. An understanding of analysis of signals over spheres is helpful, including* e.g. *the construction of spherical wavelet dictionaries. The spin operators of section 4.2.4 will be adopted without definition.*

In previous chapters we considered the mass-mapping inverse problem over tangent planes, however such planar techniques are insufficient for upcoming stage IV surveys, which are forecast to observe a large portion of the celestial sphere. Hence, to avoid projection effects, mass-mapping techniques must be extended to the sphere. More broadly, inverse problems defined naturally on the sphere are becoming increasingly of interest. As a first step towards spherical mass-mapping, in this chapter we formulate 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 (regularisation; see sections 3.1.3 and 2.1.3), the problem definition, *e.g.* the problem formulation (constrained/unconstrained) and problem setting (analysis/synthesis), and optimisation adopted to solve the problem (see section 2.2). We discuss and quantify the trade-offs between problem formulation and setting. Crucially, we consider the Bayesian interpretation of the unconstrained problem (see section 3.2.2) which, combined with recent developments in probability density theory (see section 3.2.3), 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. This chapter lays the foundations upon which the mass-mapping algorithm of chapter 5 can be extended to the celestial sphere in chapter 9.

8.1 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 (see *e.g.* Aghanim et al., 2020; Price et al., 2021b; Wallis et al., 2021), biomedical imaging (see *e.g.* Tuch, 2004; McEwen and Price, 2019), and geophysics (see *e.g.* Ritsema et al., 2011). 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 counterparts on the sphere, there are a 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.* optimisation) are the most effective.

Due to recent advances in the theory of compressed sensing (Candès et al., 2006; Candès et al., 2006; Donoho, 2006) sparsity priors (*e.g.* ℓ_1-regularisation) 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) — see sections 3.1.3 and 2.1.3 for further discussion. Recent developments in proximal convex optimisation algorithms facilitate the practical application of non-differentiable priors (see section 2.2.4), where they can be distributed and scale to high dimensional parameter spaces (Onose et al., 2016; Pratley et al., 2019). The spherical counterparts for discrete gradient spaces (McEwen et al., 2013a), wavelet families (Narcowich et al., 2006; Starck, J.-L. et al., 2006a; Baldi et al., 2009), and scale-discretized wavelet families (Wiaux et al., 2008; Leistedt et al., 2013; McEwen et al., 2013b, 2015c; McEwen et al., 2018) have been developed, and have found wide applications — see previous papers in this series (McEwen et al., 2013a; Wallis et al., 2017) for a more comprehensive overview of this topic. Somewhat restricted investigations of some aspects have already been conducted, *e.g.* considering sparsity in spherical harmonic space (Rauhut and Ward, 2012) and sparsity in various redundant dictionaries (Abrial et al., 2007; McEwen et al., 2013a; Wallis et al., 2017).

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 (Elad et al., 2006). For Euclidean settings the analysis problem typically provides greater reconstruction fidelity, a characteristic often attributed to the lower cardinality of the analysis solution space (Elad et al., 2006; Cleju et al., 2012; Nam et al., 2013). 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 (Wallis et al., 2017). There also exists a more fundamental binary classification of optimisation problems: constrained and unconstrained, corresponding to regularisation *via* hard and soft constraints respectively (Boyd, 2004), as discussed in section 2.2. 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 (see *e.g.* Robert, 2001) and thus support a principled statistical interpretation (Feeney et al., 2014; Pereyra, 2017; Cai et al., 2018b).

Traditionally, although variational approaches may support a probabilistic interpretation they typically recover point estimates and do not quantify uncertainties (see section 3.2.2). Fully probabilistic approaches (*e.g.* Markov chain Monte Carlo sampling methods, discussed in section 3.2.1) exist but are computationally expensive in the high dimensional setting of the sphere, motivating the development of hybrid techniques, *e.g.* those presented in chapter 5. Recent developments in the field of probability density theory (Pereyra, 2017) 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 (Cai et al., 2018b; Repetti et al., 2019; Price et al., 2019a,b, 2021a), with applications in a variety of domains. In this chapter, we leverage these UQ techniques (specifically those discussed in chapters 5-6) 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), such that a great many combinations of likelihood (data-fidelity) and prior (regularisation 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 (Carrillo et al., 2012, 2013; Onose

---

[1] `https://github.com/astro-informatics/s2inv`

et al., 2016) software package for flexible, efficient sparse optimisation. We use fast exact spherical harmonic (McEwen and Wiaux, 2011) and spherical wavelet transforms (Leistedt et al., 2013) to rapidly solve linear and ill-conditioned spherical inverse problems.

The remainder of this chapter is structured as follows. In Section 8.2 we provide the mathematical context which underpins analysis of spin signals on the sphere. In Section 8.3 we present variational regularisation 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 regularisation functionals to their spherical counterparts, and briefly highlight highly optimized, scalable spherical reconstruction open-source software available as a by-product of this work. In Section 8.4 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 8.5 before providing concluding remarks in Section 8.6.

8.2 Spin-signals on the sphere and rotation group

One often wishes to consider the frequency space representation of signals; whether this be embedded within regularisation 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 though projection onto the space of *spin spherical harmonics*. In this section we review the mathematical background fundamental to the analysis of signals defined on the sphere.

8.2.1 Spin spherical harmonic transforms

The space of square integrable spin-s functions ${}_sf \in \mathsf{L}^2[\mathbb{S}^2]$, 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 ${}_sf'(\omega) = \mathrm{e}^{-is\chi}{}_sf(\omega)$ where ${}_sf'$ is the rotated function (Newman and Penrose, 1966; Goldberg et al., 1967). Such functions are most naturally represented by the spin-weighted spherical harmonics ${}_sY_{\ell m} \in \mathsf{L}^2[\mathbb{S}^2]$ which are a set of complete and orthogonal basis functions for degree $\ell \in \mathbb{Z}_{>0}$ and integer $m \in \mathbb{Z}, |m| \leq \ell, |s| \leq \ell$. We adopt the Condon-Shortley phase convention (Condon et al., 1951), which results in conjugate symmetry ${}_sY^{\star}_{\ell m}(\omega) = (-1)^{s+m}{}_{-s}Y_{\ell -m}(\omega)$, where $(\cdot)^{\star}$ denotes complex conjugation.

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

$$
{}_sf_{\ell m} \equiv \langle {}_sf_{\ell m} | {}_sY_{\ell m} \rangle_{\mathbb{S}^2} = \int_{\mathbb{S}^2} d\Omega(\omega)\, {}_sf(\omega)\, {}_sY^{*}_{\ell m}(\omega), \tag{8.1}
$$

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 ${}_sY_{\ell m}(\omega)$, one can exactly synthesize the signal space representation by

$$ {}_sf(\omega) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} {}_sf_{\ell m} \, {}_sY_{\ell m}(\omega), \tag{8.2} $$

where the sum over ℓ is often truncated at L, and where it is assumed that ${}_sf_{\ell m} = 0, \forall \ell \geq L$. In this sense, signals are considered to be bandlimited at L. For notational brevity we adopt the shorthand operator notation Y and 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 that 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.

8.2.2 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 $\mathcal{R}_\rho$, for Euler angles $\rho = (\alpha, \beta, \gamma) \in \mathrm{SO}(3)$ with $\alpha \in [0, 2\pi), \beta \in [0, \pi)$, and $\gamma \in [0, 2\pi)$, with action $(\mathcal{R}_\rho {}_sf)(\omega) \equiv e^{-is\theta} {}_sf(\mathcal{R}_\rho^{-1} \boldsymbol{\omega})$. The directional scale-discretized wavelet coefficients of any square integrable spin-s function ${}_sf \in \mathsf{L}^2[\mathbb{S}^2]$ are given for scale j by the directional convolution

$$ W^{{}_s\Psi^{(j)}}(\rho) \equiv ({}_sf \circledast {}_s\Psi^{(j)})(\rho) \equiv \langle {}_sf, \mathcal{R}_\rho {}_s\Psi^{(j)} \rangle_{\mathbb{S}^2} = \int_{\mathbb{S}^2} d\Omega(\omega) {}_sf(\omega) (\mathcal{R}_\rho {}_s\Psi^{(j)})^\star(\omega), \tag{8.3} $$

where $\circledast$ represents the directional spherical convolution and ${}_s\Psi^{(j)} \in \mathsf{L}^2[\mathbb{S}^2]$ is the wavelet kernel at scale $j \in \mathbb{Z}_{>0}$, which determines the compact support of a given wavelet scale (Leistedt et al., 2013).

Typically wavelet coefficients have negligible energy concentration over the low-frequency domain in harmonic space, hence a scaling function ${}_s\Upsilon \in \mathsf{L}^2[\mathbb{S}^2]$ is introduced (McEwen et al., 2015c; McEwen and Price, 2019) with coefficients $W^{{}_s\Upsilon} \in \mathsf{L}^2[\mathbb{S}^2]$ defined by the axisymmetric convolution $\odot$ with a signal ${}_sf \in \mathsf{L}^2[\mathbb{S}^2]$ such that

$$ W^{{}_s\Upsilon}(\omega) \equiv ({}_sf \odot {}_s\Upsilon)(\omega) \equiv \langle {}_sf, \mathcal{R}_\omega {}_s\Upsilon \rangle_{\mathbb{S}^2} = \int_{\mathbb{S}^2} d\Omega(\omega') {}_sf(\omega') (\mathcal{R}_\omega {}_s\Upsilon)^\star(\omega'), \tag{8.4} $$

where $\mathcal{R}_\omega = \mathcal{R}_{(\psi,\theta,0)}$ is a simplification of $\mathcal{R}_\rho$. One can straightforwardly show that the pixel-space representation of signals may be exactly synthesized

$$
{}_sf(\omega) = \int_{\mathbb{S}^2} \mathrm{d}\Omega(\omega') W^{{}_s\Upsilon}(\omega')(\mathcal{R}_{\omega'}\, {}_s\Upsilon)(\omega) + \sum_j \int_{\mathrm{SO}(3)} \mathrm{d}\rho(\rho) W^{{}_s\Psi^{(j)}}(\rho)(\mathcal{R}_{\rho}\, {}_s\Psi^{(j)})(\omega), \tag{8.5}
$$

if, and only if, the wavelet admissibility condition holds (see *e.g.* McEwen et al., 2015c). There exist many functions which are admissible, *e.g.* spherical needlets (Marinucci et al., 2007), ridgelets (McEwen and Price, 2019) and curvelets (Chan et al., 2017), however in this work we choose to adopt the directional scale-discretized wavelet harmonic space kernel (Wiaux et al., 2008; McEwen et al., 2018). For notational brevity we define operators Ψ and Ψ^{-1} for the synthesis and analysis wavelet transforms respectively, with corresponding adjoint operators $\Psi^{\dagger}$ and $(\Psi^{-1})^{\dagger}$ (for further details see *e.g.* Wallis et al., 2017).

8.3 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 (Wallis et al., 2021) 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 (McEwen et al., 2013a), and the analysis and synthesis settings were compared in a spherical setting (Wallis et al., 2017). In this section we extend the discussion to include the constrained and unconstrained formulations, supported by a variety of proximal optimisation algorithms, and a variety of regularisation functionals.

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

$$
\Phi \in \mathbb{C}^{M \times N_{\mathbb{S}^2}} : x \in \mathbb{C}^{N_{\mathbb{S}^2}} \mapsto y \in \mathbb{C}^M, \tag{8.6}
$$

commonly referred to as the measurement operator, which simulates measurement acquisition. Suppose observations are contaminated with stochastic noise $n \in \mathbb{C}^M$ such that $y = \Phi x + n$ which is classically ill-posed (see *e.g.* Hadamard, 1902). 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 (see section 2.1.2). A diverse set of techniques exists to solve such inverse problems. This thesis is primarily concerned with variational approaches, for which we develop uncertainty quantification techniques in Section 8.4. Variational approaches consider the inverse problem as a minimisation problem over a chosen objective function, which is typically the combination of a data fidelity term and a regularisation term — selected to stabilize reconstruction with *a priori* assumptions as to the nature of the problem instance. Such optimisation problems are comprehensively discussed in section 2.2. Given an objective function over which to minimise, one must make a variety of decisions regarding optimisation formulation.

8.3.1 Constrained and unconstrained optimisation

Suppose one selects data-fidelity term $f(x)$ and regularisation functional $g(x)$, then the unconstrained optimisation problem has the Lagrangian formulation (see *e.g.* Boyd, 2004, and section 2.2.3)

$$x^\star = \underset{x \in \Omega}{\operatorname{argmin}} \Big[f(x) + \lambda g(x) \Big], \tag{8.7}$$

where $\Omega = \{\mathbb{C}^{N_{\mathbb{S}^2}},\ \mathbb{R}^{N_{\mathbb{S}^2}},\ \mathbb{R}_{>0}^{N_{\mathbb{S}^2}}\}$, and the regularisation parameter $\lambda \in \mathbb{R}_{>0}$ 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 optimisations in the unconstrained setting, the solution which minimizes the objective is in fact the *maximum a posteriori* (MAP) estimator $x^\star = x^{\text{MAP}}$.

Interestingly, it is well known that the unconstrained problem has direct links to Bayesian inference and supports a principled statistical interpretation, as discussed in section 3.2.2. 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 (Pereyra, 2017) to develop unconstrained optimisation techniques which support principled uncertainty quantification, as previously discussed in section 3.2.3. Therefore, when considering spherical imaging problems, where Bayesian sampling methods are impractical, in scientific domains, where uncertainty quantification is a desirable feature, unconstrained optimisation exhibits significant advantages. However, this advantage comes at the additional complexity of optimal selection of the regularisation parameter λ. Popular methods for selection of λ have adopted criteria such as: the Akaike information criterion (AIC; Akaike, 1998), Bayesian information criterion (BIC; Schwarz et al., 1978), or Stein's unbiased risk estimator (SURE; Stein, 1981; Pesquet et al., 2009) and others (Vidal and Pereyra, 2018). Optimal regularisation 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 (Pereyra et al., 2015) which treats the regularisation parameter as a nuisance variable (Robert, 2001) over which a majorisation-minimisation algorithm marginalizes. Effectively this method produces automatic, somewhat robust λ selection with a straightforward, natural Bayesian interpretation, facilitating principled uncertainty quantification (see section 5.2.4 for an overview of this method).

Suppose instead that one is unwilling to accept a trade-off in either the data-fidelity or regularisation 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 regularisation support or measurement operator. For such inverse problems, the problem instance is formulated as a constrained optimisation problem, in which one function is minimized subject to the constraint that the other function belongs to some constrained set (Boyd, 2004). Here we consider the common form in which the regularisation functional is minimized subject to the constraint that the solution belongs within a level-set of the data-fidelity term, *i.e.*

$$x^\star = \underset{x \in \Omega}{\operatorname{argmin}} \Big[g(x) \Big] \quad s.t. \quad f(x) \leq \delta, \tag{8.8}$$

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 optimisation restricts solutions to the sub-space $x \in \mathcal{B}_f^\delta$ where $\mathcal{B}_f^\delta$ is the f-ball centred at $z \in \Omega$ with radius δ, *i.e.* $\mathcal{B}_f^\delta(z) := \{x : f(x) \leq \delta\}$.

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 (regularisation 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.

8.3.2 Analysis and synthesis settings

Often one adopts regularisation functions which are computed on projections of the image space, *e.g.* wavelet space, harmonic space, gradient space *etc.* (see sections 2.1.3 and 3.1.3). In such settings one can formulate optimisation 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 $x^\star$ (analysis) or projected space $\alpha^\star$, which are then inverted to form pixel-space estimates $x^\star = \Psi\alpha^\star$ (synthesis) (Elad et al., 2006; Cleju et al., 2012; Nam et al., 2013).

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 regularisation functional $g_{\text{A}}(x) = \|\Psi^{-1}x\|_1$ and an ℓ_2 data-fidelity term $f_{\text{A}}(x) = \|\Phi x - y\|_2^2$. Clearly in the analysis formulation the optimisation problems are precisely those given in Section 8.3.1. However, in the synthesis settings the regularisation functional takes the form $g_{\text{S}}(\alpha) = \|\alpha\|_1$, while the data-fidelity term is given by $f_{\text{S}}(\alpha) = \|\Phi\Psi\alpha - y\|_2^2$. With these definitions the synthesis optimisation problem reads in much the same way as those presented in Section 8.3.1 and, in fact, for situations in which the measurement operator is orthogonal, *i.e.* $\Psi^{-1} = \Psi^\dagger$, these formulations are equivalent. However, they have very different geometric properties when this is not the case (Elad et al., 2006; Cleju et al., 2012; Nam et al., 2013). Notice that we adopt over-complete spherical wavelet transforms where $\Psi^{-1} \neq \Psi^\dagger$, and sampled spherical harmonic transforms which are not orthogonal, *i.e.* $\mathsf{Y}^{-1} \neq \mathsf{Y}^\dagger$ (McEwen et al., 2013a) — 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 (Elad et al., 2006; Cleju et al., 2012; Nam et al., 2013). However, in previous work it was concluded that this characteristic may not generalize to the spherical setting (Wallis et al., 2017). In Section 8.5 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 some settings can be computationally inefficient on the sphere. An example of such computational savings in the synthesis setting is demonstrated in Section 8.5.3, where an iterative Wiener filtering approach is adopted. In this case the analysis/synthesis formulations require $6, 2$ spherical harmonic transforms respectively.

8.3.3 Regularisation functionals on the sphere

Having discussed the variety of ways one may formulate and construct an optimisation on the sphere, we should now consider spherical counterparts to common regularisation functionals and how one can develop these for the spherical setting. Such regularisation functionals include *e.g.* sparsity promoting $\|\cdot\|_1$ regularisers, typically in a sparsifying dictionary Ψ, which are often motivated by the theory of compressed sensing (Candès et al., 2006; Candès et al., 2006; Donoho, 2006); Gaussian $\|\cdot\|_2^2$ regularisers, which are often iterative implementations of harmonic Wiener filters (Hiller and Chin, 1990; Kodi Ramanah et al., 2019); and spherical total-variation (TV) priors (McEwen et al., 2013a), 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 (McEwen and Wiaux, 2011), which does not uniformly sample the continuous norms. Typically this results in disproportionate weight being applied to pixels located at the poles, due to progressively increased sampling density away from the equator. To account for this spherical (directional wavelet) counterparts to the traditional norms are defined by

$$ {}_{\mathbb{S}^2}\|x\|_p = \|w \circ x\|_p \Rightarrow {}_{\Psi}\|\alpha\|_p = \Big(\sum_j \sum_n {}_{\mathbb{S}^2}\|\alpha_{j,n}\|_p^p\Big)^{\frac{1}{p}}, \tag{8.9} $$

respectively, where $w \in \mathbb{S}^2$ is the corresponding map of reciprocal pixel areas on the sphere, $\circ$ is the Hadamard product, and $j, n \in \mathbb{Z}_{>0}$ 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 ${}_{\Psi}\|\Psi^{\dagger}x\|_1$. Such a generalization permits multi-resolution algorithms (Leistedt et al., 2013) 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 (McEwen and Wiaux, 2011), however in this work we find simple weights $w_{j,n}$ which capture the pixel area to be sufficient.

8.3.4 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 regularisation functionals (*e.g.* ℓ_1-norms) for which proximal operators may be used to navigate the objective function, thus motivating proximal convex optimisation algorithms.

Convex optimisation 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 optimisation 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 optimisation 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 8.3.3, which can be incorporated into proximal optimisation 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 (McEwen and Wiaux, 2011) is adopted through the SSHT[2] 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 (Wiaux et al., 2008; McEwen et al., 2013b, 2015c) through the S2LET[3] package (Leistedt et al., 2013; Wallis et al., 2017), which are optimally sampled and support machine precision synthesis. We leverage a recently developed, highly optimized C++ OOP sparse optimisation framework SOPT[4] (Carrillo et al., 2012, 2013; Onose et al., 2016), which facilitates a variety of proximal convex optimisation algorithms, *e.g.* forward-backward (Beck and Teboulle, 2009; Combettes and Pesquet, 2011), primal dual (Boyd, 2004; Combettes et al., 2014; Komodakis and Pesquet, 2015), and the alternating direction method of multipliers (Boyd et al., 2011), 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 8.4.

---

[2] `https://astro-informatics.github.io/ssht/`

[3] `https://astro-informatics.github.io/s2let/`

[4] `http://astro-informatics.github.io/sopt/`

8.4 Spherical Bayesian uncertainty quantification

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

$$P(x|y;\mathcal{M}) \equiv \frac{P(y|x;\mathcal{M})P(x;\mathcal{M})}{\int_{\mathbb{C}^{N_{\mathbb{S}^2}}} P(y|x;\mathcal{M})P(x;\mathcal{M})dx}, \tag{8.10}$$

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 $\Phi \in \mathbb{C}^{M \times N_{\mathbb{S}^2}} : x \mapsto y$, and some understanding of the noise inherent to y (see *e.g.* Robert, 2001). 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 thesis.

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

$$x^{\text{MAP}} \equiv \underset{x}{\text{argmax}} \Big[P(x|y;\mathcal{M}) \Big] \propto \underset{x}{\text{argmin}} \Big[\underbrace{-\log(\ P(y|x;\mathcal{M})P(x;\mathcal{M})\ )}_{h(x)=f(x)+g(x)} \Big], \tag{8.11}$$

where the second line follows by the monotonicity of the logarithm function. For convex objective functions $h(x)$ this takes the form of a convex optimisation problem (Boyd, 2004), and therefore Equation 8.7 explicitly returns the MAP solution, as asserted in Section 8.3.1. Hence, leveraging state-of-the-art convex optimisation techniques one can efficiently locate the solution which maximizes the posterior odds. However this is still a point estimate which, though useful, does not obviously support uncertainty quantification. Recently, approximate contours of the latent space have been derived facilitating variational regularisation 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.

