Bayesian statistical inference and deep learning for primordial cosmology and cosmic acceleration

Doogesh Kodi Ramanah

Sorbonne Université

École Doctorale d'Astronomie et d'Astrophysique d'île-de-France (ED 127)

Institut d'Astrophysique de Paris

Bayesian statistical inference and deep learning for primordial cosmology and cosmic acceleration

Par

Doogesh KODI RAMANAH

Thèse de doctorat

Discipline: Cosmologie

Présentée et soutenue publiquement le 10 septembre 2019

Devant un jury composé de:

Bertrand LAFORGE	Professeur, Sorbonne Université	*Président du jury*
Guilhem LAVAUX	Chargé de recherche, Sorbonne Université	*Directeur de thèse*
Benjamin WANDELT	Professeur, Sorbonne Université	*Directeur de thèse*
Carlo CONTALDI	Professeur, Imperial College London	*Rapporteur*
Anthony CHALLINOR	Professeur, University of Cambridge	*Rapporteur*
Hiranya PEIRIS	Professeure, University College London	*Examinatrice*
David ALONSO	Chercheur, University of Oxford	*Examinateur*
Valérie DE LAPPARENT	Directrice de recherche, Sorbonne Université	*Invitée*

Publications related to this thesis

This thesis has used material from the following first- and co-authored publications:

- "*Wiener filtering and pure $\mathcal{E}$ / $\mathcal{B}$ decomposition of CMB maps with anisotropic correlated noise*," **D. Kodi Ramanah**, G. Lavaux, B. D. Wandelt, 2019, MNRAS, submitted (arXiv:1906.10704)

- "*Painting halos from cosmic density fields of dark matter with physically motivated neural networks*," **D. Kodi Ramanah**, T. Charnock, G. Lavaux, 2019, Phys. Rev. D, 100, 043515 (arXiv:1903.10524)

- "*Explicit Bayesian treatment of unknown foreground contaminations in galaxy surveys*," N. Porqueres, **D. Kodi Ramanah**, J. Jasche, G. Lavaux, 2019, A&A, 624, A115 (arXiv:1812.05113)

- "*Cosmological inference from Bayesian forward modelling of deep galaxy redshift surveys*," **D. Kodi Ramanah**, G. Lavaux, J. Jasche, B. D. Wandelt, 2019, A&A, 612, A69 (arXiv:1808.07496)

- "*Optimal and fast $\mathcal{E}$ / $\mathcal{B}$ separation with a dual messenger field*," **D. Kodi Ramanah**, G. Lavaux, B. D. Wandelt, 2018, MNRAS, 476, 2825 (arXiv:1801.05358)

- "*Wiener filter reloaded: fast signal reconstruction without preconditioning*," **D. Kodi Ramanah**, G. Lavaux, B. D. Wandelt, 2017, MNRAS, 468, 1782 (arXiv:1702.08852)

Résumé

Cette thèse a pour vocation le développement et l'application de nouvelles techniques d'inférence statistique bayésienne et d'apprentissage profond aux sondes cosmologiques, telles que le fond diffus cosmologique ou CMB (pour cosmic microwave background) et les grandes structures de l'Univers. Nous avons conçu des outils d'analyse de données efficaces et précis pour relever les défis statistiques imposés par les gros volumes de données complexes des missions CMB ou des relevés profonds de galaxies de la prochaine génération, dans le but d'optimiser l'exploitation des données scientifiques afin d'améliorer, à terme, notre compréhension de l'Univers.

La première partie de cette thèse concerne l'extraction des modes $\mathcal{E}$ et $\mathcal{B}$ du signal de polarisation du CMB à partir des données. À cette fin, nous avons utilisé le filtre de Wiener, un outil puissant pour la reconstruction de signaux, mais qui nécessite l'inversion de matrices denses, ce qui pose un problème non-trivial pour traiter les gros volumes de données. En conséquence, nous avons développé une méthode hiérarchique à haute performance, nommée algorithme du dual messenger, qui garantit la stabilité et la convergence numérique tout en évitant l'inversions de ces matrices. Nous avons réalisé une demonstration de faisabilité en appliquant l'algorithme aux cartes d'anisotropies de température du CMB. Nous avons ensuite adapté notre méthode pour traiter les données de polarisation du CMB et démontré son efficacité pour réaliser la séparation $\mathcal{E}/\mathcal{B}$, tout en illustrant les difficultés des approches traditionnelles pour gérer efficacement ce problème intrinsèquement mal posé. Nous avons ensuite généralisé l'algorithme du dual messenger pour la reconstruction du champ de spin sur la sphère, tout en tenant compte des modèles de bruit réalistes tels que le bruit anisotrope corrélé. En utilisant des cartes de polarisation simulées du CMB de la mission Planck comme tests, nous avons démontré les capacités de cet algorithme à reconstruire des cartes $\mathcal{E}$ et $\mathcal{B}$ pures, garanti par la suppression des modes ambigus liés au problème de couplage qui entrave les méthodes classiques de séparation $\mathcal{E}/\mathcal{B}$.

La seconde partie porte sur le développement de divers aspects de la modélisation d'inférence bayésienne pour une exploitation optimale des relevés modernes du décalage vers le rouge ou redshift des galaxies. Nous avons développé un cadre d'inférence bayésienne à grande échelle pour contraindre les paramètres cosmologiques en s'appuyant sur une nouvelle implémentation du test géométrique d'Alcock-Paczyński (AP). Notre approche, qui repose uniquement sur les symétries géométriques du principe cosmologique, extrait plusieurs ordres de grandeur d'informations supplémentaires de l'expansion cosmique par rapport aux approches classiques. En particulier, nous avons présenté nos contraintes cosmologiques sur la densité de matière et l'équation d'état de l'énergie sombre. Nous avons vérifié que notre implémentation du test AP est robuste à un modèle mal spécifié qui offre un degré de solidité remarquable au phénomène actuellement non résolu du biais des galaxies, contournant ainsi une des principales limites des analyses traditionnelles. Etant donné que le contrôle des effets systématiques est un facteur crucial pour les relevés modernes de redshift des galaxies, nous avons également présenté une fonction de vraisemblance robuste, qui résiste aux contaminations inconnues liées aux avant-plans.