8.4.1 Highest posterior density credible regions

A credible region $C_\alpha \subset \mathbb{C}^{N_{\mathbb{S}^2}}$ of the posterior latent space at credible confidence $100(1-\alpha)\%$, for $\alpha \in [0,1]$, satisfies the integral equation (Robert, 2001)

$$P(x \in C_\alpha | y; \mathcal{M}) = \int_{x \in \mathbb{C}^{N_{\mathbb{S}^2}}} P(x|y;\mathcal{M}) \mathbb{I}_{C_\alpha} dx = 1 - \alpha, \tag{8.12}$$

where $\mathbb{I}_{C_\alpha}$ is the standard set indicator function. The optimal credible region in the sense of minimal volume (Robert, 2001) is the highest posterior density (HPD) credible region defined by $C_\alpha := \{x : h(x) \leq \epsilon_\alpha\}$, where $\epsilon_\alpha \in \mathbb{R}_{>0}$ is an iso-contour of the log-posterior. Determination of the HPD region requires computation of the integral in Equation 8.12, which is computationally unfeasible in even moderate dimensional spherical settings, due to dimensionality and functional complexity considerations. Convex objectives $h(x)$ support the conservative approximate HPD credible region C'_α defined by (Pereyra, 2017)

$$C_\alpha \subseteq C'_\alpha \subset \mathbb{C}^{N_{\mathbb{S}^2}} := \Big\{x : h(x) \leq \epsilon'_\alpha\Big\} \quad \text{for} \quad \epsilon'_\alpha = h(x^{\text{MAP}}) + \sqrt{16N\log(3/\alpha)} + N, \tag{8.13}$$

which allows one to approximate C_α with only knowledge of the MAP solution x^{MAP} and the dimension $N = \mathbb{C}^{N_{\mathbb{S}^2}}$. An upper bound on the approximation error exists (Pereyra, 2017). Therefore for convex objectives, given 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 (Pereyra, 2017; Cai et al., 2018b; Price et al., 2019a,b; Repetti et al., 2019; Price et al., 2021a).

8.4.2 Bayesian hypothesis testing on the sphere

The most straightforward uncertainty quantification technique one may generate by leveraging the approximation of Equation 8.13 is that of hypothesis testing (Cai et al., 2018b; Price et al., 2021a,b). The concept of hypothesis testing is to adjust a feature of the recovered estimator x^{MAP} generating a surrogate solution x^{sur}, of which we ask is $x^{\text{sur}} \in C'_\alpha$? If $x^{\text{sur}} \notin C'_\alpha \Rightarrow x^{\text{sur}} \notin C'$, which follows from the conservative nature of the approximation in Equation 8.13, the feature of interest is considered to be statistically significant (necessary to the reconstruction) at $100(1-\alpha)\%$ confidence. Conversely $x^{\text{sur}} \in C'_\alpha$ indicates that the surrogate solution remains within the approximate credible set and we conclude that the feature is indeterminate.

8.4.3 Local credible intervals on the sphere

Suppose one recovers an optimal solution x^{MAP} through unconstrained convex optimisation (see Section 8.3.1) 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 $100(1-\alpha)\%$, 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.* $x^{\text{sur}} \notin C'_\alpha$. 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 $x_i \in \Omega_i \Rightarrow \zeta_{\Omega_i} = 1$ and $x_i \notin \Omega_i \Rightarrow \zeta_{\Omega_i} = 0$. For a given Ω_i locate upper (lower) bounds $\xi^{+}_{\Omega_i}$, $\xi^{-}_{\Omega_i}$ respectively, which saturate the HPD credible region C'_α (Cai et al., 2018b; Price et al., 2019a). This is achieved by the following optimisations,

$$\xi^{\pm}_{\Omega_i} = \max_{\xi \in \mathbb{R}_{>0}} \big\{ \pm\xi \,|\, f(x_{i,\xi}) + g(x_{i,\xi}) \leq \epsilon'_\alpha \big\}, \tag{8.14}$$

where $x_{i,\xi} = x^{\text{MAP}} \zeta_{\Omega/\Omega_i} + \xi \zeta_{\Omega_i}$ 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 (Cai et al., 2018b), which can simply be recovered *via* bisection. Though conditional local uncertainty quantification techniques such as this have demonstrated utility in certain circumstances (Cai et al., 2018b; Price et al., 2019a,b, 2021a), in the high dimensional spherical settings they can quickly become dilute (Price et al., 2021b). 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.* Price et al., 2021b).

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.

8.4.4 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 which, without loss of generality, can be written as

$$h(x) = f(x) + g(x) = \|\Phi x - y\|^{p_2}_{p_1} + \|\Psi^\dagger x\|^{q_2}_{q_1}. \tag{8.15}$$

Consider again the partition $x^{\text{sur}} = x\zeta_{\Omega/\Omega_i} + \xi\zeta_{\Omega_i}$ upon which the applications of any linear operator $\mathcal{L}$ is given trivially by linearity to be $\mathcal{L}x^{\text{sur}} = \mathcal{L}x\mathbb{I}_{\zeta_{\Omega/\Omega_i}} + \xi\mathcal{L}\mathbb{I}_{\zeta_{\Omega_i}}$. Explicitly expanding Equation 8.15 with linearity one finds

$$\|\Phi x\mathbb{I}_{\zeta_{\Omega/\Omega_i}} + \xi\Phi\mathbb{I}_{\zeta_{\Omega_i}} - y\|_{p_1}^{p_2} + \|\Psi^\dagger x\mathbb{I}_{\zeta_{\Omega/\Omega_i}} + \xi\Psi^\dagger\mathbb{I}_{\zeta_{\Omega_i}}\|_{q_1}^{q_2} \Rightarrow \|a+\xi b\|_{p_1}^{p_2} + \|c+\xi d\|_{q_1}^{q_2}, \tag{8.16}$$

for constant (per credible interval) vectors defined to be

$$a = \Phi x\mathbb{I}_{\zeta_{\Omega/\Omega_i}} - y \qquad b = \Phi\mathbb{I}_{\zeta_{\Omega_i}} \qquad c = \Psi^\dagger x\mathbb{I}_{\zeta_{\Omega/\Omega_i}} \qquad d = \Psi^\dagger\mathbb{I}_{\zeta_{\Omega_i}} \tag{8.17}$$

In this way the local credible optimisation problems in Equation 8.14 can be re-written instead as

$$\|a+\xi b\|_{p_1}^{p_2} + \|c+\xi d\|_{q_1}^{q_2} \leq \epsilon'_\alpha, \tag{8.18}$$

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 8.18 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.

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 8.18 to the binomial inequality $\|a+\xi b\|_2^2 + \|c+\xi d\|_2^2 \leq \epsilon'_\alpha$, which expands to give

$$[\|b\|_2^2 + \|d\|_2^2]\xi^2 + 2[a\cdot b + c\cdot d]\xi + [\|a\|_2^2 + \|c\|_2^2] \leq \epsilon'_\alpha. \tag{8.19}$$

One could gain some geometric insight by considering the case in which $a\cdot b + c\cdot d = 0$, 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.

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 8.18 reduces to a 2^{nd}-order polynomial $\|a+\xi b\|_2^2 + \|c+\xi d\|_1 \leq \epsilon'_\alpha$, 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

$$\|b\|_2^2\xi^2 + 2[a\cdot b + \|d\|_1]\xi + [\|a\|_2^2 + \|c\|_1] \lessapprox \epsilon'_\alpha, \tag{8.20}$$

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.

8.5 Numerical experiments

In this section we showcase the variational regularisation and uncertainty quantification techniques presented in section 8.3 and section 8.4 on a diverse set of numerical experiments. For each scenario we create mock observations $y = \Phi x$ of a ground truth signal x which are related through the forward model $\Phi : x \mapsto y$ from which we formulate an ill-posed inverse problem (*e.g.* add noise, mask, blur, *etc.*). We solve this inverse problem for an estimator $x^\star$ (x^{MAP}) of x the success of which we quantify by the recovered *signal to noise ratio*, defined by $\text{SNR} = 20 \log_{10}(\|x\|_2 / \|x - x^\star\|_2)$.

8.5.1 Earth satellite topography

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

$$\Phi_{\text{ET}} = \mathsf{D}\mathsf{Y}^{-1}\Theta\mathsf{Y} \quad \text{and} \quad \Phi_{\text{ET}}^\dagger = \mathsf{Y}^\dagger\Theta\ (\mathsf{Y}^{-1})^\dagger\mathsf{D}^\dagger, \tag{8.21}$$

where $\mathsf{Y}, \mathsf{Y}^{-1}$ are forward and inverse spherical harmonic transforms correspondingly, $\mathsf{D}, \mathsf{D}^\dagger$ 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 n is a univariate Gaussian the data-fidelity (log-likelihood) term is simply given by $\frac{1}{2\sigma^2}{}_{\mathbb{S}^2}\|\Phi_{\text{ET}} x - y\|_2^2$. Here we adopt a sparsity promoting ℓ_1-norm wavelet regularisation ${}_{\Psi}\|\Psi^\dagger(\cdot)\|_1$ (Laplace distribution log-prior), and solve both the constrained formulation through the proximal ADMM algorithm (Boyd et al., 2011) and the unconstrained formulation through the proximal forward backward algorithm (Boyd, 2004; Beck and Teboulle, 2009; Combettes and Pesquet, 2011), in both the analysis and synthesis settings. Notice the use of spherical (wavelet) space norms, outlined in Section 8.3.3, 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 8.1). 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 bottom plot of Figure 8.1). This was observed in prior analysis (Wallis et al., 2017) and reported as evidence that the synthesis setting may produce more optimal results.

Note that the synthesis setting implicitly bandlimits the observations x, therefore restricting the solution space cardinality — a factor known to impact reconstruction fidelity (Elad et al., 2006). To account for this bias we reran the analysis optimisation with an explicitly bandlimited measurement operator, results of which can be seen in Figure 8.2 (which also demonstrates the uncertainty quantification of Section 8.4.3). It was found that the analysis setting performs similarly to the synthesis setting for this problem, leading us to conclude that the optimality of optimisation formulation is at best ambiguous.

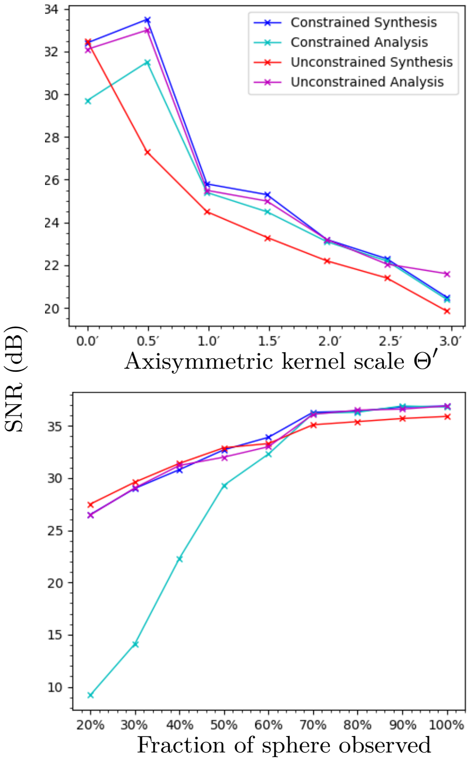


Figure 8.1: Top: Recovered SNR for a variety of problem setups versus variety of axisymmetric blurring kernel scales, for a fixed inpainting of 50% masked pixels. **Bottom:** Recovered SNR for a variety of problem setups over a variety of inpainting scenarios versus % of pixels observed. **Discussion:** Generally each setup performs similarly and it would appear that no single setting is optimal in all cases. Notice the 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 of the synthesis problem.

8.5.2 360° camera blur deconvolution

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

$$\Phi_{360^\circ} = \mathsf{Y}^{-1}\Theta\mathsf{Y} \quad \text{and} \quad \Phi^{\dagger}_{360^\circ} = \mathsf{Y}^{\dagger}\Theta\,(\mathsf{Y}^{-1})^{\dagger}, \tag{8.22}$$

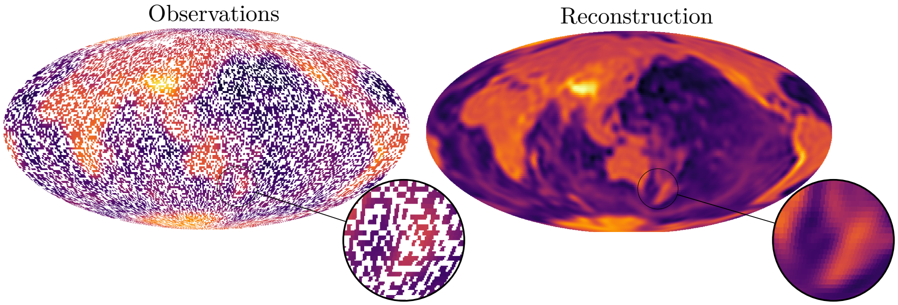


Figure 8.2: Left: Simulated observations contaminated with 30dB Gaussian *i.i.d.* noise, convolved with a ~ 268 arc-minute Gaussian blurring kernel, and with 60% of pixels masked. Original dataset was extracted from the Earth Gravitational Model (EGM2008) publicly released by the U.S. National Geospatial-Intelligence Agency (NGA) EGM development team. **Right:** Unconstrained reconstruction using ℓ_1-norm wavelet sparsity regularisation (log-prior) in the analysis setting, adopting the proximal forward-backward algorithm (Boyd, 2004; Beck and Teboulle, 2009; Combettes and Pesquet, 2011).

where $\mathsf{Y}, \mathsf{Y}^{-1}$ are forward and inverse spherical harmonic transforms correspondingly (see section 8.2), $\mathsf{\Theta}$ is the axisymmetric convolution with the harmonic representation of Θ.

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

$$x^\star = \underset{x \in \mathbb{C}^{N_{\mathbb{S}^2}}}{\operatorname{argmin}} \Big[g(x) \Big] \text{ s.t. } \frac{1}{2\sigma^2} {}_{\mathbb{S}^2}\|\mathsf{\Phi}_{360^\circ}x - y\|_2^2 \leq \delta, \tag{8.23}$$

where $\delta \in \mathbb{R}_{>0}$ is the radius of the ℓ_2-ball $\mathcal{B}^\delta_{\ell_2}$ 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 (Boyd, 2004; Combettes et al., 2014; Komodakis and Pesquet, 2015). Both priors produce similar results, with wavelet sparsity regularisation recovering SNR=17.60 dB and TV-norm marginally superior at SNR = 17.65 dB — seen in Figure 8.3.

8.5.3 CMB temperature anisotropies

Suppose one captures masked (and therefore incomplete) measurements of the cosmic microwave background (CMB; Aghanim et al., 2020) $y \in \mathbb{C}^{M_{\mathbb{S}^2}}$ that can be related through a mapping operator $\mathsf{\Phi}_{\text{CMB}} \in \mathbb{C}^{M_{\mathbb{S}^2} \times N_{\mathbb{S}^2}}$ to the full-sky CMB signal $x \in \mathbb{C}^{N_{\mathbb{S}^2}}$ which can be decomposed into harmonic coefficients $\hat{x}_{\ell m} = \langle x, Y_{\ell m} \rangle$ which (for Gaussian fields such as CMB; Aghanim et al., 2020) are uncorrelated and isotropic $\mathbb{E}[\hat{x}^*_{\ell m}\hat{x}_{\ell' m'}] = \delta_{\ell\ell'}\delta_{mm'}C_\ell$, where C_ℓ is the angular power spectrum.

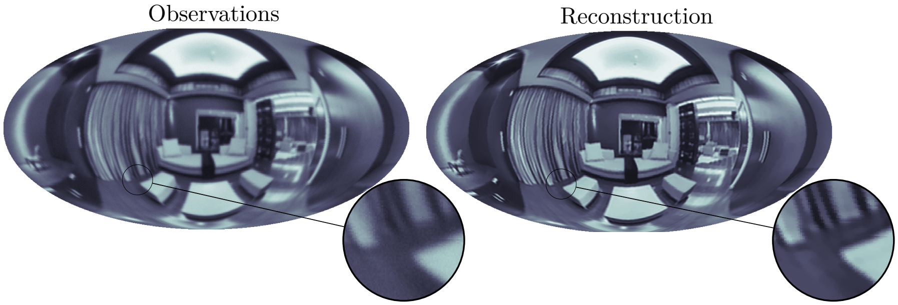


Figure 8.3: Left: Simulated observations contaminated with 30dB Gaussian noise and convolved with a ~ 78 arc-minute Gaussian blurring kernel (*e.g.* motion blur). **Right:** Spherical TV-norm $\|x\|_{\text{TV}}$ regularized reconstruction through proximal primal dual algorithm in the analysis setting. Image sourced from the SUN 360 dataset (Xiao et al., 2012).

These considerations motivate the choice of a multivariate Gaussian prior $P(\hat{x}|\mathcal{M}) = \exp\big(-\hat{x}^\dagger \mathsf{C}^{-1}\hat{x}/2\big)$ for vectorized harmonic coefficients $\hat{x}$ and covariance 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 $n \sim \mathcal{N}(0,\sigma^2) \in \mathbb{C}^{M_{\mathbb{S}^2}}$ then the whitened harmonic coefficients $\hat{x}' = \mathsf{C}^{-\frac{1}{2}}\hat{x}$ are modelled as $y = \Phi_{\text{CMB}}\hat{x}' + n$ for measurement operator

$$\Phi_{\text{CMB}} = \mathsf{D}\mathsf{Y}\mathsf{C}^{\frac{1}{2}} \quad \text{and} \quad \Phi^\dagger_{\text{CMB}} = \mathsf{C}^{\frac{1}{2}}\mathsf{Y}^\dagger\mathsf{D}^\dagger, \tag{8.24}$$

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

$$\hat{x}^{\text{MAP}\prime} = \underset{\hat{x}'}{\operatorname{argmin}}\Big[\|\hat{x}'\|_2^2 + \frac{1}{2\sigma^2}{}_{\mathbb{S}^2}\|\Phi_{\text{CMB}}\hat{x}' - y\|_2^2\Big], \tag{8.25}$$

where the pixel space signal is recovered by $x^{\text{MAP}} = \mathsf{Y}^{-1}\mathsf{C}^{\frac{1}{2}}\hat{x}^{\text{MAP}\prime}$. This is the convex optimisation 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 (see *e.g.* Figure 6 in Kodi Ramanah et al., 2019), 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 8.4.

8.5.4 Weak gravitational lensing

The following example considers spherical imaging of dark matter (Hu, 2000; Wallis et al., 2021). A more extensive analysis that applies the method presented to observational data, and not just simulations, is performed in (Price et al., 2021b) which leverages many of the methods developed in this work. At first order, gravitational lensing manifests itself into the spin-0 convergence ${}_0x(r,\omega) \in \mathbb{C}^{N_{\mathbb{S}^2}}$ (the integrated matter field along the line of sight) and the spin-2 shear ${}_2y(r,\omega) \in \mathbb{C}^{N_{\mathbb{S}^2}}$ (the ellipticity of observed images) which can be related to the lensing potential ${}_0\phi(r,\omega) \in \mathbb{C}^{N_{\mathbb{S}^2}}$

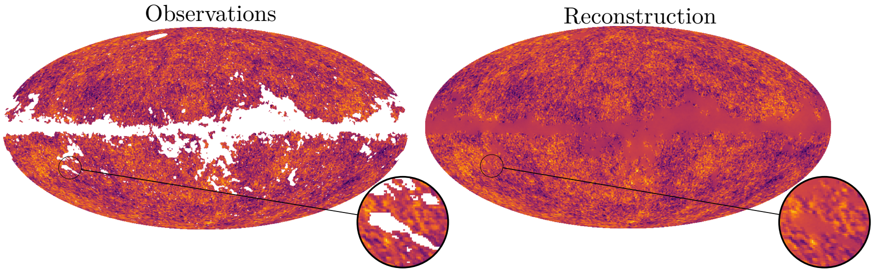


Figure 8.4: Left: Simulated Gaussian random field generated from lambda cold dark matter (Λ-CDM) best fit power spectrum, masked by a Planck survey mask (Aghanim et al., 2020) and polluted by 30dB *i.i.d.* Gaussian noise. **Right:** Unconstrained reconstruction using an ℓ_2-norm Wiener prior solved by 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.* Kodi Ramanah et al., 2019).

by ${}_0x(r,\omega) = \frac{1}{4}(\eth_+\eth_- + \eth_-\eth_+)\,{}_0\phi$ and ${}_2y(r,\omega) = \frac{1}{2}\eth_+\eth_+\,{}_0\phi$, where $\eth_\pm$ are spin-raising/lowering operators (Newman and Penrose, 1966; Goldberg et al., 1967). One can then relate ${}_0x$ and ${}_2y$ to one another in harmonic space by ${}_2\hat{y}_{\ell m} = W_\ell\,{}_0\hat{x}_{\ell m}$, for harmonic space kernel defined in the literature (see *e.g.* Price et al., 2021b; Wallis et al., 2021). As ${}_0x$ is not directly observable, typically observations of ${}_2y$ are collected and used to reconstruct ${}_0x$. Suppose one recovers observations $y \in \mathbb{C}^M$ which can be related to the $x \in \mathbb{C}^{N_{\mathbb{S}^2}}$ *via* the forward model $\Phi_{\mathrm{WL}} \in \mathbb{C}^{M \times N_{\mathbb{S}^2}} : x \mapsto y$. Consider the scenario in which observations y are contaminated with *i.i.d.* Gaussian noise $n \sim \mathcal{N}(0,\sigma^2) \in \mathbb{C}^M$, then the observations are modelled by $y = \Phi_{\mathrm{WL}}x + n$ for measurement operator

$$\Phi_{\mathrm{WL}} = \mathsf{D}\,{}_2\mathsf{Y}^{-1}\mathsf{W}\,{}_0\mathsf{Y} \quad \text{and} \quad \Phi^\dagger_{\mathrm{WL}} = {}_0\mathsf{Y}^\dagger\mathsf{W}({}_2\mathsf{Y}^{-1})^\dagger\mathsf{D}^\dagger, \tag{8.26}$$

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

$$x^{\mathrm{MAP}} = \underset{x \in \mathbb{C}^{N_{\mathbb{S}^2}}}{\operatorname{argmin}} \Big[\lambda_\Psi \|\Psi^{-1}x\|_1 + \frac{1}{2\sigma^2}{}_{\mathbb{S}^2}\|\Phi_{\mathrm{WL}}x - y\|_2^2 \Big], \tag{8.27}$$

for automatically marginalized regularisation parameter $\lambda \in \mathbb{R}_{>0}$ (see Section 8.3.1). Note that sparse priors are often adopted in the weak lensing setting, recovering state-of-the-art results (Lanusse et al., 2016; Price et al., 2021a,b). Images from this experiment can be seen in Figure 8.5. A more in depth application of the methods developed in this section to dark matter reconstruction can be found in Price et al. (2021b), in which global hypothesis testing (leveraging the techniques of Section 8.4.2) is performed to determine whether two reconstruction methods produce commensurate estimates.