Finalement, dans le but de construire des émulateurs de dynamiques complexes dans notre modèle hiérarchique bayésien, nous avons conçu un nouveau réseau de neurones qui apprend à peindre des distributions de halo réalistes sur des champs approximatifs de matière noire en 3D. Nous avons utilisé des principes physiques simples pour former le réseau de neurones afin d'apprendre la relation locale et non-triviale entre le champ de densité de la matière noire et la distribution de halo sans avoir recours à un modèle physique.

Mots-clés: Cosmologie, grandes structures de l'Univers, fond dissus cosmologique, méthodes statistiques, apprentissage profond, l'analyse des données.

Abstract

The essence of this doctoral research constitutes the development and application of novel Bayesian statistical inference and deep learning techniques to cosmological probes, such as cosmic microwave background (CMB) and large-scale structure measurements. We have designed efficient and precise data analysis tools to meet statistical challenges of massive and complex data sets from next-generation CMB missions or galaxy surveys and optimize their scientific returns to ultimately improve our understanding of the Universe. This thesis is organized in three main parts.

The first theme deals with the extraction of the $\mathcal{E}$ and $\mathcal{B}$ modes of the CMB polarization signal from the data. To this end, we employ the Wiener filter, a powerful signal reconstruction tool, but which requires the inversion of dense matrices, resulting in a highly non-trivial problem when dealing with large data sets. We have developed a high-performance hierarchical method, known as the dual messenger algorithm, which guarantees numerical stability and convergence, while obviating matrix inversions. As a proof of concept, we presented an application of the algorithm to maps of CMB temperature anisotropies. We then adapted our method for the inclusion of CMB polarization data and demonstrated its efficacy in performing $\mathcal{E}/\mathcal{B}$ separation, while illustrating the difficulties of traditional approaches in dealing efficiently with this inherently ill-conditioned problem. We subsequently extended the dual messenger algorithm for spin field reconstruction on the sphere, while accounting for complex and realistic noise models such as correlated anisotropic noise. Using simulated Planck CMB polarization maps as a test case, we demonstrated the capabilities of the algorithm in reconstructing pure $\mathcal{E}$ and $\mathcal{B}$ maps, guaranteed to be free from ambiguous modes resulting from the leakage issue that plagues conventional methods of $\mathcal{E}/\mathcal{B}$ separation.

The second theme of this doctoral work lies in the development of various aspects of Bayesian forward modelling machinery for optimal exploitation of state-of-the-art galaxy redshift surveys. We developed a large-scale Bayesian inference framework to constrain cosmological parameters via a novel implementation of the Alcock-Paczyński (AP) test. Our approach, relying purely on the geometrical symmetries of the cosmological principle, extracts several orders of magnitude more information from the cosmic expansion relative to classical approaches. In particular, we showcased our cosmological constraints on the matter density and dark energy equation of state. We verified that our implementation of the AP test is robust to a misspecified model, yielding a remarkable degree of resilience to the currently unresolved phenomenon of galaxy bias, thereby circumventing a potentially key limitation. This is consequently among the first methods to extract a large fraction of information from statistics other than that of direct density contrast correlations, without being sensitive to the amplitude of density fluctuations. With the control of systematic effects being a crucial limiting factor for modern galaxy redshift surveys, we also presented an augmented likelihood which is robust to unknown foreground and target contaminations.

Finally, with a view to building fast complex dynamics emulators in our above Bayesian hierarchical model, we have designed a novel halo painting network that learns to map approximate 3D dark matter fields to realistic halo distributions. We used simple physical principles to train the mapping network to learn the non-trivial local relation between dark matter density field and halo distributions without relying on a physical model. Combining statistical inference and deep generative modelling techniques is the plausible approach to render high-resolution analyses of upcoming galaxy redshift surveys feasible, while providing statistically interpretable results and maintaining the scientific rigour.

Keywords: Cosmology, large-scale structure of the Universe, cosmic microwave background, statistical methods, deep learning, data analysis.

Acknowledgements

This thesis marks the culmination of an absolutely wonderful journey at the Institut d'Astrophysique de Paris (IAP). Whilst it has not been plain sailing, as expected, I am delighted with the way the PhD has panned out and the friends and connections made during these three years. I must imperatively express my gratitude to all the persons who directly or indirectly have played a key role in helping to shape this PhD.

I will always be grateful to my two supervisors, Guilhem Lavaux and Ben Wandelt, for their invaluable guidance and support throughout the PhD. Guilhem was deeply involved in all of our projects, and was always available to answer even my silliest query. He has been remarkably patient and I am thankful for his hands-on approach to supervision. I am proud to be his first PhD student but he will undoubtedly get much brighter students to work with! Ben was a stark contrast to Guilhem, but his wise words at the very start helped to set the tone and paid off beautifully at the end. With Ben and Guilhem around, I felt optimistic that we would eventually crack any of the (numerous) problems that cropped up. Tom Charnock was an amazing tutor for helping me get to grips with TensorFlow for our machine learning projects. Working with him was great fun and I very much hope that we will keep working together on our endless list of projects! Guillaume Faye, my thesis Godfather, kept track of my progress and took much interest in my work, so my sincere thanks to him.

I am deeply honoured to have assembled a jury of such high standard and I thank them for having responded so promptly and positively to the invitation. In particular, my two reviewers, Carlo Contaldi and Anthony Challinor, sacrificed a substantial part of their summer to review the thesis. Carlo was instrumental in my choice to embark on the doctoral programme at IAP, a decision that leaves no room for regrets, and Anthony was always very welcoming during my visit in Cambridge. David Alonso has been extremely supportive and enthusiastic about our work and I thank him for the hospitality shown at Oxford. Hiranya Peiris, whom I admire for her impressive track record, also found some time in her packed September schedule. We are thankful to Bertrand Laforge who very kindly accepted to be the President of the jury. Many thanks to Valérie de Lapparent for her caring nature, being ever-present since the very first day, ensuring that PhD students have ideal working conditions.

Office 035 will always be a very special part of my life, having spent three years in this peaceful haven, impervious to loquacious PhD students during working hours. It was a pleasure to share the office with Matteo Rizzato, Arno Vanthieghem and occasionally Martin Pernot-Borras. I made some great friends here, and I hope our friendship will last beyond the PhD. Matteo is one of the nicest guys around, and his self-deprecating sense of humour provided us with hours of sheer delight! It was easy to forget sometimes that Arno was in the office (!), but the three of us, along with Pierre Boldrini and Shweta Dalal, shared some wonderful times together, be it on our many evening walks, dinners and drinks. I am grateful to Pierre for our gaming nights and manga/anime discussions, and for initiating me to table tennis and futsal, which added an extra dimension to life in Paris. I will also fondly remember Alexandre (Baptou) Barthélémy and his exuberant personality, and Lukas Furtak with his deadly jokes. They all made life here so much more interesting and I will sorely miss the morning and afternoon coffee breaks with them. Thanks also to Siwei-chan, Hugo, Corentin, Claire, Gohar, Adèle, Adarsh, Mo, Amael, Simon, Lucas, François, and Quentin for maintaining the vibrant atmosphere at IAP.

I am privileged to be a founding member of the Aquila consortium, which has provided me with a fantastic platform for stimulating discussions and fruitful collaborations. I express my gratitude to all the members of the consortium. Jens Jasche has been like a third supervisor to me, who always prioritizes students' progress and firmly desires a bright future for them. I thoroughly enjoyed working with Natalia Porqueres, whose determination and strong work ethic was inspiring. Franz Elsner's knowledge/expertise on CMB data analysis pipelines was very helpful. Harry

Desmond ensured that my Oxford visit went on smoothly and advised me on postdoctoral applications. I have also enjoyed interesting discussions with Florent Leclercq, Suvodip Mukherjee, Fabian Schmidt, Supranta Sarma Boruah and Florian Führer. I would like to place on record the strong support of Alan Heavens during the past few years, since my Masters at Imperial College London, and has been an inspiring presence throughout. I am grateful and fortunate to have a recommendation letter from him. I have met many interesting people on my visits: Pedro Ferreira, Blake Sherwin, Anthony Lasenby, Will Handley, Cora Uhlemann, Darsh Kodwani, Christiane Lorenz, Basundhara Ghosh and Sridevi Varma, just to name a few. It was great meeting my good friend Harry Mootoovaloo from Mauritius in various cosmology conferences. My apologies to anyone whom I have unwittingly left out.

My stay at IAP led to many interesting acquaintances. It was always nice to chat with Joe Silk, with his peerless knowledge of astrophysics and cosmology. I also thank him for twice inviting me to talk at the ICAP meetings. It was fun talking to peculiar and hilarious characters like Matt Lehnert and Cyril Pitrou, and the laid-back Eric Hivon. Christophe Gobet ensures IAP facilities are properly maintained and always acceded to my access requests even during odd hours. The IT people, especially the three Musqueteers (Laurent Domisse, Lionel Provost and Carlos Carvalho), were a massive help throughout the PhD. Laurent, whom I so often bugged, deserves a special mention for maintaining the workstation in top condition and for ensuring that my computations could run unhindered, and also for his great spirit and our ping pong rivalry (credit to Maxime Bocquier also here). It was great to share the office with the audiovisual expert Jean Mouette for some time, and get to experience first-hand the cool stuff he works on. I must imperatively convey my gratitude to Olivia Leroy, Roselys Rakotomandimby, Alexandre Dieng and Jason Payet for alleviating the administrative burden and helping me cope with the endless (and often pointless) administrative hassle, very much ingrained in the French system. Cynthia Kuyakula at the reception was a cheerful presence and regularly brightened our day with her unconventional pranks. Corinne, the lady at the canteen, was another humorous and warm addition to the daily PhD life. Florence Durret and Agnès Le Guerrer ensured that all the doctoral school formalities were also done smoothly and deserve much credit for that.

I will always be indebted to my parents for their unconditional love and support, and for always allowing me to follow my academic goals. My sister Kushma is another key pillar in my life and I cannot thank her and my brother-in-law Nitish enough for looking after and pampering me to no end! Special thoughts for my late grandparents who are always watching over me. My deepest gratitude to Pooja Surajbali for her immense love and iridescent presence in my life, and for being such a beautiful source of inspiration. I very much appreciated the company of Veeren Dullip, Santah and Kajal Jugdhur in Paris and thank them for their numerous dinner invitations. Finding a decent place in Paris is a daunting task, and in this context, I express my heartfelt thanks to Elise Doss-Tian for being such a kind landlady.

Last, but certainly not least, I convey my appreciation to Dante von Einzbern, who has been a magnificent companion for the past three years, for his pearls of wisdom and golden pieces of advice. As I move on to the next step in this tale of endless fun, and take on the Dark Fellowship at the Niels Bohr Institute in Copenhagen, I acknowledge the Dark Lord for his unwavering support. In the eternal words of Commander Erwin Smith: “Shinzō o sasageyo!”

Contents

Part I

Introduction

Context and motivation

Humanity is driven by an inherent desire and fascination to ask and answer fundamental questions about our Universe. Cosmology, as a natural science to describe and explain the origin and properties of the Universe, was born out of this epistemological and philosophical need for perspective on our existence and cosmic origins. Modern cosmology rests on firm mathematical and physical foundations, and has led to the emergence of a scientific theory that encodes the intricate machinery pertaining to the physics of the primordial Universe and its subsequent evolution, as laid out in Chapters 1 and 6.

Over the past few decades, there has been remarkable progress in cosmology, with the metamorphosis of this field of research into a precision and data-driven science. However, some crucial open questions and key unresolved issues must still be addressed, such as:

- Deciphering the nature of dark matter and dark energy;
- Detecting the elusive primordial gravitational waves to shed light on the early Universe;
- Verifying the inflationary paradigm as the precursor of cosmic structure formation;
- Probing our model of gravity for any intrinsic limitations;
- Understanding how galaxies trace the underlying dark matter, i.e. galaxy bias problem;
- Measuring the neutrino mass scale.

There are various pathways to solve or contribute towards solving the above problems, with the cosmic microwave background (CMB) and large-scale distribution of matter being particularly fertile hunting grounds to investigate and elucidate the inner-workings of the Universe.

Nevertheless, in an era of precision cosmology, the data analysis of state-of-the-art galaxy redshift surveys and CMB polarization maps with unprecedented levels of sensitivity and resolution poses complex numerical challenges. Fast and robust methods are required to render the data analysis of these large and complex data sets computationally tractable. The underlying goal of this work, therefore, is the development of precise and quantitative tools for statistical data analysis to efficiently explore the information hidden in our cosmic environment.