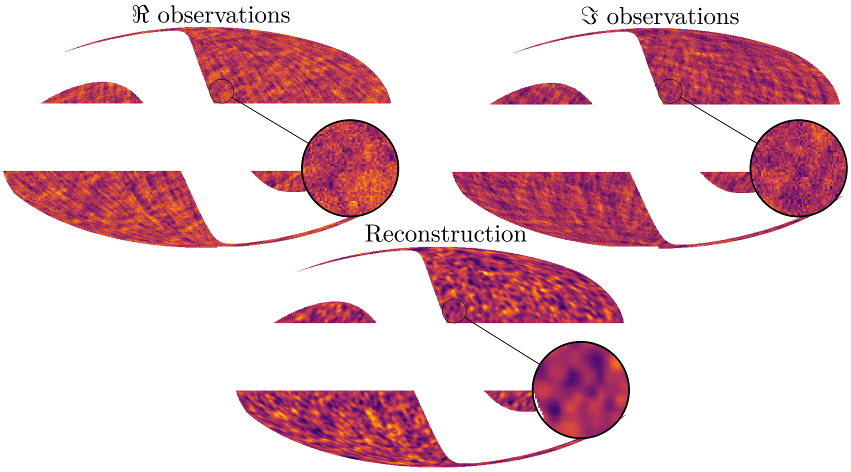


Figure 8.5: Top: Simulated weak lensing shear field generated from a ground truth N-body simulation (Takahashi et al., 2017) signal, which was further contaminated with 5dB *i.i.d.* Gaussian noise and masked using realistic pseudo-Euclid survey mask. **Bottom:** Unconstrained reconstructed dark matter mass-map using ℓ_1-norm wavelet sparsity prior solved through using the proximal forward-backward algorithm in the analysis setting. See related works (Price et al., 2021b) for a comprehensive analysis of this application, with uncertainty quantification.

8.6 Summary

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 (Boyd, 2004) in both the analysis and synthesis settings (Elad et al., 2006; Wallis et al., 2017). The implications, advantages, and disadvantages of each choice within the context of imaging on the sphere are 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 (Pereyra, 2017) to demonstrate how one can perform rapid, statistically principled uncertainty quantification on reconstructed signals (building upon work in Pereyra, 2017; Cai et al., 2018b; Price et al., 2019a,b; Repetti et al., 2019; Price et al., 2021a).

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.

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.

Chapter 9

Dark matter on the celestial sphere

This chapter is based on research presented in M. A. Price, J. D. McEwen, L. Pratley, and T. D. Kitching, "Sparse Bayesian mass-mapping with uncertainties: Full sky observations on the celestial sphere", Monthly Notices of the Royal Astronomical Society, vol. 500, no. 4, pp. 5436-5452, Jan. 2021.

Price et al. (2021b)

Note: *This chapter assumes knowledge of the following. Mathematical context of weak gravitational lensing, presented in sections 4.1.4 and 4.2.3. Bayesian methods, particularly those discussed in section 3.2.2 and 3.2.3. An understanding of analysis of signals over spheres is helpful, including* e.g. *the construction of spherical wavelet dictionaries. The spin operators of section 4.2.4 and wavelet dictionary of section 8.2.2 will be adopted without definition. A general understanding of the content provided in chapter 8 is necessary, the content of chapter 5.*

In this chapter we leverage the results of chapter 8 to extend the hierarchical Bayesian-sparse formalism developed in chapters 5 and 6 to the 2-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 (see sections 3.2.2 and 3.2.3). Throughout this chapter we refer to our estimator, formed within this framework, as the DarkMapper estimator (and by extension the DarkMapper code-base). The reconstruction formalism developed in this chapter, and any uncertainty 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 (see 2.2.1). As such, one can incorporate various experimental or systematic effects in future, *e.g.* more complex (realistic) noise models or intrinsic alignment corrections *etc.*

9.1 Introduction

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., 2021), which is classically ill-posed. The most naive mass-mapping technique for small fields of view is planar Kaiser-Squires (KS; Kaiser and Squires, 1993), which is straightforward 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 9.3).

Many, perhaps more sophisticated, approaches to mass-mapping on the plane have been developed (*e.g.* VanderPlas et al., 2011; Jee et al., 2016; Lanusse et al., 2016; Jeffrey et al., 2018) 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., 2019a,b, 2021a) that provides fully principled statistical uncertainties without the need to assume Gaussianity and without the computational overhead of MCMC methods (*e.g.* Corless et al., 2009; Schneider et al., 2015; 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 (Hu, 2000; Chang et al., 2018) to avoid errors due to projection effects, which can be large (Vallis et al., 2018; Wallis et al., 2021). Naturally one can naively invert the spherical forward model to form the *spherical Kaiser-Squires* estimator (SKS; Wallis et al., 2021) 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).

The structure of this chapter is as follows. Section 9.2 provides a cursory introduction to Bayesian analysis (see chapter 3 for further discussion) before presenting, and discussing, both the general hierarchical Bayesian formalism and our DarkMapper estimator. In this section we explicitly outline the likelihood and weakly informative priors used throughout this paper, but place emphasis on the generality of this formalism. In section 9.3, using high resolution N-body (Takahashi et al., 2017) simulations, pseudo-Euclid masking (a masking of the galactic plane and the ecliptic) and noise realizations 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 9.4 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-map. Furthermore, we perform global Bayesian uncertainty quantification on these reconstructions. Finally, in Section 9.5 we draw conclusions.

9.2 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 *e.g.* Robert (2001). 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.

9.2.1 Bayesian inference

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

$$p(\kappa|\gamma;\mathcal{M}) = \frac{p(\gamma|\kappa;\mathcal{M})p(\kappa;\mathcal{M})}{\int_{\mathbb{C}^N} p(\gamma|\kappa;\mathcal{M})p(\kappa;\mathcal{M})d\kappa}, \tag{9.1}$$

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 $\Phi : \kappa \mapsto \gamma$ 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 discussion.

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 maximisation of the posterior or — due to the monotonicity of the logarithm function — minimisation of the log-posterior,

$$\underset{\kappa}{\text{argmax}}\Big[p(\kappa|\gamma;\mathcal{M})\Big] \equiv \underset{\kappa}{\text{argmin}}\Big[-\log(\,p(\kappa|\gamma;\mathcal{M})\,)\Big]. \tag{9.2}$$

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 optimisation problem for which one may draw on the field of convex optimisation.

9.2.2 Spherical sparse mass-mapping

In this section 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 and 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

$$\Phi \in \mathbb{C}^{M \times N_{\mathbb{S}^2}} : \kappa \in \mathbb{C}^{N_{\mathbb{S}^2}} \mapsto \gamma \in \mathbb{C}^{M}. \tag{9.3}$$

In the spherical setting, by noting the spherical lensing forward model given by equation (4.48) this measurement operator, and its adjoint, naturally take the forms,

$$\Phi = \mathsf{M}\,{}_2\tilde{\mathsf{Y}}\,\mathsf{W}\,{}_0\mathsf{Y} \qquad \text{and} \qquad \Phi^{\dagger} = {}_0\mathsf{Y}^{\dagger}\,\mathsf{W}\,{}_2\tilde{\mathsf{Y}}^{\dagger}\mathsf{M}^{\dagger}, \tag{9.4}$$

where ${}_s\mathsf{Y}$ and ${}_s\tilde{\mathsf{Y}}$ represent the forward and inverse spin-s spherical harmonic transforms respectively, M is a masking operator, and W is harmonic space multiplication by the kernel $\mathcal{W}_\ell$ defined in equation (4.48). It should be noted that, from symmetry, W is trivially self-adjoint. Additionally, it is important to note that adjoint ($\dagger$) spin-s spherical harmonic transforms are not equivalent to the corresponding inverse spherical harmonic transforms — an important caveat often overlooked throughout the field.

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 $\gamma = \Phi\kappa + n$. 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 $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2) \in \mathbb{R}^{M \times M}$. Let $\Phi_i\kappa$ be the value of $\Phi\kappa$ at pixel i, then the overall likelihood is then defined as,

$$p(\gamma|\kappa) \propto \prod_{i=0}^{M} \exp\left(\frac{-(\Phi_i\kappa - \gamma_i)^2}{2\sigma_i^2}\right) = \prod_{i=0}^{M} \exp\left(\frac{-1}{2}\left(\bar{\Phi}_i\kappa - \bar{\gamma}_i\right)^2\right) = \exp\left(\frac{-\|\bar{\Phi}\kappa - \bar{\gamma}\|_2^2}{2}\right), \tag{9.5}$$

where $\|\cdot\|_2$ is the ℓ_2-norm and $\bar{\Phi} = \Sigma^{-\frac{1}{2}}\Phi$ is a composition of the measurement operator and an inverse covariance weighting as defined in Section 9.2.2. 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 9.2.2) 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 pixelization effects, reduced shear (see section 9.2), 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.

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 9.2 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 Ψ are localized 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 and Sweldens, 1995; Barreiro et al., 2000; Narcowich et al., 2006; Starck, J.-L. et al., 2006a; Marinucci et al., 2007; McEwen and Scaife, 2008; Wiaux et al., 2008; Baldi et al., 2009; McEwen et al., 2011, 2018; Chan et al., 2017; McEwen and Price, 2019) with varying localization and un-correlation properties. In this analysis we adopt the scale-discretized wavelets (Wiaux et al., 2008; Leistedt et al., 2013; McEwen et al., 2013b, 2015c) scheme as not only does it satisfy quasi-exponential localization and asymptotic un-correlation properties (McEwen et al., 2018) but also supports directionality which may often be of interest for the weak lensing setting.

Specifically we adopt a Laplace-type wavelet log-prior $\|\Psi(\cdot)\|_1$. Note that, as $\|\cdot\|_1$ is a discretization of the continuous ℓ_1-norm it must be re-weighted 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 chapter any reference to the ℓ_1-norm applied to a spherical space refers explicitly to this spherically re-weighted norm. With our choice of ℓ_1-norm regularisation the prior can be written compactly as

$$p(\kappa) \propto \exp\Big(-\mu \|\tilde{\Psi}^{\dagger}\kappa\|_1\Big), \tag{9.6}$$

where $\tilde{\Psi}^{\dagger}$ is the analysis forward-adjoint spherical wavelet transforms (see chapter 8 for details) with coefficients $\tilde{\Psi}_i^{\dagger}$, and $\mu \in \mathbb{R}_{>0}$ is the regularisation parameter (see chapter 5 for details on how this parameter is computed). It is assumed here that the spherical wavelet dictionary $\tilde{\Psi}$ 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 thought 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 $\tilde{\Psi}$ 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 (see *e.g.* Mallat, 2008).

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 *e.g.* Horowitz et al., 2019; Price et al., 2021c, for alternate iterative Wiener filtering approaches), or a flat prior which produces the *maximum likelihood estimate* (MLE).

Implementation

The minimisation of the log-posterior in equation (9.2) is (in the analysis setting) therefore precisely the same as solving,

$$\kappa^{\text{map}} = \underset{\kappa \in \mathbb{C}^{N_{\mathbb{S}^2}}}{\operatorname{argmin}} \Bigg[\underbrace{\mu \|\Psi^\dagger \kappa\|_1 + \frac{\|\bar{\Phi}\kappa - \bar{\gamma}\|_2^2}{2}}_{\text{Objective function}} \Bigg]. \tag{9.7}$$

We solve this convex optimisation problem using the S2INV code (see *e.g.* Price et al., 2021c, and chapter 8) which is largely built around the SOPT C++ object oriented framework[1] (Carrillo et al., 2012, 2013; Onose et al., 2016; Pratley et al., 2018), utilizing an adapted proximal forward-backward splitting algorithm (Combettes and Pesquet, 2011), although a variety of alternate algorithms are provided within S2INV. Wavelet transforms on the sphere are computed using S2LET[2] (McEwen and Wiaux, 2011; Leistedt et al., 2013; McEwen et al., 2015b,c; Chan et al., 2017; McEwen et al., 2018), which in turn makes use of SSHT[3] (McEwen and Wiaux, 2011; McEwen et al., 2013a) to compute spherical harmonic transforms, and SO3[4] (McEwen et al., 2015c) to compute Wigner transforms.

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 9.1, for full details of the derivation of the proximal forward-backward algorithm iterations look to Combettes and Pesquet (2011). These primary optimisations are terminated once the objective function is updated by less than a set threshold (in our experiments 10^{-6}) between iterations.

Reduced shear

Figure 9.1 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 $g(\omega) = \gamma(\omega)[1-\kappa(\omega)]^{-1}$, 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 (see *e.g.* Mediavilla et al., 2016; Price et al., 2021a; Wallis et al., 2021). 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 9.1.

---

[1] `https://github.com/astro-informatics/sopt`
[2] `http://astro-informatics.github.io/s2let/`
[3] `https://astro-informatics.github.io/ssht/`
[4] `http://astro-informatics.github.io/so3/`

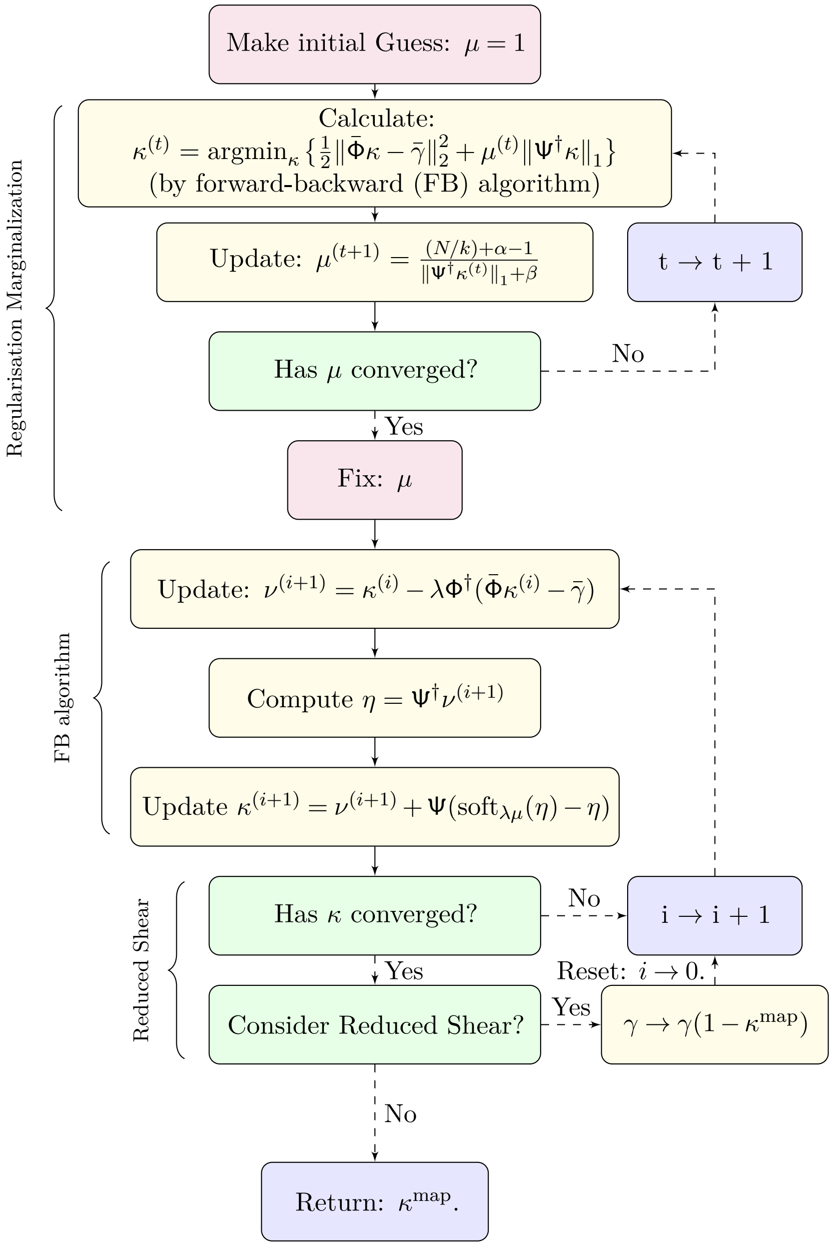


Figure 9.1: Schematic of proximal forward-backward splitting algorithm used (Combettes and Pesquet, 2011). Note that the first iterative block represents the Majorise-Minimisation (MM) algorithm marginalization over the regularisation parameter (which here is treated as a nuisance parameter), the second iterative block represents the primary proximal forward backward iterations, and the final (optional) block represents the reduced shear outer iterations. Note that the $\text{soft}_{\lambda,\mu}(\eta)$ 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).

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.

Computational efficiency

As discussed in section 9.2, 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 dominated 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 marginalisation preliminary iterations and subsequent annealing iterations to optimise convergence, we find $\mathcal{O}(10^2)$ iterations are sufficient for convergence. We consider axisymmetric wavelets ($N = 1$), 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 $\mathcal{O}(n_{\text{samples}}/10^2)$ 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 $\mathcal{O}(\text{days})$.

9.3 Simulated observations

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 chapter 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.

9.3.1 Data-set

Throughout this analysis we perform reconstructions and uncertainty quantification on simulated convergence maps generated from the high resolution Takahashi N-body simulation datasets (Takahashi et al., 2017)[5]. These mock convergence maps are generated *via* multiple-lens plane ray-tracing, and are provided for a range of co-moving 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 $\Omega_m = 1, \Omega_\Lambda = 0.279, \Omega_{\text{cdm}} = 0.233, \Omega_b = 0.046, h = 0.7, \sigma_8 = 0.82$ 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.

9.3.2 Method

As in previous work (Price et al., 2019a,b, 2021a) we begin by applying the measurement operator Φ (see equation 9.4) 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 9.3.2) for each pixel i individually and used to construct a known diagonal covariance Σ.[6] 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 section 9.2, to these noisy artificial measurements γ_n to create estimates of the fiducial convergence map κ. For DarkMapper we simply adopt dyadic axisymmetric spherical wavelets ($N = 1$ and $\lambda = 2$ for simplicity), with scale-discretized 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-discretized wavelets are only one possible choice of spherical wavelets (see Section 9.2). Other wavelets on the sphere could be adopted and are interchangeable within this reconstruction formalism, provided they support exact synthesis of a signal from its wavelet coefficients.

---

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

[6] Note we here do note consider off diagonal terms which may arise due to intrinsic galaxy alignments though in future this can be incorporated, albeit perhaps not straightforwardly.

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,

$$\text{Recovered SNR} = 20 \times \log_{10}\left(\frac{\|\kappa\|_2}{\|\kappa - \kappa^{\text{map}}\|_2}\right), \tag{9.8}$$

from which it is clear that the larger the recovered SNR the more accurate[7] the convergence estimator. Additionally, we record the Pearson correlation coefficient between recovered convergence estimators $\kappa^{\text{map}} \in \mathbb{C}^{N_{\mathbb{S}^2}}$ and the fiducial convergence $\kappa \in \mathbb{C}^{N_{\mathbb{S}^2}}$ as a measure of topological fidelity of the estimator. The Pearson correlation coefficient is defined to be

$$r = \frac{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa^{\text{map}}(i) - \bar{\kappa}^{\text{map}}\}\{\kappa(i) - \bar{\kappa}\}}{\sqrt{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa^{\text{map}}(i) - \bar{\kappa}^{\text{map}}\}^2}\sqrt{\sum_{i=1}^{N_{\mathbb{S}^2}}\{\kappa(i) - \bar{\kappa}\}^2}}, \tag{9.9}$$

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 optimisation algorithms.

Noise simulation

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 square arc-minute), 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,

$$\sigma_i = \sqrt{\frac{\sigma_e^2}{A \times (180/\pi)^2 \times 3600 \times n_{\text{gal}}}}, \tag{9.10}$$

where $3600(180/\pi)^2$ converts steradians to arcmin^2 — this relation is simply a

---

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

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) capture more observations have smaller noise variance. In practice the value of σ_i (and therefore the 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.

9.3.3 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 9.2 and numerically in Table 9.1. 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 9.8) 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 $n_{\text{gal}} = 30$, the increase in performance SNR (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.