Window to probe early Universe physics

Observations of the CMB temperature anisotropies by the Planck satellite (e.g. Planck Collaboration et al., 2018a) and its predecessors, such as COsmic microwave Background Explorer (COBE) (Smoot et al., 1992) and the Wilkinson Microwave Anisotropy Probe (WMAP) (e.g. Dunkley et al., 2009), have made revolutionary contributions in constraining the currently favoured standard model of cosmology, as described in Chapter 1.3. With the Planck mission exhausting the accessible information content of the temperature sector, the next generation of CMB experiments is tailored for detailed measurements of the maps of polarization anisotropies. Some notable examples include the Simons Observatory (Ade et al., 2019) and CMB-S4 (Abazajian et al., 2016) experiments.

Cosmological inference from observations of the CMB polarization, however, requires the separation of the contributions of the gradient and curl (or $\mathcal{E}$ and $\mathcal{B}$) components of the polarization signal to the data. The $\mathcal{E}$-mode power spectrum provides an independent probe of the scalar modes measured via the temperature anisotropies (e.g. Abazajian et al., 2016; Suzuki et al., 2016; Henning et al., 2018; Louis et al., 2017). The inclusion of $\mathcal{E}$-mode polarization data in parameter inference pipelines allows us to derive more stringent constraints (e.g. Galli et al., 2014), whilst the scientific potential of $\mathcal{B}$-mode anisotropy observations is extremely promising.

The $\mathcal{B}$-mode component of the polarization is the focus of growing interest in the community. First, they are an independent confirmation of the lensing effect detected in the temperature and $\mathcal{E}$-mode anisotropies, as $\mathcal{B}$ modes are produced from the gravitational lensing of $\mathcal{E}$ modes by the dark matter distribution along the line of sight (Zaldarriaga and Seljak, 1998). These lensing-induced $\mathcal{B}$ modes have been observed by high-resolution ground-based CMB experiments (e.g. Hanson et al., 2013; The POLARBEAR Collaboration et al., 2017). Second, and most importantly, the detection of larger angular scale $\mathcal{B}$ modes, directly sourced by primordial gravitational waves, remains a key but elusive objective of modern cosmology due to their extremely low amplitude, with groundbreaking implications (e.g. Guzzetti et al., 2016; Kamionkowski and Kovetz, 2016).

A measurement of this primordial $\mathcal{B}$-mode signal on large angular scales would be regarded as a direct validation of the inflationary paradigm as the precursor of this stochastic background of gravitational waves. The amplitude of this background would directly constrain the energy scale of inflation, thereby ruling out some inflationary models, while also constraining the reionization period. In a nutshell, this would open a window to probe the physical mechanisms at work in the very early Universe, which remains at the frontiers of knowledge,[1] thereby dramatically improving our understanding of this obscure and mysterious period.

The detection of this faint primordial signal is hindered by various complexities, such as foreground contaminations, $\mathcal{B}$ modes induced by gravitational lensing of $\mathcal{E}$ modes, and non-trivial noise and systematics inherent to the instrument and observation strategy. In addition, there is a key $\mathcal{E}/\mathcal{B}$ leakage issue (e.g. Lewis, Challinor, and Turok, 2002; Bunn et al., 2003), whereby the $\mathcal{B}$ modes are contaminated by the relatively larger $\mathcal{E}$-mode power, which arises due to mode coupling introduced by incomplete sky coverage. To address this crucial aspect, in Part II, we develop a powerful statistical reconstruction tool which stringently ensures no such leakage contaminations, while accounting for anisotropic correlated noise.

Detailed mapping of cosmic matter distribution

The exploration of the Universe on large scales relies primarily on the use of large galaxy surveys, i.e. a compilation of the positions and optical properties of galaxies in the sky. These surveys are either photometric, when only wide band observations are available, or spectroscopic, for which the emission of each galaxy has been finely described at different wavelengths. From their respective luminous properties, we infer the galaxy redshift, i.e. its total apparent receding velocity.

[1] It is worth mentioning that spectral distortions in the CMB would provide a complementary and independent probe of early Universe physics (e.g. Chluba, 2014).

The statistics of the galaxy distribution measured by the above surveys have emerged as key tools to investigate various cosmological models and to improve our understanding of the late-time Universe through the growth and clustering of cosmic structures. In essence, distinct signatures of rich physics, encompassing quantum field theory and general relativity, are imprinted on the various processes of structure formation, such as the generation of the initial seed perturbations and dynamics of gravitational instability, during the course of cosmic history.

Ongoing and upcoming galaxy surveys will map the spatial distribution of galaxies in unprecedented volumes of the observable Universe. Some prominent examples include wide photometric surveys, such as the Dark Energy Survey (DES, The Dark Energy Survey Collaboration, 2005) and Large Synoptic Survey Telescope (LSST, Ivezic et al., 2008), and deep spectroscopic surveys, such as Dark Energy Spectroscopic Instrument (DESI, Levi et al., 2013) and Euclid (Laureijs et al., 2011; Racca et al., 2016; Amendola et al., 2016). Such state-of-the-art galaxy surveys are bringing us closer to a full census of the galaxy distribution in our patch of Universe. As such, this leads to a crucial consequence: We are running out of exploitable information on our Universe.

To make further progress in cosmology, we must inevitably transcend traditional data analysis approaches at the level of two- or three-point statistics of the galaxy distribution. In Part III, we employ a Bayesian forward modelling machinery for the joint inference of the underlying dark matter distribution and cosmological parameters from galaxy observations. Since we perform our analysis at the level of the field, the latter naturally encodes all the higher order statistics, resulting in a substantial information gain. We demonstrate that such an approach yields orders of magnitude improvement in the cosmological constraints over conventional methods. Moreover, we make use of recent advances in deep generative modelling to construct fast emulators of complex dynamics in Part IV, in preparation for future high-resolution analyses using our forward modelling framework.

Structure of this thesis

This thesis is written in three main parts, organized as follows. Part II focuses on the CMB polarization, with the key concepts underlying modern cosmology and physical mechanisms of CMB anisotropies, within the context of this thesis, outlined in Chapter 1. An introduction to the Wiener filter, as well as its associated numerical challenges, is provided in Chapter 2, followed by a description of the conceptual framework of our dual messenger algorithm. The applications of our novel statistical reconstruction tool to CMB temperature anisotropies and polarization are illustrated in Chapters 3 and 4, respectively. This part of the thesis culminates in the extension of our dual messenger algorithm to reconstruct pure $\mathcal{E}$ and $\mathcal{B}$ maps, while accounting for anisotropic correlated noise, as described in Chapter 5.