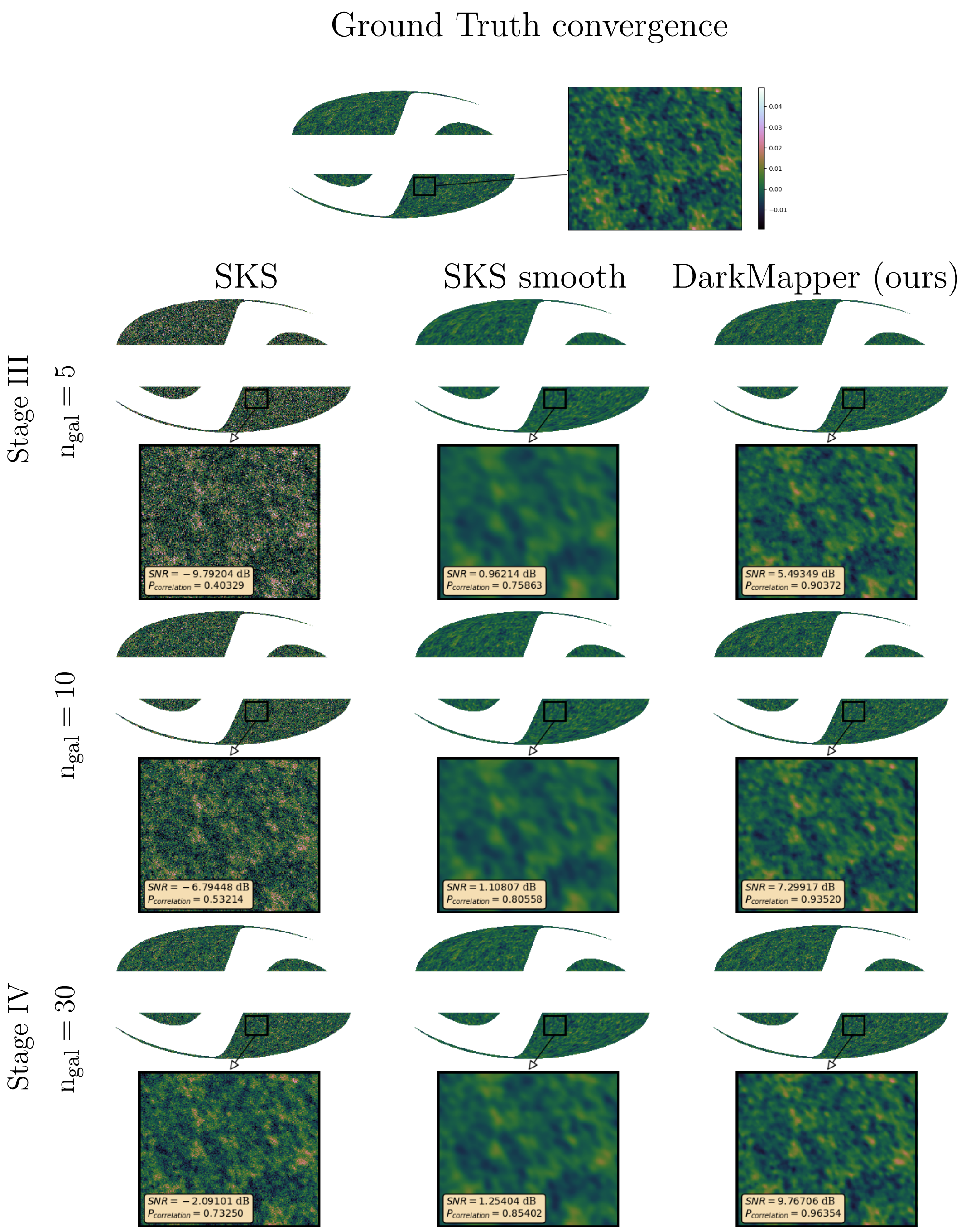


Figure 9.2: The top row displays the ground truth Takahashi convergence map as described in Section 9.3.1 with close up of a small region. **Top to bottom:** Increasing number density of galaxies (n_{gal}) and therefore decreasing noise levels. At the top we have $n_{\text{gal}} = 5$ which is representative of current Stage III surveys, at the bottom we have $n_{\text{gal}} = 30$ which has been forecast for upcoming Stage IV surveys, *e.g.* Euclid or LSST. **Left to right:** The spherical Kaiser-Squires (Wallis et al., 2021) estimator without the *ad hoc* smoothing kernel post-processing, the optimally smoothed spherical 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 colour-scale to aid comparison (Green, 2011).

$\ell_{\max} = 2048$		SKS		SKS (smooth)		DarkMapper		Difference
Setting	n_{gal}	SNR (dB)	P_{cor}	SNR (dB)	P_{cor}	SNR (dB)	P_{cor}	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)
Ideal	100	2.956	0.887	n/a	n/a	12.132	0.980	+9.176 (n/a)

Table 9.1: Numerical results from reconstructions of Takahashi simulations as discussed in Sections 9.3.1 and 9.3.2. 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_{cor}. 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 (see *e.g.* Vallis et al., 2018; Wallis et al., 2021) 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 $n_{\text{gal}} = 100$ setting for the SKS estimator and so was recorded as n/a.

9.4 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 (Flaugher and DES Collaboration, 2015; Abbott et al., 2018; Morganson and DES Collaboration, 2018), CFHTLens (Erben et al., 2012), and the KiDS450 (Fenech Conti et al., 2017; Hildebrandt 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)

9.4.1 Joint spherical dark matter 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 $w(i) \in \mathbb{R}_{>0}$ 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

$$\Re[\gamma(i)] = \frac{\sum_j w(j)[e_1(j) - c_1(j)]}{\sum_j w(j)[1 + \text{mcorr}(j)]}, \qquad \Im[\gamma(i)] = \frac{\sum_j w(j)[e_2(j) - c_2(j)]}{\sum_j w(j)[1 + \text{mcorr}(j)]}, \tag{9.11}$$

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 γ respectively. This joint global dataset was then projected onto an equi-angularly 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 9.2 and 9.3.2)

To this spherical shear map γ_{obs} (with corresponding data covariance Σ_{obs}) we apply DarkMapper outlined in Section 9.2 with the same parameter choices outlined in Section 9.3 (see Price et al., 2021a, for a planar equivalent). Additionally, we provide the SKS (Wallis et al., 2021) 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 $= \Theta = 25$ arcmins)[8]. The results of all reconstruction algorithms can be seen both globally, and with enhanced regions in Figure 9.3, where all subplots share the same colour-scale (Green, 2011). It is immediately apparent that the SKS estimator, in the absence of smoothing, is overwhelmingly dominated by noise (hence the motivation for post-processing).

In contrast to this, the SKS estimator with a $\Theta = 25$ 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 9.3. 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.

---

[8] All data products aforementioned within this section are publicly available and may be found at `https://doi.org/10.5281/zenodo.3980652`.

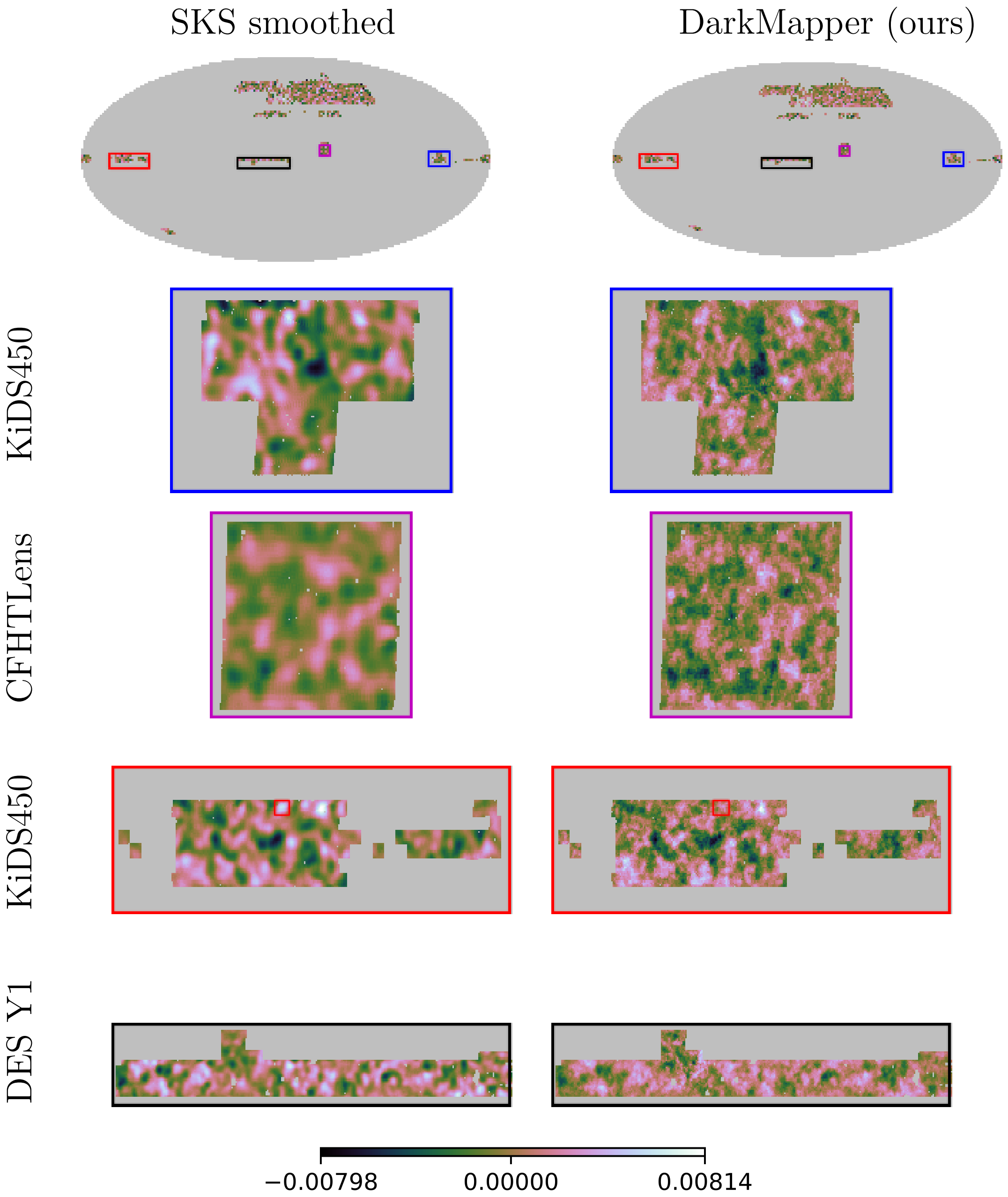


Figure 9.3: Top to bottom: Global reconstruction of the majority of public weak lensing datasets, magnified view of a variety of patches. **Left to right:** Spherical Kaiser-Squires (SKS) estimator without Gaussian smoothing kernel, SKS estimator with FWHM $= \Theta = 25$ arcmin smoothing kernel (as 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 colour-scale to aid comparison (Green, 2011). As these plots are enhanced regions of Mollweide projections they exhibit very slight visual distortions which are not present in the true datasets, which is a limitation of planar projections (see *e.g.* Wallis et al., 2021). These data-sets can be found online at `https://doi.org/10.5281/zenodo.3980652`.

9.4.2 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 9.3 but not in the corresponding DarkMapper estimator. We did so by performing local hypothesis testing of structure as described in chapter 5 (see *e.g.* Price et al., 2021a, for more comprehensive details).

In all cases the hypothesis test of local structure could not reject the existence 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) and the extremely high dimensionality $\sim \mathcal{O}(10^7)$ of the reconstruction (which directly increases the level set threshold ϵ'_α in equation 10.12; see section 3.2.3).

9.4.3 Global uncertainty quantification

For high dimensional cases it is often more informative to consider global features of the reconstruction (Price et al., 2021a), see chapters 5 and 8 for further discussion. 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 $\Theta \sim 30$ arcminute smoothing. In this way we can directly address the question of which smoothing scales produces solutions $\kappa^{\text{sur}}(\Theta)$ which 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 9.2. 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 rigorous 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,

Surrogate κ^{sur}		Analysis ($\epsilon'_{99\%} = 842789$)		Hypothesis test
Estimator	Θ (arcmin)	Obj(κ^{sur})	Obj(κ^{sur})/$\epsilon'_{99\%}$	$\kappa^{\text{sur}} \in C'_{99\%}$
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 9.2: 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 testing. **Discussion:** Clearly, the SKS estimator without smoothing is unequivocally 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.

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.

9.5 Summary

In this chapter we have extended the previously presented (Price et al., 2021a) sparse Bayesian reconstruction formalism to the spherical setting (see chapter 5) by adopting the general framework presented in chapter 8, resulting in a sparse spherical Bayesian mass-mapping algorithm which we refer to as DarkMapper. This algorithm is general and accommodates any log-concave posterior. Throughout this analysis 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., 2021) 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 discussion, 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 of SKS as strongly as possible). We extend approximate Bayesian uncertainty quantification methods (Pereyra, 2017; Cai et al., 2018b; Price et al., 2019a,b; Repetti et al., 2019; Price et al., 2021a) 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 (Flaugher and DES Collaboration, 2015; Abbott et al., 2018; Morganson and DES Collaboration, 2018), CFHTLens (Erben et al., 2012), and the KiDS450 (Fenech Conti et al., 2017; Hildebrandt et al., 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 arcminutes 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 an avenue for future investigation (see chapter 10).

Chapter 10

Variational regularisation on the ball

This chapter is based on research presented in M. A. Price, and J. D. McEwen, "Bayesian variational regularisation on the ball", submitted to IEEE Signal Processing Letters, 2021.

Price and McEwen (2021)

Note: *This chapter is highly technical and assumes knowledge of the following. Convex optimisation algorithms and proximal analysis, as presented in chapter 2. Bayesian methods, particularly those discussed in section 3.2.2 and 3.2.3. A general understanding of the content provided in chapter 5 is helpful. Look to Leistedt and McEwen (2012) for further details of Fourier-Laguerre wavelet construction* etc.

Having developed a framework for general inverse problems over the sphere in chapter 8, and leveraging said framework for spherical mass-mapping in chapter 9, we will now develop an analogous framework within which one may consider inverse problems defined over the full 3D ball, *e.g.* universal dark matter cartography (see *e.g.* Massey et al., 2007a,b). In this final chapter we develop scalable techniques, with associated open-source software, which leverage variational regularisation methods to solve ill-posed and/or ill-conditioned inverse problems natively on the 3D ball (see section 2.1.2 and chapter 2 more broadly). Furthermore, leveraging recent developments in the theory of probability density theory (Pereyra, 2017), we demonstrate how convex variational regularisation techniques can be combined with advances in probability density theory to construct computationally efficient signal reconstruction techniques on the ball with principled uncertainty quantification, or 'Bayesian variational regularisation' (see sections 3.2.2 and 3.2.3).

10.1 Introduction

Inverse problems on Euclidean manifolds have been researched extensively and associated techniques have found effective application in countless domains. However, increasingly often one wishes to consider inverse problems defined on curved, non-Euclidean manifolds, *e.g.* diffusion magnetic resonance imaging (MRI; Tuch, 2004) and 2D dark matter reconstructions on the sphere, and many aspects of geophysics

(Simons et al., 2011; Marignier et al., 2020; Kendall et al., 2021), astrophysics (Heavens, 2003; Leistedt et al., 2015), and molecular modelling (Boomsma and Frellsen, 2017) on the 3D ball, for which very few techniques have been developed.

Inverse problems are often solved by Bayesian Markov chain Monte Carlo (MCMC) sampling methods (see section 3.2.1) or variational approaches (optimisation *etc.*, see section 2.2). MCMC methods are highly computationally demanding on the ball, due to the computational complexity of transforms on curved manifolds, and are unfeasible for many applications. Variational methods, which solve inverse problems through classical optimisation techniques, are typically scalable and robust, supporting both theoretical guarantees, and can support principled uncertainty quantification (see section 3.2.3 and chapter 5). Such techniques are thus perfectly suited to scientific analysis on the ball, where computational efficiency and probabilistic interpretations are highly desirable. Variational methods have been considered over the sphere (McEwen et al., 2013a; Wallis et al., 2017; Price et al., 2021c), often leveraging ideas from compressed sensing (Candès et al., 2006; Donoho, 2006), and typically promoting sparsity in spherical wavelet dictionaries, (*e.g.* Leistedt et al., 2013; Chan et al., 2017; McEwen and Price, 2019), to recover state-of-the-art results. Spherical techniques have been used tomographically (as concentric spherical shells) to model radially distributed datasets, however holistic approaches, which perform analysis natively on the underlying manifold (the ball), are crucially missing. Wavelet transforms on the 3D ball have been developed to support radially distributed problems (Michel, 2005; Fengler et al., 2006; Lanusse et al., 2012; Leistedt and McEwen, 2012; Durastanti et al., 2014; Khalid et al., 2016), however these dictionaries have, to our best knowledge, not been leveraged to perform variational inference on the ball.

10.2 Bayesian variational regularisation on ball

In this section we develop mathematical techniques for the analysis of spin signals on the 3D ball and wavelets on the directional ball, scalable convex optimisation algorithms on the ball, and variational regularisation techniques which support principled Bayesian uncertainty quantification. Throughout we adopt separable eigenfunctions on the ball, with radial basis functions given by the Laguerre polynomials (Pollard, 1947; Weniger, 2008) and angular basis functions given by the spin spherical harmonics ${}_sY_{\ell m}$ (Newman and Penrose, 1966; Goldberg et al., 1967; McEwen and Wiaux, 2011; McEwen et al., 2015c). As spin spherical harmonic transform are more common in the associated literature, we will focus primarily on the novel radial components (Leistedt and McEwen, 2012), and the Bayesian interpretation (Pereyra, 2017; Cai et al., 2018b; Price et al., 2021a,b). Many of the spherical analysis details can be found in chapter 8, should they be required.

10.2.1 Spin signals on the ball

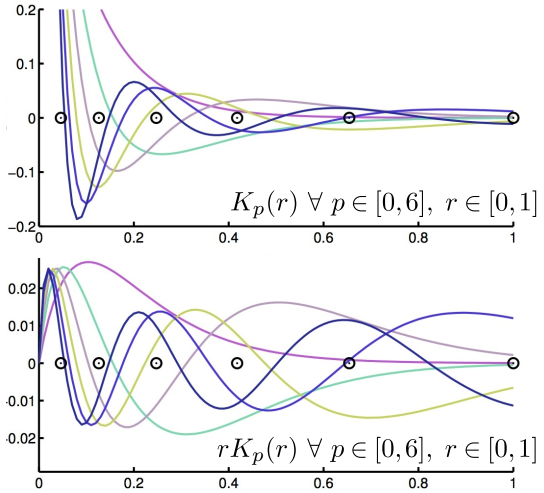


Figure 10.1: First six spherical Laguerre basis functions $K_p(r)$ for $r \in [0,1]$. Functions $rK_p(r)$ can be viewed as basis functions in Cartesian co-ordinates, satisfying the usual orthogonality relation $\langle rK_p(r)|rK_q(r)\rangle_{\mathbb{R}_{>0}} = \delta_{pq}$. This figure is originally from Leistedt and McEwen (2012), to which the reader is referred for further details.

Here we discuss the construction of Spherical-Laguerre basis functions on the 3D ball, developed in previous work (Leistedt and McEwen, 2012) and adopted throughout this chapter. First let us define the Laguerre basis functions along the radial half-line $K_p(r)$ as

$$K_p(r) \equiv \sqrt{\frac{p!}{(p+2)!}}\frac{e^{\frac{-r}{2\tau}}}{\sqrt{\tau^3}} L_p^{(2)}(\frac{r}{\tau}), \quad (10.1)$$

where $L_p^{(2)}$ is the p^{th}-associated 2^{nd}-order Laguerre polynomial (Pollard, 1947; Weniger, 2008), and $\tau \in \mathbb{R}_{>0}$ is a scale factor that adds a scaling flexibility. These basis functions (see Figure 10.1) are orthonormal on $\mathbb{R}_{>0}$, *i.e.* $\langle K_p|K_q\rangle_{\mathbb{R}_{>0}} = \delta_{pq}$, and complete, by Gram-Schmidt orthogonalization and exploiting polynomial completeness on $\mathsf{L}^2(\mathbb{R}_{>0}, r^2 e^{-r}\,\mathrm{d}r)$ (Leistedt and McEwen, 2012). Any square-integrable function $f \in \mathsf{L}^2(\mathbb{R}_{>0})$ can be projected into this basis as

$$f_p = \langle f|K_p\rangle = \int_{\mathbb{R}_{>0}} \mathrm{d}r r^2 f(r) K_p(r), \tag{10.2}$$

which supports exact synthesis by

$$f(r) = \sum_{p=0}^{\infty} f_p K_p(r). \tag{10.3}$$

Real-world functions are typically to a good approximation bandlimited, *i.e.* the Fourier-Laguerre coefficients of signals $f \in \mathbb{R}_{>0}$ are such that $f_p = 0, \forall p \geq P$, and so this summation is truncated at P. We adopt the Gauss-Laguerre quadrature (see *e.g.* Press et al., 2007), which is commonly used to numerically evaluate integrals over the radial half-line, and was used to develop an exact sampling theorem on Spherical-Laguerre space (Leistedt and McEwen, 2012).

Suppose we adopt these radial basis functions which we then combine with the spin-s spherical harmonic angular basis functions ${}_sY_{\ell m}(\omega)$ (Newman and Penrose, 1966; Goldberg et al., 1967) for $s \in \mathbb{Z}$ and $\omega = (\theta, \psi) \in \mathbb{S}^2$, where $\theta = [0, \pi)$ is the colatitude and $\psi \in [0, 2\pi)$ is the longitude. In such a case, we can straightforwardly define the Spherical-Laguerre basis functions ${}_sZ_{\ell m p}(\boldsymbol{r}) = K_p(r)\,{}_sY_{\ell m}(\omega)$ for

$\boldsymbol{r} = (r, \omega) \in \mathbb{B}^3 := \mathbb{R}_{>0} \times \mathbb{S}^2$, which are orthogonal $\langle {}_sZ_{\ell mp} | {}_sZ_{\ell' m' p'} \rangle_{\mathbb{B}^3} = \delta_{\ell\ell'}\delta_{mm'}\delta_{pp'}$, and onto which any square integrable spin-s function on the ball ${}_sf \in \mathsf{L}^2(\mathbb{B}^3)$ can be projected by

$$
{}_sf_{\ell mp} = \langle {}_sf | {}_sZ_{\ell mp} \rangle_{\mathbb{B}^3} = \int_{\mathbb{B}^3} \mathrm{d}\mu(\boldsymbol{r}) {}_sf(\boldsymbol{r}) {}_sZ^*_{\ell mp}(\boldsymbol{r}), \tag{10.4}
$$

where $\mathrm{d}\mu(\boldsymbol{r}) = \mathrm{d}^3\boldsymbol{r} = r^2 \sin\theta \, \mathrm{d}r \, \mathrm{d}\theta \, \mathrm{d}\psi$ is the rotation invariant measure (Haar measure) on the ball. By considering the separability and completeness of angular and radial basis functions this projection supports exact synthesis, such that

$$
{}_sf(\boldsymbol{r}) = \sum_{p=0}^{P-1} \sum_{\ell=0}^{L-1} \sum_{m=-\ell}^{\ell} {}_sf_{\ell mps} Z_{\ell mp}(\boldsymbol{r}), \tag{10.5}
$$

where $L, P \in \mathbb{Z}_{>0}$ are the angular (McEwen and Wiaux, 2011) and radial (Leistedt and McEwen, 2012) bandlimits respectively. In this work, by considering the relations presented in this subsection, fast adjoint Spherical-Laguerre transforms were constructed, facilitating variational regularisation on the ball (see Section 10.3).

10.2.2 Directional scale-discretized spin wavelets on the ball

Here we extend the Spherical-Laguerre wavelets on the 3D ball developed in previous work (Leistedt and McEwen, 2012) to 4D directional scale-discretized wavelets on the ball. Furthermore, we extend the discussion to include spin-signals, which arise in various areas of physics *e.g.* quantum mechanics and weak gravitational lensing (Price et al., 2021b). Consider the radial translation operator $\mathcal{T}_r$ for $r \in \mathbb{R}_{>0}$ (see *e.g.* Leistedt and McEwen, 2012; McEwen and Leistedt, 2013, for further details), and rotation $\mathcal{R}_\rho$, for Euler angles $\rho = (\alpha, \beta, \gamma) \in \mathrm{SO}(3)$ with $\alpha \in [0, 2\pi)$, $\beta \in [0, \pi)$, and $\gamma \in [0, 2\pi)$, with action $(\mathcal{R}_\rho {}_sf)(\omega) \equiv e^{-is\theta} {}_sf(\mathcal{R}_\rho^{-1}\boldsymbol{\omega})$. Further define the concatenation of these transforms to be the 4D transformation $\mathcal{L}_{\boldsymbol{h}} = \mathcal{T}_r \mathcal{R}_\rho$ for $\boldsymbol{h} = (r, \rho) \in \mathbb{H}^4 := \mathbb{R}_{>0} \times \mathrm{SO}(3)$. Leveraging this composite transformation one can straightforwardly define the directional wavelet coefficients $W^{{}_s\Psi^{jj'}} \in \mathsf{L}^2(\mathbb{H}^4)$ of any square integrable spin-s function ${}_sf \in \mathsf{L}^2[\mathbb{B}^3]$ by the directional convolution $\circledast$

$$
W^{{}_s\Psi^{jj'}}(\boldsymbol{h}) \equiv ({}_sf \circledast {}_s\Psi^{jj'})(\boldsymbol{h}) \equiv \langle {}_sf, \mathcal{L}_{\boldsymbol{h}} {}_s\Psi^{jj'} \rangle_{\mathbb{B}^3} = \int_{\mathbb{B}^3} \mathrm{d}\mu(\boldsymbol{r}) {}_sf(\boldsymbol{r}) (\mathcal{L}_{\boldsymbol{h}} {}_s\Psi^{jj'})^\star(\boldsymbol{r}), \tag{10.6}
$$

where ${}_s\Psi^{jj'} \in \mathsf{L}^2[\mathbb{B}^3]$ is the wavelet kernel at angular and radial scales $j, j' \in \mathbb{Z}_{>0}$ respectively. These scales determine the volume over which a given wavelet function has compact support (Leistedt and McEwen, 2012). Typically, wavelet coefficients do not capture low frequency signal content, which instead is captured by axisymmetric scaling functions ${}_s\Upsilon \in \mathsf{L}^2(\mathbb{B}^3)$ with coefficients $W^{{}_s\Upsilon} \in \mathsf{L}^2(\mathbb{B}^3)$ defined by the axisymmetric convolution $\odot$ with a spin-s signal ${}_sf \in \mathsf{L}^2(\mathbb{B}^3)$ such that

$$
W^{{}_s\Upsilon}(\boldsymbol{r}) \equiv ({}_sf \odot {}_s\Upsilon)(\boldsymbol{r}) \equiv \langle {}_sf, \mathcal{L}_{\boldsymbol{r}} {}_s\Upsilon \rangle_{\mathbb{B}^3} = \int_{\mathbb{B}^3} \mathrm{d}\mu(\boldsymbol{r}') {}_sf(\boldsymbol{r}') (\mathcal{L}_{\boldsymbol{r}} {}_s\Upsilon)^\star(\boldsymbol{r}'), \tag{10.7}
$$

where $\mathcal{L}_{\boldsymbol{r}}$ is an axisymmetric simplification of the full 4D transformation $\mathcal{L}_{\boldsymbol{h}}$. For suitable choices of wavelet and scaling generating functions (those which satisfy wavelet admissibility) these projections support exact synthesis by

$$
{}_s f(\boldsymbol{r}) = \int_{\mathbb{B}^3} \mathrm{d}\mu(\boldsymbol{r}') W^{{}_s\Upsilon}(\boldsymbol{r}')(\mathcal{L}_{\boldsymbol{r}'}{}_s\Upsilon)(\boldsymbol{r}) + \sum_{j=J_0}^{J} \sum_{j'=J'_0}^{J'} \int_{\mathbb{H}^4} \mathrm{d}\mu(\boldsymbol{h}) W^{{}_s\Psi^{jj'}}(\boldsymbol{h})(\mathcal{L}_{\boldsymbol{h}}{}_s\Psi^{jj'})(\boldsymbol{r}), \tag{10.8}
$$

where $\mathrm{d}\mu(\boldsymbol{h}) = \mathrm{d}^4\boldsymbol{h} = r^2 \sin\beta \, \mathrm{d}r \, \mathrm{d}\alpha \, \mathrm{d}\beta \, \mathrm{d}\gamma$ is the Haar measure on $\mathbb{H}^4$. By construction (Leistedt and McEwen, 2012) this wavelet dictionary exhibits both good frequency and spatial localisation, permits exact synthesis, and leverages optimal sampling theories for efficient transforms. Furthermore, by adopting adjoint Spherical-Laguerre transforms (see subsection 10.2.1) fast adjoint 4D wavelet transforms on the ball were constructed.

10.2.3 Efficient transformations over the ball

Variational methods on the ball require an additional level of complexity over those defined on the spherical manifolds, which are already significantly computationally expensive. The forward and inverse Spherical-Laguerre transforms are computed through the FLAG[1] package (Leistedt and McEwen, 2012) with computational complexity $\sim \mathcal{O}(L^4)$, built on spin spherical harmonic transforms provided by the SSHT[2] package (McEwen and Wiaux, 2011; McEwen et al., 2013a). Similarly, forward and inverse wavelet transforms on the 3D ball are computed through the FLAGLET[3] package (Leistedt and McEwen, 2012) with computational complexity $\sim \mathcal{O}(NL^4)$, where $N \in \mathbb{Z}_{>0}$ is the wavelet directionality, built on the wavelet transforms provided by the S2LET[4] package (Wiaux et al., 2008; Leistedt et al., 2013; McEwen et al., 2015c; Chan et al., 2017; McEwen et al., 2018; McEwen and Price, 2019). Both these transforms on the ball (FLAG and FLAGLET) scale at least quartically with bandlimit. Therefore, even optimally sampled transforms on the ball are very computationally expensive, motivating attention to scalable implementations.

10.2.4 Maximum a posteriori estimation

Consider measurements $y \in \mathbb{R}^M$, *e.g.* observations on the sky with some radial component, which may be related to some intrinsic underlying field of interest on the ball $x \in \mathbb{R}^{N_{\mathbb{B}^3}}$ by a sensing operator $\Phi \in \mathbb{R}^{M \times N_{\mathbb{B}^3}} : x \mapsto y$. Further suppose measurements are polluted with noise n, then our measurement model is generally given by $y = \Phi x + n$, which is both classically ill-posed in the sense of Hadamard (Hadamard, 1902) and may be seriously ill-conditioned. There are many methods for inferring x from y, in this work we will consider a Bayesian variational approach, so as

---

[1] `https://astro-informatics.github.io/flag/`

[2] `https://astro-informatics.github.io/ssht/`

[3] `http://astro-informatics.github.io/flaglet/`

[4] `https://astro-informatics.github.io/s2let/`

to benefit from the computational efficiency of variational methods (a key component on the ball) whilst retaining the principled statistical interpretation provided by Bayesian methods (see *e.g.* Robert, 2001)

In a Bayesian sense, given a sufficient understanding of our physical system (including *e.g.* the forward model and the noise distribution *etc.*) we can assign a likelihood distribution $P(y|x;\Phi)$, which acts as a data-fidelity constraint on our solutions. Furthermore, suppose we have some *a priori* knowledge as to the nature of our latent variable, *e.g.* x is presumed to be sparse in a given dictionary, then we can straightforwardly define a Bayesian prior distribution $P(x)$, which acts as a regularisation functional to stabilize our inference. With these distributions defined we can construct our posterior distribution through Bayes' theorem

$$P(x|y;\Phi) \propto P(y|x;\Phi)P(x), \tag{10.9}$$

where we drop the normalization term (Bayesian evidence) as it does not affect our solution, and for simplicity. A reasonable choice of solution, in a Bayesian sense, is that which maximizes the posterior odds (*i.e.* the most likely one), called the *maximum a posteriori* (MAP) solution (see section 3.2.2 for further details), given by

$$x^{\text{MAP}} \equiv \underset{x\in\mathbb{R}^{N_{\mathbb{B}^3}}}{\text{argmax}} \Big[P(x|y;\Phi) \Big] \propto \underset{x\in\mathbb{R}^{N_{\mathbb{B}^3}}}{\text{argmin}} \Big[\underbrace{-\log(\, P(y|x;\Phi)P(x)\,)}_{h(x)=f(x)+g(x)} \Big] \tag{10.10}$$

where the second line comes from the monotonicity of the logarithm function. The final line highlights that MAP estimation, for the common class of log-concave distributions, yields convex objectives $h(x)$, and this is equivalent to unconstrained convex optimisation. Such optimisation problems typically leverage 1^{st}-order information to efficiently converge to global (from convexity) extremal solutions (Combettes and Pesquet, 2011). For convex but non-differentiable objectives (*e.g.* sparsity priors) gradient information is accessed through the proximal projection (Moreau, 1962), and thus extremal solutions are efficiently recovered *via* proximal optimisation algorithms (Boyd et al., 2011; Combettes and Pesquet, 2011). Such algorithms permit strong guarantees of both convergence and rate of convergence, however they still only recover point estimates and do not naively support uncertainty quantification.

10.2.5 Uncertainty quantification of MAP estimation

Bayesian methods often consider credible regions (regions of high probability concentration) $C_\alpha \subset \mathbb{C}^{N_{\mathbb{B}^3}}$ of the full posterior distribution, at $100(1-\alpha)\%$ confidence, by evaluating

$$P(x \in C_\alpha|y;\Phi) = \int_{x\in\mathbb{R}^{N_{\mathbb{B}^3}}} P(x|y;\Phi)\mathbb{I}_{C_\alpha} dx = 1-\alpha, \tag{10.11}$$

which is computationally intractable in high dimensional settings, such as data on the ball, even for moderate resolutions. In our method we adopt a recently derived conservative approximation (which is valid for all log-concave posteriors or convex objectives) to the highest posterior density (HPD) credible set $C'_\alpha \supseteq C_\alpha$ (Pereyra, 2017) defined by

$$C'_\alpha \subset \mathbb{C}^{N_{\mathbb{B}^3}} := \left\{ x : h(x) \leq \epsilon'_\alpha \right\} \quad \text{for} \quad \epsilon'_\alpha = h(x^{\text{MAP}}) + \sqrt{16N\log(3/\alpha)} + N, \quad (10.12)$$

which allows one to approximate C_α with knowledge only of the MAP solution x^{MAP} and the dimension $N_{\mathbb{B}^3}$. This approximation is discussed at length in section 3.2.3, to which we recommend the interested reader. This is a crucial realization for variational methods on complex manifolds (such as the ball), as the necessity for scalable, computationally efficient approaches is paramount. Furthermore, the approximation error is bounded above (Pereyra, 2017) thus affording sensitivity guarantees (*i.e.* ϵ'_α cannot become arbitrarily larger than ϵ_α). The error in this approximation has been assessed in a variety of application domains (Cai et al., 2018b; Price et al., 2019a) and has been benchmarked against proximal MCMC methods (Pereyra, 2016).

A number of uncertainty quantification techniques have recently been developed which are built around this approximation, in a variety of settings, many of which exploit linearity (Price et al., 2021c) to facilitate extremely rapid computation. In this chapter we consider, for the first time on the ball, perhaps the most straightforward uncertainty quantification technique, Bayesian hypothesis testing (Cai et al., 2018b; Repetti et al., 2019; Price et al., 2021a). Bayesian hypothesis testing is conducted as follows. A feature of x^{MAP} is adjusted to construct a surrogate solution x^{SUR} from which it is determined if this solution belongs to the credible set at confidence $100(1-\alpha)\%$. If x^{SUR} does not belong to C'_α then it necessarily does not belong to C_α (from the conservative nature of the approximation in Equation 10.12) and therefore the feature is statistically significant at $100(1-\alpha)\%$ confidence. Conversely, if $x^{\text{SUR}} \in C'_\alpha$ then the statistical significance of the feature of interest is indeterminate. In this section we consider features $\Omega \subset x^{\text{MAP}}$ to be local sub-structure and thus hypothesis tests in this case relate to the physicality of local structure, *i.e.* whether these structures are aberrations or physical signals.

One can straightforwardly leverage Bayesian hypothesis testing to constrain the maximum and minimum intensities a partition of x^{MAP} can take, such that the resulting surrogate x^{SUR} saturates the approximate level-set threshold ϵ'_α. In this sense one can recover local voxel level Bayesian error bars coined *local credible intervals* (Cai et al., 2018b; Price et al., 2019a; Repetti et al., 2019; Price et al., 2021c). The concept of Bayesian hypothesis testing can further be leveraged to consider hypothesis tests which quantify the uncertainty in *e.g.* feature location (Price et al., 2019b) and global features (Price et al., 2021b). These uncertainty quantification techniques are comprehensively considered in chapters 5 - 9.

10.3 Numerical experiment

In this section we consider a noisy inpainting directional deconvolution inverse problem, which is (seriously) ill-posed and ill-conditioned. Such an example is representative of a diverse set of practical applications. Consider again the problem setup outlined in Section 10.2.4, where we model the acquisition of observations by the sensing operator

$$\mathsf{\Phi} = \mathsf{M}\mathsf{B}^{-1}\mathsf{K}\mathsf{B} \quad \text{and} \quad \mathsf{\Phi}^{\dagger} = \mathsf{B}^{\dagger}\mathsf{K}\mathsf{B}^{-\dagger}\mathsf{M}^{\dagger}, \tag{10.13}$$

where B and B^{-1} represent forward and inverse spin-0 Spherical-Laguerre transforms (see Section 10.2.1), K is multiplication with a skewed Gaussian kernel in Spherical-Laguerre space (which is trivially self-adjoint), M represents masking, and † denotes the operator adjoint. It is important to note that $\mathsf{B}^{-1} \neq \mathsf{B}^{\dagger}$ which is a poorly motivated approximation often adopted in settings involving spherical harmonic transforms. Additionally, we define as a baseline the naive direct inversion $x^{\text{DIR}} = \mathsf{\Phi}^{-1}y$ for $\mathsf{\Phi}^{-1} = \mathsf{B}^{-1}\mathsf{K}^{-1}\mathsf{B}\mathsf{M}^{\dagger}$, where K^{-1} is simply division by the Spherical-Laguerre space convolutional kernel. As we are considering ill-posed inverse problems (Hadamard, 1902) the naive inverse x^{DIR} can give (potentially non-physical) solutions which lie far from the true signal. Moreover, the noise contribution, which is typically highly oscillatory, may (and often does) dominate the solution.

We consider the case in which n is independent and identically distributed noise drawn from a univariate Gaussian distribution $n \sim \mathcal{N}(0, \sigma^2)$. Our likelihood function is thus given by a Gaussian distribution with zero mean and variance σ^2. Suppose our prior knowledge indicates that x is likely to be sparsely distributed when projected into the ball wavelet dictionary $\mathsf{\Psi}$, described in Section 10.2.2. A prior distribution which naturally promotes sparsity is the Laplacian distribution, which one might adopt, such that the posterior is given by

$$P(x|y;\mathsf{\Phi}) \propto \exp\left(\frac{-_{\mathbb{B}^3}\|\mathsf{\Phi}x - y\|_2^2}{2\sigma^2}\right)\exp\left(-\lambda_{\Psi}\|\mathsf{\Psi}^{\dagger}x\|_1\right), \tag{10.14}$$

where $_{\mathbb{B}^3}\|\cdot\|$ and $_{\Psi}\|\cdot\|$ are the standard ℓ_p-norms weighted by pixel-size so as to better approximate the continuous ℓ_p-norms on the ball. By following the logic presented in Section 10.2.4 one finds the MAP estimate is given by

$$x^{\text{MAP}} = \underset{x \in \mathbb{R}^{N_{\mathbb{B}^3}}}{\text{argmin}}\left[\frac{_{\mathbb{B}^3}\|\mathsf{\Phi}x - y\|_2^2}{2\sigma^2} + \lambda_{\Psi}\|\mathsf{\Psi}^{\dagger}x\|_1\right], \tag{10.15}$$

with regularisation parameter $\lambda \in \mathbb{R}_{>0}$ which we marginalize over (Pereyra et al., 2015), to maintain a principled Bayesian interpretation.

10.3.1 Experiment details

We generate a ground truth signal x^{TRUE} by smoothing a random signal on the ball, effectively generating a pseudo-Gaussian random field, which is bandlimited at L in the angular domain and $P = L$ along the radial line. This ground truth is mapped by Φ to simulated observations which are subsequently polluted with *i.i.d.* noise, drawn from a univariate Gaussian distribution, to form simulated observations y, such that the input signal to noise ratio,

$$\text{SNR} = 20 \times \log_{10}\left(\frac{\|\Phi x^{\text{TRUE}}\|_2}{\|\Phi x^{\text{TRUE}} - y\|_2}\right), \tag{10.16}$$

is 30dB. An analogous SNR definition is used to quantify the reconstruction fidelity between x and a recovered solution x^*. Both the naive inversion (SNR$= -2.707$dB), and MAP (SNR$= 10.293$dB) estimators are recovered, and are presented in Figure 10.2. Note that the variational solution is recovered in the analysis unconstrained setting through the proximal forward-backward algorithm (Beck and Teboulle, 2009; Combettes and Pesquet, 2011). This dramatic improvement in reconstruction fidelity is compounded by the fact that our estimator also supports principled Bayesian uncertainty quantification, namely hypothesis testing of structure *e.g.* the diffuse, high intensity region Ω highlighted in Figure 10.2 was correctly determined to be physical at 99% confidence.

10.4 Summary

Whilst there are many methods which consider reconstruction over the 3D ball by analysing individual concatenated spherical shells, to the best of our knowledge, this is the first work which develops variational regularisation methods natively on the ball. Leveraging recent developments in probability concentration theory, we demonstrate how MAP estimation (unconstrained optimisation) permits principled uncertainty quantification. Our Bayesian variational approach benefits from the computational efficiency of convex optimisation whilst facilitating principled uncertainty quantification. We demonstrate that our variational approach is effective at solving seriously ill-posed and ill-conditioned inverse problems on the ball, recovering very accurate, robust estimates of the underlying ground truth. In future collaborative work we will apply these methods to more realistic simulations and observational data, in a variety of application domains. As a by-product of this work an open source, flexible, scalable object oriented C++ software package, B3INV[5] was created which is constructed on the convex optimisation package SOPT[6] (Carrillo et al., 2012, 2013; Onose et al., 2016; Pratley et al., 2018). Additionally, fast adjoint operators were constructed and collected into the FLAG and FLAGLET codebases.

---

[5] `https://github.com/astro-informatics/b3inv`

[6] `http://astro-informatics.github.io/sopt/`

This methodology lays the foundations upon which full 3D universal dark matter mass-mapping may now be considered, which is a substantial step between the first forays of Massey et al. (2007a,b) and the vision outlined by Fritz Zwicky nearly a century ago.