Part III involves the statistical reconstruction of the large-scale structures of the Universe within a Bayesian inference framework. Chapter 6 provides an introduction to the theory of cosmic structure formation, with an emphasis on Lagrangian perturbation theory and numerical simulations, as relevant to this thesis. The conceptual underpinnings of Bayesian probability theory are subsequently outlined in Chapter 7. Chapter 8 presents the development and implementation of the ALTAIR forward modelling machinery for cosmological parameter inference from galaxy redshift

surveys. We present an effective solution to the unknown foreground contamination problem in galaxy surveys via the development of a robust likelihood in Chapter 9.

Part IV describes the use of machine learning to construct fast complex dynamics emulators for Bayesian hierarchical models. Basic concepts underlying neural networks are introduced in Chapter 10, with Chapter 11 reviewing the rationale of the recently proposed generative modelling framework of adversarial networks. In Chapter 12, we present our halo painting model, encoding physically motivated neural networks, to map approximate dark matter simulations to 3D halo distributions.

Finally, in Part V, we summarize the key aspects of the work presented in this thesis. Prospective applications of the tools developed and possible avenues for future investigations are also laid out. Supplementary material is provided in the appendices, with Appendix A providing a mathematical description of the convergence properties of the standard messenger and dual messenger algorithms. Appendix B provides further details pertaining to spin field reconstruction and spherical harmonic transforms, as relevant to Chapters 4 and 5.

Part II

Polarization of the cosmic microwave background

Chapter 1

Cosmology and the cosmic microwave background

1.1 The ΛCDM model of the Universe

In this section, we review the theoretical foundations of modern cosmology, which have given rise to the so-called concordance (ΛCDM) model of the Universe. A comprehensive and detailed introduction to cosmology can be found in Peebles (1980), Kolb and Turner (1990), Peacock (1999), and Dodelson (2003).

1.1.1 The cosmological principle

The *cosmological principle* is the cornerstone of modern cosmology. This is the hypothesis that the Universe is spatially homogeneous and isotropic on sufficiently large scales (Weinberg and Dicke, 1973). Homogeneity implies that we have the same physical conditions at every point of any given hypersurface, while isotropy implies that we observe identical physical conditions in all spatial directions from a given point on the hypersurface (Mukhanov, 2005), i.e. there are no preferred directions in the Universe.

Observational evidence

The isotropy of the Universe is well-established by various cosmological observations, with the observed quasi-isotropic CMB radiation (Smoot et al., 1992; Dunkley et al., 2009; Planck Collaboration et al., 2018a, e.g.) providing the most remarkable quantitative evidence. There is, however, no consensus concerning the scales above which the Universe is homogeneous, with distinct definitions for the homogeneity scale proposed in the literature. Past studies of the mean density of galaxies in the recent galaxy surveys have yielded a homogeneity scale $\sim 70h^{-1}$ Mpc (e.g. Scrimgeour et al., 2012; Hogg et al., 2005), while investigations based on fractal nature of the Universe have led to varying results (e.g. Sylos Labini et al., 2009; Sarkar et al., 2009; Yadav, Bagla, and Khandai, 2010). Nevertheless, Alonso et al. (2015) constrained the fractal nature of the galaxy distribution from the Two Micron All Sky Survey Photometric Redshift (2MASS) survey (Bilicki et al., 2014) while relying only on the angular distances of galaxies and found no significant deviations from statistical homogeneity and isotropy in the large-scale structures of the Universe, in agreement with other complementary methods (e.g. Appleby and Shafieloo, 2014; Avila et al., 2018; Avila et al., 2019).

General relativistic Universe

The rationale underlying this simplifying hypothesis is encoded in the assumption of a spatially symmetric tensor $g_{\mu\nu}$, for any time slice, to describe the expansion dynamics of the Universe. Einstein's theory of general relativity describes the gravitational force as an emergent property of space and time, with the metric tensor characterizing the geometric relation between two events in a 4D spacetime manifold (Einstein, 1915). This metric allows us to calculate the separation $\mathrm{d}s$ between these events, separated by $\mathrm{d}x^\mu$, as $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$ and is used to solve Einstein's field equations (Carroll, 2004):

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{R}{2}g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} + \Lambda g_{\mu\nu}, \tag{1.1}$$

where $G_{\mu\nu}$ is the Einstein tensor, $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar, $T_{\mu\nu}$ is the energy-momentum tensor of the matter component, Λ denotes the cosmological constant, c is the speed of light in vacuum and G corresponds to the gravitational constant.

Friedmann-Lemâitre-Robertson-Walker metric

Under the assumption of the cosmological principle, the physical description of the Universe relies on the Friedmann-Lemaître-Robertson-Walker (FLRW) metric (Friedmann, 1922; Friedmann, 1924; Lemaître, 1927; Lemaître, 1931; Lemaître, 1933; Robertson, 1935; Robertson, 1936a; Robertson, 1936b; Walker, 1937):

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right), \tag{1.2}$$

where t is the cosmic time and $a(t)$ is the cosmic scale factor, normalized to unity at present time, i.e. $a_0 = 1$. $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin\theta^2\mathrm{d}\phi^2$ is the metric on the 2D sphere, with spherical comoving coordinates (r, θ, ϕ). k is the reduced curvature parameter, distinguishing between different geometries for spatial hypersurfaces: $k < 0$ (open Universe), $k = 0$ (flat Universe) and $k > 0$ (closed Universe). In essence, the cosmic scale factor describes the dynamical behaviour of the Universe. In Chapter 8, we constrain our cosmological model while relying purely on the geometrical symmetries of the cosmological principle.

Friedmann equations

Friedmann equations, derived from Einstein's equations of general relativity using the FLRW metric, encode the dynamics of the homogeneous Universe. They can be expressed as follows:

$$H^2 = \frac{8\pi G}{3c^4}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}, \tag{1.4}$$

where $H \equiv \dot{a}/a$ is the Hubble parameter,[1] ρ is the total energy density and P is the pressure.

[1] We denote differentiation with respect to t using a dot.

An immediate consequence of the Friedmann equations is the *continuity* equation,[2] which determines the evolution of the energy density:

$$\dot{\rho} = -3H(\rho + P). \tag{1.5}$$

Integrating the above equation leads to the description of the evolution of energy density for each component in the Universe: $\rho_i \propto a^{-3(1+w_i)}$, with a constant equation of state, $w_i \equiv P_i/\rho_i$. Radiation (relativistic particles such as photons) is characterized by $w = 1/3$, non-relativistic matter satisfies $w = 0$ and the cosmological constant interpreted as a fluid with $w = -1$. The above equation being linear, the time evolution of the energy density of each component can be solved independently from the other.