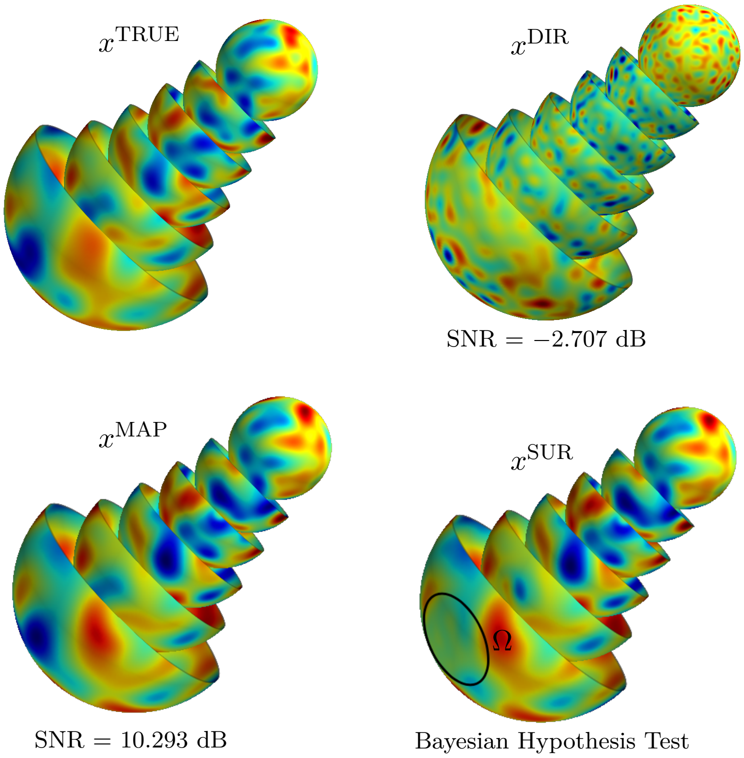


Figure 10.2: Description: Variational inference results for 3D directional deconvolution with inpainting using pseudo-Gaussian simulations for $L = P = 32$ (upsampled to 128), roughly approximating physical fields *e.g.* atmospheric fields – these methods trivially extend to realistic simulations. Specifically the ground truth is smoothed with a directional kernel, 50% of observations are masked, and the remaining observations are corrupted with 30dB *i.i.d.* Gaussian noise $n \sim \mathcal{N}(0, \sigma^2)$. **Panels:** Input ground truth (top left), naive inversion (baseline, top right), *maximum a posteriori* (MAP) estimator using ball wavelet ℓ_1 sparsity (bottom left), and Bayesian hypothesis test of local structure Ω (bottom right, see Section 10.2.5). **Discussion:** Notice that naive direct inversion recovers a poor estimator (SNR = −2.707dB). Alternatively, treating the problem as a Bayesian variational problem not only recovers a very good estimate (SNR = 10.293dB), but also supports principled uncertainty quantification. The Bayesian hypothesis test of local substructure Ω in the bottom right sub-figure correctly determines the physicality of this feature at 99%-confidence. The MAP estimate and uncertainties were recovered in ~ 2 minutes of non-dedicated compute on a MacBook Air 2016 respectively, which can trivially be sped up through *e.g.* OpenMP and MPI.

Chapter 11

Conclusions & future work

While a great deal of cosmological information is concentrated in 2^{nd}-order statistics of the cosmic shear, potentially decisive non-Gaussian information is encoded into higher order statistics of the convergence; hence, the lensing community's ongoing motivation to develop methods by which the convergence may be inferred. Existing methods are limited in a variety of ways, *e.g.* scalability, statistical interpretability, or degradation of non-Gaussian information, as chapter 4 discusses at length. Effectively, those methods with scientific utility, *e.g.* sampling methods, are computationally intractable at high-resolutions, or are fundamentally reliant on posterior Gaussianity. Moreover, for upcoming stage IV wide-field surveys the flat-sky approximation does not hold, therefore the lensing inverse problem must necessarily be considered over the celestial sphere ($\mathbb{S}^2$). All but a small subset of convergence inference methods are limited to Euclidean tangent planes ($\mathbb{R}^2$), limiting their scientific applicability going forward. The most general lensing inverse problem is, of course, most naturally defined over the 3D ball ($\mathbb{B}^3$), for which no algorithms currently exist.

Merging the paradigms of optimisation and Bayesian inference, this thesis addresses this methodological void by developing computationally efficient mass-mapping techniques which support: uncertainty quantification, are not reliant on Gaussianity, operate over $\{\mathbb{R}^2, \mathbb{S}^2, \mathbb{B}^3\}$, and can support an expansive set of posterior models. Exploiting recent advances in the theory of probability concentration, convex optimisation, and proximal calculus we develop next-generation Bayesian mass-mapping techniques; over $\mathbb{R}^2$ in chapter 5, $\mathbb{S}^2$ in chapter 9, and pave the way for mass-mapping over $\mathbb{B}^3$ in chapter 10. The super-resolution planar techniques are applied the hotly debated Abel-520 merging cluster, quantifying the degree to which purported evidence of dark-matter self-interactions is plausible. Initial reconstructions of two Abel-520 observational catalogues reported high σ detections of an anomalous dark core[1], however in our Bayesian framework neither data-set contain sufficient information to determine the physicality of this dark core at 99% confidence. Additionally, in chapter 9 our spherical reconstruction algorithm is

---

[1] In this context, a dark core is an over-dense region of the convergence with no apparent optical counterpart, often taken to indicate the existence of a frictional force or 'self-interaction' between dark matter, by which the dark matter component of a cluster is stripped from its optical counterpart.

applied to all public weak lensing observational data, generating the largest spherical map of dark matter to date[2].

Scalability of such techniques stems from a conservative approximation of posterior credible regions, derived from the concentration properties of log-concave densities in high-dimensions (see section 3.2.3). The numerical error of such approximations is explored in chapter 6, and found to be of $\mathcal{O}(10\%)$ which is traded for a factor of $\mathcal{O}(10^6)$ gain in computational efficiency. In chapter 7 the notion of approximate posterior credible regions is leveraged to quantify the uncertainty in both the location and count of convergence peaks (over-dense regions). Such uncertainty quantification techniques are directly applicable to a variety of tangential domains, *e.g.* location of features such as tumours in magnetic resonance images. In chapter 10 we abstract the aforementioned inference techniques to the full 3D ball, setting the scene for principled 3D dark matter reconstruction in the near future.

In summary, the primary original contributions of this thesis are the following:

- ▷ Development of highly scalable, statistically principled Bayesian techniques for imaging of dark matter, amongst other applications.
- ▷ Development of a variety of novel Bayesian uncertainty quantification techniques supported by such imaging approaches.
- ▷ Quantifying the plausibility of evidence suggesting self-interactions of dark matter in the Abel-520 merging cluster.
- ▷ Abstracting these techniques from $\mathbb{R}^2$ to the celestial sphere $\mathbb{S}^2$, a necessity for scientific analysis of imminent next-generation weak lensing surveys.
- ▷ Reconstructing, what was at the time, the most comprehensive map of the celestial dark matter distribution to date.
- ▷ Developing techniques which, for the first time, facilitate efficient variational Bayesian inference on $\mathbb{B}^3$.

Despite the substantial progress made in this thesis, the lensing inversion problem is far from solved. Though the computational foundations for 3D mass-mapping have been laid, it is as yet unclear how best to incorporate radial (photometric redshift) information into the statistical model. Perhaps more fundamentally, operations over $\mathbb{B}^3$ scale quartically with resolution, motivating further research into computational efficiency. Throughout this thesis we do not consider the problem of prior misspecification. Techniques developed during this thesis are reliant only on

---

[2] This has since been overtaken by the subsequent introduction of new DES data.

posterior log-concavity, and thus objective (negative log-posterior) convexity. These two realizations motivate potential future research into *e.g.* convex machine learning priors, which can more accurately model one's *a priori* knowledge of a given system.

The cosmological phenomenon of gravitational lensing is uniquely sensitive to the total mass distribution throughout the universe, both visible and invisible, and is thus a novel probe as to the nature of both dark matter and energy. Of the two linear order manifestations of gravitational lensing, the convergence field constitutes one of the principle cosmological observables, containing information which directly constrains *e.g.* modified theories of gravity, exotic theoretical physics, and neutrino hierarchies. With next-generation stage IV weak lensing surveys now imminent, techniques (such as those presented in this thesis) which extract such information in an efficient, statistically principled manner are of critical importance.

Appendices

Appendix A

Circular bisection details

A.1 N-splitting circular bisection

In this appendix we consider the N-splitting Circular Bisection (N-splitting) algorithm for iteratively sampling the Bayesian $100(1-\alpha)\%$ confidence iso-contour of the position of a feature in a reconstructed convergence map — or the *Bayesian Location* at $100(1-\alpha)\%$ confidence.

As in the text, we begin by defining the number of directions to sample n_T from which we then form the angular increment $\Delta\Theta = 2\pi/n_T$. Starting from a vector n_0 oriented along the positive y-axis define the $(i+1)^{\text{th}}$ pointing to be the vector $\mathsf{n}_{i+1} = \mathcal{R}_{\Delta\Theta}\mathsf{n}_i$ where $i \in (1, n_T)$, and where $\mathcal{R}_{\Delta\Theta}$ is rotation by angle $\Delta\Theta$ clockwise on 2D Euclidean space — a irreducible representation of which is the standard clockwise rotation matrix,

$$\mathcal{R}_\theta = \begin{bmatrix} \cos(\Delta\Theta) & \sin(\Delta\Theta) \\ -\sin(\Delta\Theta) & \cos(\Delta\Theta) \end{bmatrix}. \tag{A.1}$$

Now we know the direction along which we wish to sample we construct the $(i+1)^{\text{th}}$ bisection problem which is

$$d_\alpha^{i+1} = \min_d \Big\{ \, d \in \Gamma_{i+1} \mid f(\kappa_d^{\text{sgt}}) + g(\kappa_d^{\text{sgt}}) > \epsilon'_\alpha \, \Big\}, \tag{A.2}$$

where κ_d^{sgt} is a surrogate convergence map with the feature of interest inserted into perturbed location $d\mathsf{n}_{i+1}$ and $\Gamma_{i+1} \subset \mathbb{R}^2$ lies on the directional line of infinite extent centered at the original peak location with unit vector n_{i+1}, *i.e.* $\Gamma_{i+1} = \{a\mathsf{n}_{i+1} | a \in \mathbb{R}_{>0}\}$. A pictorial representation of how the problem is set up is provided in Figure A.1.

For bisection we must first make an initial guess d_0 which we define to be square root of the number of pixels contained within the mask, as this is a typical measure of the length of a masked region. This choice is particularly logical as, if a feature of interest can be removed entirely from its masked location without saturating the level-set threshold ϵ'_α then it by definition must be inconclusive, *i.e.* the data is insufficient evidence to say that the peak is physical. To optimize the convergence of this algorithm further (for high sampling rates, low angular increments $\Delta\Theta \leq \pi/4$)

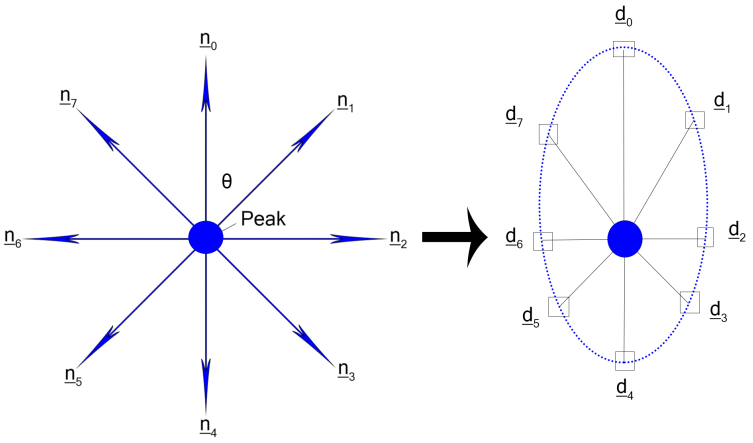


Figure A.1: Representation of how the problem is broken up in N-splitting circular bisection. First the n_i directions are specified (*left*) at equiangular separations θ about the peak location (*blue ball*). Bisection iterations are conducted as in equation (A.2) along each of the directions, recovering a set of samples d_i of the Bayesian location iso-contour at $100(1-\alpha)\%$ confidence (*right*). Provided a sufficient number of samples are taken, this boundary will fully represent the iso-contour. We find typically ≈ 16 samples are needed for 512×512 convergence reconstructions though more or less may be needed depending on the resolution and application.

we also propagate information between pointing's. For bisection problems associated with pointing $i > 1$ the initial guess is now set to be twice the previous optimal value $d_\alpha^{\prime,i}$. This increases the computational efficiency by $\sim 20\%$, in most cases.

Propagating information in this way relies on the assumption that the iso-contour we are searching for is somewhat smooth and continuous, which is the case for most convergence reconstructions. If there is uncertainty as to the smoothness of the iso-contour it is recommended that information is not propagated and the number of pointings is increased to correctly map the iso-contour structure.

A.1.1 Convergence properties

Standard inverse nesting algorithms iteratively sample the entire sub-space of the reconstructed domain bounded by the iso-contour at $100(1-\alpha)\%$ confidence, making them inefficient when one is only interested in the boundary.

Consider the case where the iso-contour of a reconstructed 512×512 convergence map is a circular region of radius R. Here inverse nesting will have to sample a square region out to R, which is to say the total number of samples T_{nest} will at least be $R^2 - 1$, where 1 is removed for the central location. For our N-splitting algorithm we define n_T pointings, and assume that the iso-contour is relatively smooth. As the first bisection problem n_0 makes a large first guess it typically takes $4-5$ iterations to converge with a single pixel accuracy. The subsequent $n_T - 1$ bisection problems converge within $3-4$ iterations. Therefore the total number of calculations $T_{\text{N-split}}$ is conservatively $T_{\text{N-split}} = 5 + 4(n_T - 1)$, which is essentially independent from R.

There is in fact a small inverse dependence which is incorporated in the number of iterations needed for convergence, though this dependence is found to be small.

Comparing the computational efficiency of the two algorithms $\mathbb{E}_{512}$ where,

$$\mathbb{E}_{512} \equiv \frac{T_{\text{N-split}}}{T_{\text{nest}}} = \frac{5+4(n_T-1)}{R^2}. \tag{A.3}$$

Typically, we find an angular separation between pointings of $\pi/4$ (*i.e.* 16 pointings) is sufficient to accurately recover the iso-contour. Additionally, the circular radius is typically $15-30$ pixels which indicates that,

$$\frac{5+4\times 15}{30^2} = 0.072 \leq \mathbb{E}_{512} \leq \frac{5+4\times 15}{15^2} = 0.289, \tag{A.4}$$

i.e. N-splitting circular bisection on 512×512 dimensional reconstructions is $\sim 4-14$ times faster than inverse nesting. However, in the future we may be interested in recovering high dimensional 2048×2048 convergence maps. In this setting the number of iterations for N-splitting to converge is assumed to change by 1-2, and the number of pointings to faithfully recover the iso-contour will be increase by a factor of ~ 2. Additionally, the radius of the circle R increases by a factor of 4. Thus,

$$\frac{5+4\times 31}{120^2} = 0.009 \leq \mathbb{E}_{2048} \leq \frac{5+4\times 31}{60^2} = 0.0360, \tag{A.5}$$

i.e. the conservative increase in computational efficiency of N-splitting over inverse nesting for 2048×2048 becomes a factor of $\approx 30-112$. Further optimisations are possible, such as trivially parallelizing the bisection problems of each pointing. Doing so removes the scaling with the number of pointings, but now information about starting positions cannot be propagated.

Appendix B

Public lensing catalogues

B.1 Data availability

All observational data utilized throughout chapter 9 are 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`.

Dark Energy Survey (DES) acknowledgements

This project used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology FacilitiesCouncil of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

Kilo Degree Survey (KiDS) acknowledgements

Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017 and 177.A-3018.

We use cosmic shear measurements from the Kilo-Degree Survey (Kuijken et al., 2015; Fenech Conti et al., 2017; Hildebrandt et al., 2017), hereafter referred to as KiDS. The KiDS data are processed by THELI (Erben et al., 2013) and Astro-WISE (McFarland et al., 2013; de Jong et al., 2015). Shears are measured using lensfit (Miller et al., 2013), and photometric redshifts are obtained from PSF-matched photometry and calibrated using external overlapping spectroscopic surveys (see *e.g.* Hildebrandt et al., 2017).

Bibliography

T. M. C. Abbott et al. The Dark Energy Survey: Data Release 1. *Astrophysical Journal, Supplement*, 239(2):18, Dec 2018. doi: 10.3847/1538-4365/aae9f0.

P. Abrial et al. Morphological component analysis and inpainting on the sphere: Application in physics and astrophysics. *Journal of Fourier Analysis and Applications*, 13:729–748, 12 2007. doi: 10.1007/s00041-006-6908-x.

Jonas Adler and Ozan Öktem. Deep Bayesian Inversion. *arXiv e-prints*, art. arXiv:1811.05910, November 2018.

Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE transactions on medical imaging*, 37(6):1322–1332, 2018.

Nabila Aghanim, Yashar Akrami, Frederico Arroja, Mark Ashdown, J Aumont, Carlo Baccigalupi, M Ballardini, Anthony J Banday, RB Barreiro, Nicola Bartolo, et al. Planck 2018 results-i. overview and the cosmological legacy of planck. *Astronomy & Astrophysics*, 641:A1, 2020.

Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer, 1998.

J. Alsing, A. Heavens, A. H. Jaffe, A. Kiessling, B. Wandelt, and T. Hoffmann. Hierarchical cosmic shear power spectrum inference. *Monthly Notices of the Royal Astronomical Society*, 455:4452–4466, February 2016. doi: 10.1093/mnras/stv2501.

Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440–4458, 2019.

D. J. Bacon, A. R. Refregier, and R. S. Ellis. Detection of weak gravitational lensing by large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 318: 625–640, October 2000. doi: 10.1046/j.1365-8711.2000.03851.x.

Paolo Baldi, Gérard Kerkyacharian, Domenico Marinucci, Dominique Picard, et al. Asymptotics for spherical needlets. *The Annals of Statistics*, 37(3):1150–1171, 2009.

James M. Bardeen. Gauge-invariant cosmological perturbations. *Phys. Rev. D*, 22: 1882–1905, Oct 1980. doi: 10.1103/PhysRevD.22.1882. URL `https://link.aps.org/doi/10.1103/PhysRevD.22.1882`.

R. B. Barreiro, M. P. Hobson, A. N. Lasenby, A. J. Banday, K. M. Górski, and G. Hinshaw. Testing the Gaussianity of the COBE-DMR data with spherical wavelets. *MNRAS*, 318:475–481, 2000.

M. Bartelmann. TOPICAL REVIEW Gravitational lensing. *Classical and Quantum Gravity*, 27(23):233001, December 2010. doi: 10.1088/0264-9381/27/23/233001.

M. Bartelmann and P. Schneider. Weak gravitational lensing. *Physics Reports*, 340: 291–472, January 2001. doi: 10.1016/S0370-1573(00)00082-X.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

Francis Bernardeau, Camille Bonvin, and Filippo Vernizzi. Full-sky lensing shear at second order. *Physical Review D*, 81(8):083002, 2010.

Jose M Bernardo. Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):113–128, 1979.

Sergey Bobkov and Mokshay Madiman. Concentration of the information in data with log-concave distributions. *The Annals of Probability*, pages 1528–1543, 2011a.

Sergey Bobkov and Mokshay Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Transactions on Information Theory*, 57(8):4940–4954, 2011b.

Wouter Boomsma and Jes Frellsen. Spherical convolutions and their application in molecular modelling. In *Advances in Neural Information Processing Systems*, volume 2, page 6, 2017.

S. Boyd. Convex Optimization. 2004.

Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers.* Now Publishers Inc, 2011.

Willard S Boyle and George E Smith. Charge coupled semiconductor devices. *Bell System Technical Journal*, 49(4):587–593, 1970.

Maruša Bradač, Marco Lombardi, and Peter Schneider. Mass-sheet degeneracy: Fundamental limit on the cluster mass reconstruction from statistical (weak) lensing. *Astronomy & Astrophysics*, 424(1):13–22, 2004.

T. G. Brainerd, R. D. Blandford, and I. Smail. Weak Gravitational Lensing by Galaxies. *apj*, 466:623, August 1996. doi: 10.1086/177537.

E. F. Bunn, M. Zaldarriaga, M. Tegmark, and A. de Oliveira-Costa. E/B decomposition of finite pixelized CMB maps. *prd*, 67(2):023501, January 2003. doi: 10.1103/PhysRevD.67.023501.

Xiaohao Cai, Marcelo Pereyra, and Jason D McEwen. Uncertainty quantification for radio interferometric imaging–i. proximal mcmc methods. *Monthly Notices of the Royal Astronomical Society*, 480(3):4154–4169, 2018a.

Xiaohao Cai, Marcelo Pereyra, and Jason D McEwen. Uncertainty quantification for radio interferometric imaging: Ii. map estimation. *Monthly Notices of the Royal Astronomical Society*, 480(3):4170–4182, 2018b.

Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452, 2006.

Emmanuel J Candès et al. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

R. E. Carrillo, J. D. McEwen, and Y. Wiaux. Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 426(2):1223–1234, October 2012. doi: 10.1111/j.1365-2966.2012.21605.x.

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*, 20:591–594, June 2013. doi: 10.1109/LSP.2013.2259813.

George Casella. An introduction to empirical bayes data analysis. *The American Statistician*, 39(2):83–87, 1985.

A. Challinor. Lecture Notes on Advanced Cosmology. December 2009.

A. Challinor and H. Peiris. Lecture notes on the physics of cosmic microwave background anisotropies. In M. Novello and S. Perez, editors, *American Institute of Physics Conference Series*, volume 1132 of *American Institute of Physics Conference Series*, pages 86–140, May 2009. doi: 10.1063/1.3151849.

J. Y. H. Chan, B. Leistedt, T. D. Kitching, and J. D. McEwen. Second-Generation Curvelets on the Sphere. *IEEE Transactions on Signal Processing*, 65:5–14, January 2017. doi: 10.1109/TSP.2016.2600506.

C Chang, M Jarvis, B Jain, SM Kahn, D Kirkby, A Connolly, S Krughoff, E-H Peng, and JR Peterson. The effective number density of galaxies for weak lensing measurements in the lsst project. *Monthly Notices of the Royal Astronomical Society*, 434(3):2121–2135, 2013.

C. Chang et al. Dark Energy Survey Year 1 results: curved-sky weak lensing mass map. *Monthly Notices of the Royal Astronomical Society*, 475:3165–3190, April 2018. doi: 10.1093/mnras/stx3363.

Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335, 1995.

Nicolae Cleju, Maria G Jafari, and Mark D Plumbley. Choosing analysis or synthesis recovery for sparse reconstruction. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 869–873. IEEE, 2012.

Douglas Clowe, Maruša Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, aug 2006. doi: 10.1086/508162. URL `https://doi.org/10.1086/508162`.

Douglas Clowe, Maxim Markevitch, M. Bradač, Anthony H. Gonzalez, Sun Mi Chung, Richard Massey, and Dennis Zaritsky. On dark peaks and missing mass: A weak-lensing mass reconstruction of the merging cluster system a520. *The Astrophysical Journal*, 758(2):128, 2012. URL `http://stacks.iop.org/0004-637X/758/i=2/a=128`.

P. Coles and L. Y. Chiang. Characterizing the nonlinear growth of large-scale structure in the Universe. *Nature*, 406:376–378, July 2000. doi: 10.1038/35019009.

P. L. Combettes and Valérie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4:1168–1200, 2005.

Patrick L Combettes and Jean-Christophe Pesquet. *Proximal splitting methods in signal processing*. Springer, 2011.

Patrick L Combettes, Laurent Condat, J-C Pesquet, and BC Vũ. A forward-backward view of some primal-dual optimization methods in image recovery. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4141–4145. IEEE, 2014.

Edward Uhler Condon, EU Condon, and GH Shortley. *The theory of atomic spectra*. Cambridge University Press, 1951.

D. Copeland, A. Taylor, and A. Hall. The impact of baryons on the sensitivity of dark energy measurements. *Monthly Notices of the Royal Astronomical Society*, 480:2247–2265, October 2018. doi: 10.1093/mnras/sty2001.

V. L. Corless, L. J. King, and D. Clowe. A new look at massive clusters: weak lensing constraints on the triaxial dark matter haloes of a1689, a1835 and a2204.

Monthly Notices of the Royal Astronomical Society, 393:1235–1254, March 2009. doi: 10.1111/j.1365-2966.2008.14294.x.

R. T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946. doi: 10.1119/1.1990764. URL `https://doi.org/10.1119/1.1990764`.

I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004. doi: https://doi.org/10.1002/cpa.20042. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20042`.

Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988.

Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992.

Jelte T. A. de Jong et al. The first and second data releases of the Kilo-Degree Survey. *Astronomy and Astrophysics*, 582:A62, Oct 2015. doi: 10.1051/0004-6361/201526601.

Joseph DeRose, Risa H Wechsler, Matthew R Becker, Michael T Busha, Eli S Rykoff, Niall MacCrann, Brandon Erickson, August E Evrard, Andrey Kravtsov, Daniel Gruen, et al. The buzzard flock: Dark energy survey synthetic sky catalogs. *arXiv preprint arXiv:1901.02401*, 2019.

Persi Diaconis and Donald Ylvisaker. Conjugate priors for exponential families. *The Annals of statistics*, pages 269–281, 1979.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

David L Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 58(2):1094–1121, 2012.

C. Durastanti, Y. Fantaye, F. Hansen, D. Marinucci, and I. Z. Pesenson. Simple proposal for radial 3d needlets. *Phys. Rev. D*, 90:103532, Nov 2014.

Alain Durmus, Eric Moulines, and Marcelo Pereyra. Efficient bayesian computation by proximal markov chain monte carlo: when langevin meets moreau. *SIAM Journal on Imaging Sciences*, 11(1):473–506, 2018.

F. W. Dyson. *The scientific papers of the Honourable Henry Cavendish*. 1921.

F. W. Dyson, A. S. Eddingtion, and C. Davidson. Die Grundlage der allgemeinen Relativitätstheorie. *Phil. Trans. Roy. Soc.*, 220:291, 1920.

A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17, 1905a.

A. Einstein. Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig? *Annalen der Physik*, 18, 1905b.

A. Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffended heuristischen Gesichtspunkt. *Annalen der Physik*, 17, 1905c.

A. Einstein. zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 17, 1905d.

A. Einstein. Über den EinfluÃ§ der Schwerkraft auf die Ausbreitung des Lichtes. *Annalen der Physik*, 35:898–908, 1911.

A. Einstein. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 49:769–822, 1916.

A. Einstein. Lens-like action of a star by the deviation of light in the gravitational field. *Science*, 1936.

M. Elad et al. Analysis versus synthesis in signal priors. In *2006 14th European Signal Processing Conference*, pages 1–5, 2006.

Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer Publishing Company, Incorporated, 1st edition, 2010. ISBN 144197010X, 9781441970107.

Elsner, F. and Wandelt, B. D. Efficient wiener filtering without preconditioning. *A&A*, 549:A111, 2013. doi: 10.1051/0004-6361/201220586. URL `https://doi.org/10.1051/0004-6361/201220586`.

T. Erben et al. Cfhtlens: The canada-france-hawaii telescope lensing survey - imaging data and catalogue products. *Monthly Notices of the Royal Astronomical Society*, 433, 10 2012. doi: 10.1093/mnras/stt928.

T. Erben et al. CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey - imaging data and catalogue products. *Monthly Notices of the Royal Astronomical Society*, 433(3):2545–2563, Aug 2013. doi: 10.1093/mnras/stt928.

Samuel Farrens. Sparsity tutorial. 2018.

Stephen M. Feeney et al. Sparse inpainting and isotropy. *Journal of Cosmology and Astroparticle Physics*, 2014(1):050, January 2014. doi: 10.1088/1475-7516/2014/01/050.

I. Fenech Conti, R. Herbonnet, H. Hoekstra, J. Merten, L. Miller, and M. Viola. Calibration of weak-lensing shear in the Kilo-Degree Survey. *Monthly Notices of the Royal Astronomical Society*, 467(2):1627–1651, May 2017. doi: 10.1093/mnras/stx200.

M.J. Fengler, D. Michel, and V. Michel. *ZAMM - Journal of Applied Mathematics and Mechanics*, 86(11):856–873, 2006. doi: https://doi.org/10.1002/zamm.200510277.

Daniel Fink. A compendium of conjugate priors. *See http://www. people. cornell. edu/pages/df36/CONJINTRnew% 20TEX. pdf*, 46, 1997.

B. Flaugher and DES Collaboration. The Dark Energy Camera. *Astronomical Journal*, 150(5):150, Nov 2015. doi: 10.1088/0004-6256/150/5/150.

Janis Fluri, Tomasz Kacprzak, Raphael Sgier, Alexandre Refregier, and Adam Amara. Weak lensing peak statistics in the era of large scale cosmological surveys. *Journal of Cosmology and Astroparticle Physics*, 2018(10):051, 2018.

A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, 1922. doi: 10.1007/BF01332580.

B Scott Gaudi. Microlensing surveys for exoplanets. *Annual Review of Astronomy and Astrophysics*, 50:411–453, 2012.

Andrew Gelman and Xiao-Li Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185, 1998.

Malay Ghosh. Objective priors: An introduction for frequentists. *Statistical Science*, 26(2):187–202, 2011.

Apostolos A Giannopoulos and Vitali D Milman. Concentration property on probability spaces. *Advances in Mathematics*, 156(1):77–106, 2000.

Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.

I. Gkioulekas. Sparsity Based Regularization. mar 2010.

J. N. Goldberg, A. J. Macfarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan. Spin-s spherical harmonics. *Journal of Mathematical Physics*, 8(11):2155–2161, 1967. doi: 10.1063/1.1705135. URL `https://doi.org/10.1063/1.1705135`.

I.J. Good. The algebra of probable inference. *The Mathematical Gazette*, 46(357): 246–247, 1962. doi: 10.2307/3614046.

MV Gorenstein, EE Falco, and II Shapiro. Degeneracies in parameter estimates for models of gravitational lens systems. *The Astrophysical Journal*, 327:693–711, 1988.

Krzysztof M Gorski, Eric Hivon, AJ Banday, Benjamin D Wandelt, Frode K Hansen, Mstvos Reinecke, and Matthia Bartelmann. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal*, 622(2):759, 2005.

Amara Graps. An introduction to wavelets. *IEEE computational science and engineering*, 2(2):50–61, 1995.

D. A. Green. A colour scheme for the display of astronomical intensity images. *Bulletin of the Astronomical Society of India*, 39:289–295, Jun 2011.

Peter J Green, Krzysztof Łatuszyński, Marcelo Pereyra, and Christian P Robert. Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25(4):835–862, 2015.

N. Grimm and J. Yoo. Jacobi Mapping Approach for a Precise Cosmological Weak Lensing Formalism. *ArXiv e-prints*, May 2018.

Michael U Gutmann, Jukka Corander, et al. Bayesian optimization for likelihood-free inference of simulator-based statistical models. *Journal of Machine Learning Research*, 2016.

Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52, 1902.

JOHN Hartigan. Invariant prior distributions. *The Annals of Mathematical Statistics*, pages 836–845, 1964.

W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

A. Heavens. 3D weak lensing. *Monthly Notices of the Royal Astronomical Society*, 343:1327–1334, August 2003. doi: 10.1046/j.1365-8711.2003.06780.x.

A. Heavens. Weak lensing: Dark Matter, Dark Energy and Dark Gravity. *Nuclear Physics B Proceedings Supplements*, 194:76–81, October 2009. doi: 10.1016/j.nuclphysbps.2009.07.005.

H. Hildebrandt et al. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 465(2):1454–1498, Feb 2017. doi: 10.1093/mnras/stw2805.

AD Hiller and Roland T Chin. Iterative wiener filters for image restoration. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 1901–1904. IEEE, 1990.

G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. *Astrophysical Journal, Supplement*, 208(2):19, Oct 2013. doi: 10.1088/0067-0049/208/2/19.

Michael Paul Hobson, George P Efstathiou, and Anthony N Lasenby. *General relativity: an introduction for physicists*. Cambridge University Press, 2006.

Benjamin Horowitz, Uros Seljak, and Grigor Aslanyan. Efficient optimal reconstruction of linear fields and band-powers from cosmological data. *Journal of Cosmology and Astroparticle Physics*, 2019(10):035, 2019.

Wayne Hu. Weak lensing of the cmb: A harmonic approach. *Phys. Rev. D*, 62: 043007, Jul 2000. doi: 10.1103/PhysRevD.62.043007. URL `https://link.aps.org/doi/10.1103/PhysRevD.62.043007`.

Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929. ISSN 0027-8424. doi: 10.1073/pnas.15.3.168. URL `https://www.pnas.org/content/15/3/168`.

MJ Irwin, RL Webster, PC Hewett, RT Corrigan, and RI Jedrzejewski. Photometric variations in the q2237+ 0305 system-first detection of a microlensing event. *The Astronomical Journal*, 98, 1989.

B. Jain and L. Van Waerbeke. Statistics of Dark Matter Halos from Gravitational Lensing. *The Astrophysical Journal Letters*, 530:L1–L4, February 2000. doi: 10.1086/312480.

Edwin T Jaynes. Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241, 1968.

Edwin T Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.

M. J. Jee, A. Mahdavi, H. Hoekstra, A. Babul, J. J. Dalcanton, P. Carroll, and P. Capak. A Study of the Dark Core in A520 with the Hubble Space Telescope: The Mystery Deepens. *The Astrophysical Journal*, 747:96, March 2012. doi: 10.1088/0004-637X/747/2/96.

M. J. Jee, H. Hoekstra, A. Mahdavi, and A. Babul. Hubble Space Telescope/Advanced Camera for Surveys Confirmation of the Dark Substructure in A520. *The Astrophysical Journal*, 783:78, March 2014. doi: 10.1088/0004-637X/783/2/78.

M. James Jee, William A. Dawson, Andra Stroe, David Wittman, Reinout J. van Weeren, Marcus Braggen, MaruaÂąa Brada, and Huub Rattgering. Mc2: Mapping the dark matter distribution of the "toothbrush" cluster rx j0603.3+4214 with hubble space telescope and subaru weak lensing. *The Astrophysical Journal*, 817 (2):179, 2016. URL http://stacks.iop.org/0004-637X/817/i=2/a=179.

William H Jefferys and James O Berger. Ockham's razor and bayesian analysis. *American Scientist*, 80(1):64–72, 1992.

N. Jeffrey et al. Improving Weak Lensing Mass Map Reconstructions using Gaussian and Sparsity Priors: Application to DES SV. *ArXiv e-prints*, January 2018.

Niall Jeffrey, François Lanusse, Ofer Lahav, and Jean-Luc Starck. Deep learning dark matter map reconstructions from DES SV weak lensing data. *Monthly Notices of the Royal Astronomical Society*, 492(4):5023–5029, 01 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa127. URL https://doi.org/10.1093/mnras/staa127.

Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30:175 – 193, 1906. doi: 10.1007/BF02418571. URL https://doi.org/10.1007/BF02418571.

E. Jullo, S. Pires, M. Jauzac, and J.-P. Kneib. Weak lensing galaxy cluster field reconstruction. *Monthly Notices of the Royal Astronomical Society*, 437:3969, February 2014.

Sergey Kabanikhin. Definitions and examples of inverse and ill-posed problems. 16: 317–357, 07 2008.

Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

N. Kaiser and G. Squires. Mapping the dark matter with weak gravitational lensing. *apj*, 404:441–450, February 1993. doi: 10.1086/172297.

Robert E Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American statistical Association*, 91(435):1343–1370, 1996.

Lambros S Katafygiotis and Konstantin M Zuev. Geometric insight into the challenges of solving high-dimensional reliability problems. *Probabilistic Engineering Mechanics*, 23(2-3):208–218, 2008.

I. Kayo, M. Takada, and B. Jain. Information content of weak lensing power spectrum and bispectrum: including the non-Gaussian error covariance matrix. *Monthly Notices of the Royal Astronomical Society*, 429:344–371, February 2013. doi: 10.1093/mnras/sts340.

Jens Keiner, Stefan Kunis, and Daniel Potts. Using nfft 3—a software library for various nonequispaced fast fourier transforms. *ACM Trans. Math. Softw.*, 36(4): 19:1–19:30, August 2009. ISSN 0098-3500. doi: 10.1145/1555386.1555388. URL `http://doi.acm.org/10.1145/1555386.1555388`.

E. Kendall, A.M.G. Ferreira, S.-J. Chang, M. Witek, and D. Peter. Constraints on the upper mantle structure beneath the pacific from 3-d anisotropic waveform modelling. *Journal of Geophysical Research: Solid Earth*, 2021. doi: https://doi.org/10.1029/2020JB020003.

John Maynard Keynes. *A Treatise on Probability*. Dover Publications, 1921.

Zubair Khalid, Rodney A. Kennedy, and Jason D. McEwen. Slepian spatial-spectral concentration on the ball. *Applied and Computational Harmonic Analysis*, 40(3): 470–504, 2016. ISSN 1063-5203.

O. Khvolson. *Astronomischen Nachrichten*, 221:329–330, 1924.

Martin Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, 2015.

T. D. Kitching, J. Alsing, A. F. Heavens, R. Jimenez, J. D. McEwen, and L. Verde. The limits of cosmic shear. *Monthly Notices of the Royal Astronomical Society*, 469:2737–2749, August 2017. doi: 10.1093/mnras/stx1039.

Y. G.. Klimov. *Soviet Phys.*, 8:119, 1963.

A. A. Klypin, S. Trujillo-Gomez, and J. Primack. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. *The Astrophysical Journal*, 740:102, October 2011. doi: 10.1088/0004-637X/740/2/102.

Christopher S Kochanek. The implications of lenses for galaxy structure. *The Astrophysical Journal*, 373:354–368, 1991.

CS Kochanek, CR Keeton, and BA McLeod. The importance of einstein rings, 2001.

Doogesh Kodi Ramanah, Guilhem Lavaux, and Benjamin D Wandelt. Wiener filtering and pure E/B decomposition of CMB maps with anisotropic correlated noise. *Monthly Notices of the Royal Astronomical Society*, 490(1):947–961, 09 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz2608. URL `https://doi.org/10.1093/mnras/stz2608`.

Nikos Komodakis and Jean-Christophe Pesquet. Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, 32(6):31–54, 2015. doi: 10.1109/MSP.2014.2377273.

Léon VE Koopmans and Tommaso Treu. The structure and dynamics of luminous and dark matter in the early-type lens galaxy of 0047–281 at z= 0.485. *The Astrophysical Journal*, 583(2):606, 2003.

H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, Berkeley, Calif., 1951. University of California Press. URL `https://projecteuclid.org/euclid.bsmsp/1200500249`.

Konrad Kuijken et al. Gravitational lensing analysis of the Kilo-Degree Survey. *Monthly Notices of the Royal Astronomical Society*, 454(4):3500–3532, Dec 2015. doi: 10.1093/mnras/stv2140.

Solomon Kullback. *Information theory and statistics*. Courier Corporation, 1997.

F. Lanusse, A. Rassat, and J. L. Starck. Spherical 3D isotropic wavelets. *Journal of Astronomy & Astrophysics*, 540:A92, April 2012. doi: 10.1051/0004-6361/201118568.

F. Lanusse, J.-L. Starck, A. Leonard, and S. Pires. High resolution weak lensing mass mapping combining shear and flexion. *Astronomy and Astrophysics*, 591:A2, June 2016. doi: 10.1051/0004-6361/201628278.

Pierre Simon Laplace. Memoir on the Probability of the Causes of Events. *Statistical Science*, 1(3):364 – 378, 1986. doi: 10.1214/ss/1177013621. URL `https://doi.org/10.1214/ss/1177013621`.

Rémi Laumont, Valentin de Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra. Bayesian imaging using Plug & Play priors: when Langevin meets Tweedie. *arXiv e-prints*, art. arXiv:2103.04715, March 2021.

R. Laureijs et al. Euclid Definition Study Report. *arXiv e-prints*, art. arXiv:1110.3193, October 2011.

Daniel TL Lee and Akio Yamamoto. Wavelet analysis: theory and applications. *Hewlett Packard journal*, 45:44–44, 1994.

B. Leistedt, J. D. McEwen, P. Vandergheynst, and Y. Wiaux. S2LET: A code to perform fast wavelet analysis on the sphere. *Astronomy and Astrophysics*, 558: A128, October 2013. doi: 10.1051/0004-6361/201220729.

Boris Leistedt and Jason D. McEwen. Exact Wavelets on the Ball. *IEEE Transactions on Signal Processing*, 60(12):6257–6269, December 2012. doi: 10.1109/TSP.2012.2215030.

Boris Leistedt, Jason D. McEwen, Thomas D. Kitching, and Hiranya V. Peiris. 3D weak lensing with spin wavelets on the ball. *Physical Review D*, 92(12):123010, December 2015. doi: 10.1103/PhysRevD.92.123010.

G. Lemaître. L'Univers en expansion. *Annales de la Société Scientifique de Bruxelles*, 53, 1933.

Nathan P Lemoine. Moving beyond noninformative priors: why and how to choose weakly informative priors in bayesian analyses. *Oikos*, 128(7):912–928, 2019.

Adrienne Leonard, François Lanusse, and Jean-Luc Starck. Glimpse: accurate 3d weak lensing reconstructions using sparsity. *Monthly Notices of the Royal Astronomical Society*, 440(2):1281–1294, 2014.

Sidney Liebes Jr. Gravitational lenses. *Physical Review*, 133(3B):B835, 1964.

C.-A. Lin. Cosmology with weak-lensing peak counts. *ArXiv e-prints*, December 2016.

Dennis Victor Lindley. *Bayesian statistics: A review*. SIAM, 1972.

O. Lodge. *Nature*, 104:354, 1919.

Sebastian Lunz, Ozan Öktem, and Carola-Bibiane Schönlieb. Adversarial regularizers in inverse problems. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL `https://proceedings.neurips.cc/paper/2018/file/d903e9608cfbf08910611e4346a0ba44-Paper.pdf`.

C.-P. Ma and E. Bertschinger. Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. *apj*, 455:7, December 1995. doi: 10.1086/176550.

David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

Prasanta Chandra Mahalanobis. On the generalized distance in statistics. National Institute of Science of India, 1936.

Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

R. Mandelbaum. Weak lensing for precision cosmology. *ArXiv e-prints*, October 2017.

Augustin Marignier, Ana M. G. Ferreira, and Thomas Kitching. The probability of mantle plumes in global tomographic models. *Geochemistry, Geophysics, Geosystems*, 21(9):e2020GC009276, 2020. doi: https://doi.org/10.1029/2020GC009276.

D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, and N. Vittorio. Spherical needlets for cosmic microwave background data analysis. *Monthly Notices of the Royal Astronomical Society*, 383(2):539–545, 12 2007. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2007.12550.x. URL `https://doi.org/10.1111/j.1365-2966.2007.12550.x`.

Maxim Markevitch, AH Gonzalez, D Clowe, A Vikhlinin, W Forman, C Jones, S Murray, and W Tucker. Direct constraints on the dark matter self-interaction cross section from the merging galaxy cluster 1e 0657–56. *The Astrophysical Journal*, 606(2):819, 2004.

Richard Massey, Jason Rhodes, Richard Ellis, Nick Scoville, Alexie Leauthaud, Alexis Finoguenov, Peter Capak, David Bacon, Hervé Aussel, Jean-Paul Kneib, et al. Dark matter maps reveal cosmic scaffolding. *Nature*, 445(7125):286, 2007a.

Richard Massey, Jason Rhodes, Alexie Leauthaud, Peter Capak, Richard Ellis, Anton Koekemoer, Alexandre Réfrégier, Nick Scoville, James E Taylor, Justin Albert, et al. Cosmos: three-dimensional weak lensing and the growth of structure. *The Astrophysical Journal Supplement Series*, 172(1):239, 2007b.

J. McEwen and Matthew A. Price. Scale-discretised ridgelet transform on the sphere. *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5, 2019.

J. D. McEwen and B. Leistedt. Fourier-Laguerre transform, convolution and wavelets on the ball. *arXiv e-prints*, art. arXiv:1307.1307, July 2013.

J. D. McEwen and A. M. M. Scaife. Simulating full-sky interferometric observations. *Monthly Notices of the Royal Astronomical Society*, 389(3):1163–1178, 2008. doi: 10.1111/j.1365-2966.2008.13690.x.

J. D. McEwen and Y. Wiaux. A Novel Sampling Theorem on the Sphere. *IEEE Transactions on Signal Processing*, 59:5876–5887, December 2011. doi: 10.1109/TSP.2011.2166394.