1.1.2 Constituents of the Universe

For each of the species from the above section, we can define their respective density parameters as the ratio of their energy density to the critical energy density, within a flat Universe, as $\Omega_i \equiv \rho_i/\rho_{\rm crit}$, where $\rho_{\rm crit} = 3H^2/8\pi G$ defines the critical point between an expanding and contracting Universe. This implies that if the sum of cosmological fluids is $\rho_{\rm crit}$, there is vanishing spatial curvature and all spatial hypersurfaces are flat. The first Friedmann equation can subsequently be expressed as:

$$\Omega_{\rm m} + \Omega_\Lambda + \Omega_{\rm r} + \Omega_{\rm k} = 1. \tag{1.6}$$

Using their respective equation of state and the conservation equation, we can express the evolution of the energy density of each component in terms of the scale factor:

$$\Omega_i(a) = \Omega_i^{(0)} \left(\frac{a}{a_0}\right)^{-3(1+w_i)} \left(\frac{H_0}{H}\right)^2, \tag{1.7}$$

where the present-day values of the Hubble and density parameters are denoted by H_0 and $\Omega_i^{(0)}$, respectively. Along with Eq. (1.6), this yields the familiar form of the Hubble function expressed in terms of the scale factor:

$$H^2(a) = H_0^2 \left(\Omega_\Lambda^{(0)} + \frac{\Omega_{\rm k}^{(0)}}{a^2} + \frac{\Omega_{\rm m}^{(0)}}{a^3} + \frac{\Omega_{\rm r}^{(0)}}{a^4}\right). \tag{1.8}$$

The ΛCDM scenario is based on a physical description of the gravitational force according to general relativity, requiring two key ingredients: A component of dark energy via the cosmological constant Λ in Einstein's equations to explain the accelerating cosmic expansion, and a component of cold (i.e. non-relativistic) dark matter (CDM) to explain the growth of cosmic structures and observed distribution of galaxies at present time. Hence, our standard cosmological model is known as the ΛCDM model and it is remarkably well-described using only six independent cosmological parameters: the physical baryon ($\Omega_{\rm b}$), dark matter ($\Omega_{\rm dm}$) and dark energy (Ω_Λ) densities corresponding to the energy content, and the scalar spectral index $n_{\rm s}$ and curvature fluctuation

[2] This is also a direct result of matter conservation, $\nabla_\mu T^{\mu\nu} = 0$, where $T^{\mu\nu}$ is the stress-energy tensor of a perfect fluid.

amplitude A_s pertaining to the initial power spectrum of fluctuations (cf. Section 1.4.4), and the reionization optical depth τ.

From the latest Planck observations (Planck Collaboration et al., 2018b), we have the following cosmological constraints: $\Omega_\Lambda \approx 0.68$, $\Omega_{dm} \approx 0.27$, $\Omega_b \approx 0.05$ for a Universe with zero spatial curvature, i.e. $\Omega_k = 0$. Hence, the matter content in the Universe is dominated by cold dark matter, which does not interact with light. Some dark matter candidates are weakly interacting massive particles such as neutrinos or neutralinos, and axions (e.g. Fornengo, 2016; Freese, 2017). The energy content of the Universe is dominated by dark energy, acting as a repulsive force that drives the acceleration of the cosmic expansion. Its nature is as yet unknown, with vacuum energy, dynamical scalar field and modified theories of gravity being some common hypotheses (e.g. Debono and Smoot, 2016).

1.1.3 Cosmological redshift and distance measures

As a consequence of the description of gravity as a geometric property of the spacetime manifold within general relativity, the notion of "distance" has no unique meaning in any arbitrary curved and non-stationary spacetimes, yielding distinct ways of specifying cosmological distances. Any given measure, nevertheless, is based on the separation of two spacetime events on radial null geodesics given by $ds^2 = -c^2dt^2 + a^2(t)dr^2 = 0$. As such, distances in general relativity are determined by the travelling times of photons emitted by a source at t_s and observed at t_o (e.g. Bartelmann and Schneider, 2001).

Due to the expansion of the Universe, the photons of a given wavelength λ_s emitted by a comoving light source at time t_s will be observed with redshifted wavelength λ_o at t_o. Photons, being massless particles, travel on null geodesics ($ds^2 = 0$) of zero proper time, implying that $c^2dt^2 = a^2(t)dq^2$, where dq is the line element on the spatial hypersurface. The *comoving distance* d_{com} between the source and observer can then be calculated as follows:

$$d_{com} = \int_0^{d_{com}} dq = \int_{t_s}^{t_o} dt \, \frac{c}{a(t)}. \tag{1.9}$$

By construction, the comoving distance between the source and observer is constant, such that the following relation holds:

$$\int_{t_s}^{t_o} dt \, \frac{c}{a(t)} = \int_{t_s+\lambda_o/c}^{t_o+\lambda_o/c} dt \, \frac{c}{a(t)}, \tag{1.10}$$

implying that photons emitted within a given time interval will have covered the same comoving distance. Assuming $a(t)$ to be approximately constant over these small time intervals, we obtain the following relation:

$$\frac{\lambda_o}{\lambda_s} = \frac{a(t_o)}{a(t_s)} \equiv 1 + z, \tag{1.11}$$

where the relative change in wavelength is given by $1 + z$, and z is known as the *cosmological redshift*.[3]

[3] This is the redshift due to the expansion of space and not due to the relative velocities of the source and observer.

The comoving distance can be expressed in terms of redshift as follows:

$$d_{\rm com}(z_o, z_s) \equiv \int_{r(z_s)}^{r(z_o)} \mathrm{d}r = \int_{t(z_s)}^{t(z_o)} \mathrm{d}t \, \frac{c}{a(t)} = \int_{a(z_s)}^{a(z_o)} \mathrm{d}a \, \frac{c}{a^2 H(a)} = \int_{z_s}^{z_o} \mathrm{d}z \, \frac{c}{H(z)}, \tag{1.12}$$

where $H(z)$ is obtained by combining the Hubble rate $H(a)$ from Eq. (1.8) with relation (1.11). The above *comoving - redshift* transformation is of special relevance to Chapter 8. The comoving distance factors out the expansion of the universe, such that it is not affected by the expansion of space. In contrast, the *proper distance* corresponds to a distance measurement at a specific cosmic time, which would be affected by cosmic expansion. Hence, at present time, the comoving distance is the proper distance. There are some other common distance measures in cosmology, such as angular and luminosity distances, but are not particularly pertinent to this thesis.