J. D. McEwen, Y. Wiaux, and D. M. Eyers. Data compression on the sphere. *Astronomy and Astrophysics*, 531:A98, 2011. doi: 10.1051/0004-6361/201015728.

J. D. McEwen, G. Puy, J.-P. Thiran, P. Vandergheynst, D. Van De Ville, and Y. Wiaux. Sparse Image Reconstruction on the Sphere: Implications of a New Sampling Theorem. *IEEE Transactions on Image Processing*, 22:2275–2285, June 2013a. doi: 10.1109/TIP.2013.2249079.

J. D. McEwen, M. Buttner, B. Leistedt, H. V. Peiris, and Y. Wiaux. A Novel Sampling Theorem on the Rotation Group. *IEEE Signal Processing Letters*, 22: 2425–2429, December 2015a. doi: 10.1109/LSP.2015.2490676.

J. D. McEwen, M. Buttner, B. Leistedt, H. V. Peiris, and Y. Wiaux. A Novel Sampling Theorem on the Rotation Group. *IEEE Signal Processing Letters*, 22: 2425–2429, December 2015b. doi: 10.1109/LSP.2015.2490676.

J. D. McEwen, B. Leistedt, M. Büttner, H. V. Peiris, and Y. Wiaux. Directional spin wavelets on the sphere. *ArXiv e-prints*, September 2015c.

Jason D. McEwen, Pierre Vandergheynst, and Yves Wiaux. On the computation of directional scale-discretized wavelet transforms on the sphere. In *Proceedings of the SPIE*, volume 8858 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 88580I, Sep 2013b. doi: 10.1117/12.2022889.

Jason D. McEwen, Claudio Durastanti, and Yves Wiaux. Localisation of directional scale-discretised wavelets on the sphere. *Applied and Computational Harmonic Analysis*, 44(1):59 – 88, 2018. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2016.03.009. URL http://www.sciencedirect.com/science/article/pii/S1063520316000324.

John P McFarland, Gijs Verdoes-Kleijn, Gert Sikkema, Ewout M Helmich, Danny R Boxhoorn, and Edwin A Valentijn. The astro-wise optical image pipeline. *Experimental Astronomy*, 35(1-2):45–78, 2013.

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

Yves Meyer. *Wavelets and operators*, volume 1. Cambridge university press, 1995.

V. Michel. Wavelets on the 3 dimensional ball. *PAMM*, 5(1):775–776, 2005. doi: https://doi.org/10.1002/pamm.200510362.

L. Miller et al. Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey. *Monthly Notices of the Royal Astronomical Society*, 429(4):2858–2880, Mar 2013. doi: 10.1093/mnras/sts454.

R. A. Millikan. A direct photoelectric determination of planck's "h". *Phys. Rev.*, 7: 355–388, Mar 1916. doi: 10.1103/PhysRev.7.355. URL `https://link.aps.org/doi/10.1103/PhysRev.7.355`.

MIT. Electrical engineering and computer science. *MIT OpenCourseWare, Massachusetts Institute of Technology*, 2012. URL `https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-253-convex-analysis-and-optimization-spring-2012/lecture-notes/MIT6_253S12_lec02.pdf`.

G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38 (8), 1965.

J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C.R. Acad. Sci. Paris Ser. A Math.*, 255:2897–2899, 1962. URL `https://ci.nii.ac.jp/naid/10025986654/en/`.

E. Morganson and DES Collaboration. The Dark Energy Survey Image Processing Pipeline. *Publications of the ASP*, 130(989):074501, Jul 2018. doi: 10.1088/1538-3873/aab4ef.

D. Munshi and P. Coles. The integrated bispectrum and beyond. *Journal of Cosmology and Astroparticle Physics*, 2:010, February 2017. doi: 10.1088/1475-7516/2017/02/010.

D. Munshi, P. Valageas, L. van Waerbeke, and A. Heavens. Cosmology with weak lensing surveys. *Physics Reports*, 462:67–121, June 2008. doi: 10.1016/j.physrep.2008.02.003.

D. Munshi, L. van Waerbeke, J. Smidt, and P. Coles. From weak lensing to non-Gaussianity via Minkowski functionals. *Monthly Notices of the Royal Astronomical Society*, 419:536–555, January 2012. doi: 10.1111/j.1365-2966.2011.19718.x.

Sangnam Nam, Mike E Davies, Michael Elad, and Rémi Gribonval. The cosparse analysis model and algorithms. *Applied and Computational Harmonic Analysis*, 34 (1):30–56, 2013.

Francis J Narcowich, Pencho Petrushev, and Joseph D Ward. Localized tight frames on spheres. *SIAM Journal on Mathematical Analysis*, 38(2):574–594, 2006.

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

E. T. Newman and R. Penrose. Note on the bondi-metzner-sachs group. *Journal of Mathematical Physics*, 7(5):863–870, 1966. doi: 10.1063/1.1931221. URL `https://doi.org/10.1063/1.1931221`.

Alexandru Onose, Rafael E. Carrillo, Audrey Repetti, Jason D. McEwen, Jean-Philippe Thiran, Jean-Christophe Pesquet, and Yves Wiaux. Scalable splitting algorithms for big-data interferometric imaging in the SKA era. *Monthly Notices of the Royal Astronomical Society*, 462(4):4314–4335, Nov 2016. doi: 10.1093/mnras/stw1859.

N. Padmanabhan, U. Seljak, and U. L. Pen. Mining weak lensing surveys. *New Astronomy*, 8:581–603, August 2003. doi: 10.1016/S1384-1076(03)00055-1.

Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.

P. Paykari, F. Lanusse, J.-L. Starck, F. Sureau, and J. Bobin. Prism: Sparse recovery of the primordial power spectrum. *Astronomy and Astrophysics*, 566, 02 2014.

A. Peel, F. Lanusse, and J.-L. Starck. Sparse Reconstruction of the Merging A520 Cluster System. *The Astrophysical Journal*, 847:23, September 2017a. doi: 10.3847/1538-4357/aa850d.

A. Peel, C.-A. Lin, F. Lanusse, A. Leonard, J.-L. Starck, and M. Kilbinger. Cosmological constraints with weak-lensing peak counts and second-order statistics in a large-field survey. *Astronomy and Astrophysics*, 599:A79, March 2017b. doi: 10.1051/0004-6361/201629928.

A. Peel, V. Pettorino, C. Giocoli, J.-L. Starck, and M. Baldi. Breaking degeneracies in modified gravity with higher (than 2nd) order weak-lensing statistics. *ArXiv e-prints*, May 2018.

Marcelo Pereyra. Proximal markov chain monte carlo algorithms. *Statistics and Computing*, 26(4):745–760, Jul 2016. ISSN 1573-1375. doi: 10.1007/s11222-015-9567-4. URL https://doi.org/10.1007/s11222-015-9567-4.

Marcelo Pereyra. Maximum-a-posteriori estimation with bayesian confidence regions. *SIAM Journal on Imaging Sciences*, 10(1):285–302, 2017. doi: 10.1137/16M1071249. URL https://doi.org/10.1137/16M1071249.

Marcelo Pereyra, Jose M. Bioucas-Dias, and MÃąrio A. T. Figueiredo. *Maximum-a-posteriori estimation with unknown regularisation parameters*, pages 230–234. 12 2015. doi: 10.1109/EUSIPCO.2015.7362379.

Jean-Christophe Pesquet, Amel Benazza-Benyahia, and Caroline Chaux. A sure approach for digital signal/image deconvolution problems. *IEEE Transactions on Signal Processing*, 57(12):4616–4632, 2009.

C. Pichon, E. Thiébaut, S. Prunet, K. Benabed, S. Colombi, T. Sousbie, and R. Teyssier. ASKI: full-sky lensing map-making algorithms. *Monthly Notices of the Royal Astronomical Society*, 401:705–726, January 2010. doi: 10.1111/j.1365-2966.2009.15609.x.

D. Piras, B. Joachimi, B. M. Schäfer, M. Bonamigo, S. Hilbert, and E. van Uitert. The mass dependence of dark matter halo alignments with large-scale structure. *Monthly Notices of the Royal Astronomical Society*, 474:1165–1175, February 2018. doi: 10.1093/mnras/stx2846.

S. Pires, J.-L Starck, and A. Amara. Weak Gravitational Lensing. *Monthly Notices of the Royal Astronomical Society*, 395:1265, August 2009.

Sandrine Pires, J-L Starck, Adam Amara, Alexandre Réfrégier, and Romain Teyssier. Cosmological model discrimination from weak lensing data. In *AIP Conference Proceedings*, volume 1241, pages 1118–1127. American Institute of Physics, 2010.

Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. *ArXiv e-prints*, July 2018.

Henri Poincaré. *Calcul des probabilités*, volume 1. Gauthier-Villars, 1912.

Harry Pollard. Representation of an analytic function by a laguerre series. *Annals of Mathematics*, 48(2):358–365, 1947.

Luke Pratley, Jason D. McEwen, Mayeul d'Avezac, Rafael E. Carrillo, Alexandru Onose, and Yves Wiaux. Robust sparse image reconstruction of radio interferometric observations with PURIFY. *Monthly Notices of the Royal Astronomical Society*, 473(1):1038–1058, Jan 2018. doi: 10.1093/mnras/stx2237.

Luke Pratley et al. A Fast and Exact w-stacking and w-projection Hybrid Algorithm for Wide-field Interferometric Imaging. *The Astrophysical Journal*, 874(2):174, April 2019. doi: 10.3847/1538-4357/ab0a05.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition, 2007. ISBN 0521880688.

M. A. Price and J. D. McEwen. Bayesian variational regularization on the ball. *arXiv preprint arXiv:2105.05518*, 2021.

M. A. Price, X Cai, J. D. McEwen, M Pereyra, and T. D. Kitching. Sparse Bayesian mass mapping with uncertainties: local credible intervals. *Monthly Notices of the Royal Astronomical Society*, 492(1):394–404, December 2019a. doi: 10.1093/mnras/stz3453.

M. A. Price, J. D. McEwen, X Cai, and T. D. Kitching. Sparse Bayesian mass mapping with uncertainties: peak statistics and feature locations. *Monthly Notices of the Royal Astronomical Society*, 489(3):3236–3250, December 2019b. doi: 10.1093/mnras/stz2373.

M. A. Price, J. D. McEwen, X Cai, T. D. Kitching, and C. G. R. Wallis. Sparse Bayesian mass-mapping with uncertainties: hypothesis testing of structure. *Monthly Notices of the Royal Astronomical Society*, 07 2021a. ISSN 0035-8711. doi: 10.1093/mnras/stab1983. URL https://doi.org/10.1093/mnras/stab1983. stab1983.

M. A. Price, J. D. McEwen, L. Pratley, and T. D. Kitching. Sparse Bayesian mass-mapping with uncertainties: Full sky observations on the celestial sphere. *Monthly Notices of the Royal Astronomical Society*, 500(4):5436–5452, January 2021b. doi: 10.1093/mnras/staa3563.

M. A. Price, L. Pratley, and J. D. McEwen. Sparse image reconstruction on the sphere: a general approach with uncertainty quantification. *submitted to IEEE Transactions on Image Processing*, art. arXiv:2105.04935, 5 2021c.

Holger Rauhut and Rachel Ward. Sparse legendre expansions via ℓ_1-minimization. *Journal of approximation theory*, 164(5):517–533, 2012.

S. Refsdal. On the possibility of determining hubble's parameter and the masses of galaxies from the gravitational lens effect. *Monthly Notices of the Royal Astronomical Society*, 128(4):307–310, 1964.

S. Refsdal and H. Bondi. The gravitational lens effect. *Monthly Notices of the Royal Astronomical Society*, 128(4):295–306, 1964.

Audrey Repetti, Marcelo Pereyra, and Yves Wiaux. Scalable bayesian uncertainty quantification in imaging inverse problems via convex optimization. *SIAM Journal on Imaging Sciences*, 12(1):87–118, 2019.

J Ritsema, a A Deuss, HJ Van Heijst, and JH Woodhouse. *Geophysical Journal International*, 184(3):1223–1236, 2011.

C.-P. Robert. *The Bayesian Choice*, volume 340. 2001. doi: https://doi.org/10.1007/0-387-71599-1.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. ISSN 13507265. URL http://www.jstor.org/stable/3318418.

Gareth O Roberts, Andrew Gelman, Walter R Gilks, et al. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.

H. P. Robertson. Kinematics and World-Structure. *apj*, 82:284, November 1935. doi: 10.1086/143681.

Donald B Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, pages 1151–1172, 1984.

D. Saadeh, S. M. Feeney, A. Pontzen, H. V. Peiris, and J. D. McEwen. How Isotropic is the Universe? *Physical Review Letters*, 117(13):131302, September 2016. doi: 10.1103/PhysRevLett.117.131302.

Robert Schlaifer and Howard Raiffa. *Applied statistical decision theory*. 1961.

M. D. Schneider, D. W. Hogg, P. J. Marshall, W. A. Dawson, J. Meyers, D. J. Bard, and D. Lang. Hierarchical Probabilistic Inference of Cosmic Shear. *The Astrophysical Journal*, 807:87, July 2015. doi: 10.1088/0004-637X/807/1/87.

P. Schneider. Weak Gravitational Lensing. *ArXiv Astrophysics e-prints*, September 2005.

P. Schröder and W. Sweldens. Spherical wavelets: efficiently representing functions on the sphere. In *Computer Graphics Proceedings (SIGGRAPH '95)*, pages 161–172, 1995.

Thomas Schuster, Barbara Kaltenbacher, Bernd Hofmann, and Kamil S Kazimierski. *Regularization methods in Banach spaces*, volume 10. Walter de Gruyter, 2012.

Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

Frederik J. Simons et al. Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. *Geophysical Journal International*, 187(2):969–988, November 2011. doi: 10.1111/j.1365-246X.2011.05190.x.

Devinderjit Sivia and John Skilling. *Data analysis: a Bayesian tutorial*. OUP Oxford, 2006.

J. Soldner. On the Deflection of a Light Ray from its Rectilinear Motion. *Astronomisches Jahrbuch*, page 161, March 1804.

Jean-Luc Starck, Fionn Murtagh, and Jalal M Fadili. *Sparse image and signal processing: wavelets, curvelets, morphological diversity*. Cambridge university press, 2010.

Starck, J.-L., Yassir Moudden, Pierrick Abrial, and Mai Nguyen. Wavelets, ridgelets and curvelets on the sphere. *Astronomy & Astrophysics*, 446(3):1191–1204, 2006a.

Starck, J.-L., Pires, S., and Réfrégier, A. Weak lensing mass reconstruction using wavelets. *A&A*, 451(3):1139–1150, 2006b. doi: 10.1051/0004-6361:20052997. URL `https://doi.org/10.1051/0004-6361:20052997`.

Starck, J.-L., Themelis, K. E., Jeffrey, N., Peel, A., and Lanusse, F. Weak-lensing mass reconstruction using sparsity and a gaussian random field. *A&A*, 649: A99, 2021. doi: 10.1051/0004-6361/202039451. URL `https://doi.org/10.1051/0004-6361/202039451`.

Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pages 1135–1151, 1981.

J M Stewart. Perturbations of friedmann-robertson-walker cosmological models. *Classical and Quantum Gravity*, 7(7):1169, 1990. URL `http://stacks.iop.org/0264-9381/7/i=7/a=013`.

Mervyn Stone. Strong inconsistency from uniform priors. *Journal of the American Statistical Association*, 71(353):114–116, 1976. ISSN 01621459. URL `http://www.jstor.org/stable/2285744`.

Masahiro Takada and Bhuvnesh Jain. Cosmological parameters from lensing power spectrum and bispectrum tomography. *Monthly Notices of the Royal Astronomical Society*, 348(3):897–915, 2004. doi: 10.1111/j.1365-2966.2004.07410.x. URL `http://dx.doi.org/10.1111/j.1365-2966.2004.07410.x`.

Ryuichi Takahashi, Takashi Hamana, Masato Shirasaki, Toshiya Namikawa, Takahiro Nishimichi, Ken Osato, and Kosei Shiroyama. Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments. *Astrophysical Journal*, 850(1):24, Nov 2017. doi: 10.3847/1538-4357/aa943d.

Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*, volume 89. siam, 2005.

P. L. Taylor, T. D. Kitching, and J. D. McEwen. Preparing for the cosmic shear data flood: Optimal data extraction and simulation requirements for stage IV dark energy experiments. *prd*, 98(4):043532, August 2018. doi: 10.1103/PhysRevD.98.043532.

Peter L Taylor, Thomas D Kitching, Jason D McEwen, and Thomas Tram. Testing the cosmic shear spatially-flat universe approximation with generalized lensing and shear spectra. *Physical Review D*, 98(2):023522, 2018.

V.N. Temlyakov. Greedy algorithms andm-term approximation with regard to redundant dictionaries. *Journal of Approximation Theory*, 98(1):117 – 145, 1999. ISSN 0021-9045. doi: https://doi.org/10.1006/jath.1998.3265. URL http://www.sciencedirect.com/science/ARTICLE/pii/S0021904598932658.

R. Teyssier, S. Pires, S. Prunet, D. Aubert, C. Pichon, A. Amara, K. Benabed, S. Colombi, A. Refregier, and J.-L. Starck. Full-sky weak-lensing simulation with 70 billion pARTICLEs. *Astronomy and Astrophysics*, 497:335–341, April 2009. doi: 10.1051/0004-6361/200810657.

Robert Tibshirani. Regression shrinkage and selection via the lasso, 1996.

R. Trotta. Bayesian Methods in Cosmology. *ArXiv e-prints*, January 2017.

M. A. Troxel and M. Ishak. The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology. *Physics Reports*, 558: 1–59, February 2015. doi: 10.1016/j.physrep.2014.11.001.

David S Tuch. Q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 52(6):1358–1372, 2004.

J. A. Tyson, F. Valdes, and R. A. Wenk. Detection of systematic gravitational lens galaxy image alignments - Mapping dark matter in galaxy clusters. *apjl*, 349: L1–L4, January 1990. doi: 10.1086/185636.

Clemens Valens. A really friendly guide to wavelets. *ed. Clemens Valens*, 1999.

Z. M. Vallis, C. G. R. Wallis, and T. D. Kitching. On the effect of projections on convergence peak counts and Minkowski functionals. *Astronomy and Computing*, 24:84, Jul 2018. doi: 10.1016/j.ascom.2018.06.004.

Ludovic Van Waerbeke, Y Mellier, T Erben, JC Cuillandre, F Bernardeau, R Maoli, E Bertin, HJ Mc Cracken, O Le Fevre, B Fort, et al. Detection of correlated galaxy ellipticities on cfht data: first evidence for gravitational lensing by large-scale structures. *arXiv preprint astro-ph/0002500*, 2000.

J. T. VanderPlas, A. J. Connolly, B. Jain, and M. Jarvis. Three-dimensional Reconstruction of the Density Field: An SVD Approach to Weak-lensing Tomography. *The Astrophysical Journal*, 727:118, February 2011. doi: 10.1088/0004-637X/727/2/118.

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948, 2013. doi: 10.1109/GlobalSIP.2013.6737048.

Ana Fernandez Vidal and Marcelo Pereyra. Maximum likelihood estimation of regularisation parameters. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1742–1746. IEEE, 2018.

A. G. Walker. On milne's theory of world-structure*. *Proceedings of the London Mathematical Society*, s2-42(1):90–127, 1937. doi: 10.1112/plms/s2-42.1.90. URL `http://dx.doi.org/10.1112/plms/s2-42.1.90`.

C. G. R. Wallis, Y. Wiaux, and J. D. McEwen. Sparse Image Reconstruction on the Sphere: Analysis and Synthesis. *IEEE Transactions on Image Processing*, 26: 5176–5187, November 2017. doi: 10.1109/TIP.2017.2716824.

Christopher G R Wallis, Matthew A Price, Jason D McEwen, Thomas D Kitching, Boris Leistedt, and Antoine Plouviez. Mapping dark matter on the celestial sphere with weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 509(3):4480–4497, 11 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab3235. URL `https://doi.org/10.1093/mnras/stab3235`.

Dennis Walsh, Robert F Carswell, and Ray J Weymann. 0957+ 561 a, b: twin quasistellar objects or gravitational lens? *Nature*, 279(5712):381, 1979.

Steven Weinberg. *Cosmology*. Oxford university press, 2008.

Ernst Joachim Weniger. On the analyticity of Laguerre series. *Journal of Physics A Mathematical General*, 41(42):425207, October 2008. doi: 10.1088/1751-8113/41/42/425207.

Y. Wiaux, J. D. McEwen, P. Vandergheynst, and O. Blanc. Exact reconstruction with directional wavelets on the sphere. *Monthly Notices of the Royal Astronomical Society*, 388:770–788, August 2008. doi: 10.1111/j.1365-2966.2008.13448.x.

Norbert Wiener et al. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*, volume 8. MIT press Cambridge, MA, 1964.

David M Wittman, J Anthony Tyson, David Kirkman, Ian Dell'Antonio, and Gary Bernstein. Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales. *Nature*, 405(6783):143, 2000.

Jianxiong Xiao, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Recognizing scene viewpoint using panoramic place representation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2695–2702. IEEE, 2012.

Q. Yang. Advanced Convex Optimization. feb 2015.

J. Yoo, N. Grimm, E. Mitsou, A. Amara, and A. Refregier. Gauge-invariant formalism of cosmological weak lensing. *Journal of Cosmology and Astroparticle Physics*, 4: 029, April 2018. doi: 10.1088/1475-7516/2018/04/029.

Ya. B. Zel'dovich. *Soviet Ast.*, 8:13, 1964.

F. Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.

F. Zwicky. Nebulae as gravitational lenses. *Physical Review*, 51(4):290, 1937a.

F. Zwicky. On the probability of detecting nebulae which act as gravitational lenses. *Physical Review*, 51(8):679, 1937b.