1.2 Cosmic inflationary paradigm

We present a concise description of the early Universe physics which give rise to the initial conditions for the evolution of cosmic structures described in Chapter 6, and whose signatures are imprinted on the CMB.

1.2.1 Generation of cosmological fluctuations

The inflationary and Hot Big Bang scenario constitute our contemporary physical model for the cosmic initial conditions, strongly favoured by current observations. An in-depth review of the inflationary paradigm is provided in Baumann (2011). Cosmic inflation accounts for some shortcomings in the standard Hot Big Bang model, such as the observed statistical homogeneity and isotropy of the Universe and the horizon problem (Guth, 1981; Linde, 1982; Albrecht and Steinhardt, 1982). During the inflationary epoch, the cosmological equation of state was governed by the *inflaton field*, which is a potential-dominated quantum scalar field with negative pressure, resulting in an exponential growth of the cosmic scale factor. In this scenario, the accelerated expansion of the very early Universe magnified the microscopic vacuum fluctuations of the inflaton field into macroscopic cosmological perturbations. These initial quantum fluctuations generated during the inflationary phase led to inhomogeneities in the matter density (cf. Section 6.2.1) and in the CMB, as discussed in the following section.

1.2.2 Imprints of early Universe physics

According to the Hot Big Bang scenario (e.g. Kolb and Turner, 1990; Peacock, 1999), the Universe was in an initial hot and dense state, composed of a primordial plasma of photons, electrons, protons in thermal equilibrium. Thomson scattering of the photons with the free electrons caused the Universe to be opaque. The expansion of the Universe led to a gradual decrease in temperature, and when sufficiently low at $T \sim 3000$ K, protons and electrons could combine to form the lightest atoms in an epoch known as *recombination*, resulting in photon-matter decoupling. Consequently, the photons could propagate freely to produce the *cosmic microwave background* emission and the surface of *last scattering*. In a nutshell, the CMB is the remnant light from the earliest moments of

the Universe and its detection for the first time by Penzias and Wilson (1965) provided substantial evidence in favour of the Hot Big Bang model and ushered in a new era in cosmology.

Temperature anisotropies

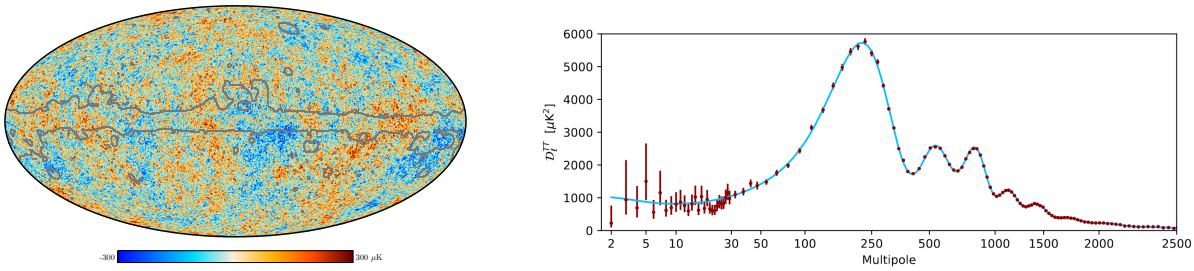
Various signatures of early Universe physics are imprinted on the CMB radiation, both on its temperature and polarization, as a result of the *scalar* and *tensor* perturbations generated during the inflationary epoch. The scalar modes correspond to perturbations in the energy density of cosmological fluid at the time of last scattering. These are the only fluctuations which may lead to structure formation via gravitational instability (Hu and White, 1997) and therefore, are the seeds of structure formation. In overdense regions, gravitational forces overwhelm pressure, causing the matter to flow towards the density peaks. The energy of photons is reduced as they climb out of such potential wells, implying that overdense regions are effectively initially cold. Conversely, underdense regions are hot, and these temperature fluctuations related to the density fluctuations of the baryons at recombination are encoded in the quasi isotropic CMB radiation. The temperature anisotropies of the CMB are discussed further in Section 1.3.

Acoustic peaks

Before recombination, as the primordial plasma cools down, the matter gradually starts to dominate over the radiation, leading to gravitational collapse of overdense regions and subsequent growth of the density perturbations. The density of such regions increases, but so does the radiation pressure. Nevertheless, due to the baryons being strongly coupled to radiation, the amplitude of the density fluctuations decreases, causing the temperature to drop. This rise in radiation pressure can then withstand the tendency towards collapse. This proceeds in a cyclic scheme, yielding the so-called *acoustic oscillations*. This oscillatory phase of the density perturbations manifest as peaks in the power spectrum of the temperature anisotropies (cf. right panel of Fig. 1.1). In physical terms, these oscillations correspond to the heating and cooling of the photon-baryon fluid, with compression and rarefaction induced by an acoustic or standing sound wave.

Baryon acoustic oscillations

After recombination and matter-radiation decoupling, the above oscillatory mechanism ceases, and baryons can collapse in potential wells generated by (non-baryonic) dark matter. In this case, there is no oscillation since dark matter does not couple to radiation and this eventually leads to the growth of cosmic large-scale structures, clusters of galaxies and galaxies. Moreover, the acoustic waves mentioned above are also imprinted weakly on the clustering of galaxies and matter (Peebles and Yu, 1970; Bond and Efstathiou, 1984), as observed by galaxy redshift surveys. These features, which manifest themselves as a small excess in number of pairs of galaxies separated by a given scale, are known as *baryon acoustic oscillations* (BAOs). They were first measured via the correlation function of luminous red galaxies by Eisenstein et al. (2005), with the BAO scale being roughly 150 Mpc in radius at recombination.



(A) Map of temperature anisotropies

(B) Angular power spectrum of temperature fluctuations

FIGURE 1.1: *Left panel:* Map of the temperature anisotropies of the CMB, as observed by the Planck satellite (Planck Collaboration et al., 2018a). The gray outline indicates the extent of the confidence mask, which takes into account the galactic emission from the Milky Way. Note that in standard CMB data analysis pipelines, the masked regions are artificially augmented using *inpainting* techniques (e.g. Starck, Fadili, and Rassat, 2013). *Right panel:* Angular power spectra of the CMB temperature anisotropies reconstructed from the Planck observations (Planck Collaboration et al., 2018a). The blue line indicates the best fit of the standard model of cosmology to the Planck data.

Primordial gravitational waves

Tensor modes can be viewed as gravitational waves, and mathematically correspond to transverse-traceless perturbations to the metric. Such a plane gravitational wave would distort the space in the plane of the fluctuation, resulting in a quadrupolar stretching pattern of space. As a consequence of this spatial distortion induced by gravity waves, the CMB radiation acquires a quadrupole moment, with Thomson scattering subsequently generating a polarization pattern from this quadrupole anisotropy (Hu and White, 1997). These polarized anisotropies of the CMB are discussed in Section 1.4. Quantum fluctuations from the inflationary era should, therefore, result in primordial gravitational waves, which would be imprinted on the CMB polarization maps. These ripples in the space-time fabric constitute a major prediction of the inflationary paradigm. The detection of such gravity waves, as yet elusive by virtue of their extremely low amplitude, is the ultimate goal of next-generation CMB experiments. The origin of CMB polarization and their key significance are detailed in Sections 1.4.1 and 1.4.4, respectively.

1.3 CMB temperature anisotropies

In this section, we provide a statistical description of the temperature anisotropies of the CMB and outline their remarkable contribution in constraining our standard cosmological model.

1.3.1 Statistics of temperature fluctuations

Due to the photons being tightly coupled to the matter before recombination, the CMB has a nearly uniform blackbody spectrum, with inhomogeneities corresponding to the hot and cold spots. These temperature variations, of order 10^{-5}, were first detected by COsmic microwave Background Explorer (COBE) (Smoot et al., 1992), with more refined measurements eventually by the Wilkinson Microwave Anisotropy Probe (WMAP) (e.g. Dunkley et al., 2009) and Planck satellites (e.g. Planck

Collaboration et al., 2018a). The map of CMB temperature anisotropies from the final release of the Planck mission (Planck Collaboration et al., 2018a) is depicted in the left panel of Fig. 1.1.

The spectrum of the temperature anisotropies is measured as a function of angular scale in the sky. The temperature field can be decomposed into spherical harmonic modes on the sky as follows:

$$\Delta T(\hat{\boldsymbol{n}}) = \sum_{\ell,m} a^{\mathcal{T}}_{\ell m} Y_{\ell m}(\hat{\boldsymbol{n}}), \tag{1.13}$$

where $\Delta T(\hat{\boldsymbol{n}})$ denotes the CMB temperature relative to the mean value ($T = 2.725$ K) at a direction $\hat{\boldsymbol{n}}$ in the sky, and the value of ℓ indicates a particular angular scale, i.e. inverse of the scale between two points on the sky, while m denotes the orientation of the modes on the sphere. As such, large angular scales on the celestial sphere correspond to the small values of ℓ. The spherical harmonic coefficients encode all the information on the amplitude of the fluctuations for a given scale.

Under the assumption of statistical isotropy, there is no dependence on m, such that each harmonic coefficient $a^{\mathcal{T}}_{\ell m}$, for a given scale ℓ, has the following variance:

$$\langle a^{\mathcal{T}}_{\ell m} a^{*\mathcal{T}}_{\ell' m'} \rangle = \delta_{\ell\ell'}\delta_{mm'} C^{TT}_{\ell}, \tag{1.14}$$

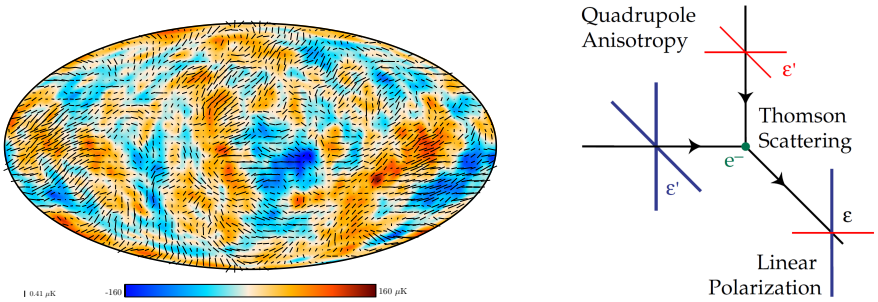
where C^{TT}_{ℓ} corresponds to the *angular power spectrum* of the temperature anisotropies, and $\langle \ldots \rangle$ denotes the average over all possible realizations of the Universe. However, since we have only one observable Universe and there are only $(2\ell + 1)$ modes for each scale ℓ, this definition implies that we have only a finite amount of information to sample the $a_{\ell m}$ distribution, leading to a fundamental and inherent uncertainty in the measurement of the C_ℓ's. This uncertainty, known as the *cosmic variance*, is given by:

$$\left(\frac{\Delta C_\ell}{C_\ell}\right) = \sqrt{\frac{2}{2\ell + 1}}, \tag{1.15}$$

implying that this variance is larger for low ℓ, i.e. on large angular scales. The right panel of Fig. 1.1 displays the angular power spectrum of the temperature anisotropies obtained from the final release of the Planck mission (Planck Collaboration et al., 2018a). The Planck measurements correspond to the red dots with the associated error bars accounting for both measurement errors and the cosmic variance. The blue line represents the best fit of the ΛCDM model with the observations, and this remarkable agreement constitutes one of the key successes of our standard model of cosmology as the currently accepted scenario for the origin and evolution of our Universe. The Planck satellite has mapped the temperature anisotropies to a precision limited only by the cosmic variance, and as such, provides the ultimate state-of-the-art measurement, thereby mostly exhausting their information content pertaining to primordial fluctuations. Future CMB experiments, therefore, aim to optimally exploit the cosmological information encoded in the polarized anisotropies.

1.3.2 Primary and secondary anisotropies

The division of the CMB anisotropies into *primary* and *secondary* anisotropies is of particular interest. The former emerge from the properties of the primordial photon-baryon fluid before recombination. The amplitude and characteristics of these temperature fluctuations are governed by a



(A) Map of CMB polarization (B) Thomson scattering

FIGURE 1.2: *Left panel:* Map of the polarization anisotropies of the CMB, as observed by the Planck satellite (Planck Collaboration et al., 2018a). Temperature is depicted as colour and polarization as a vector field, where the orientation and length of the rods indicate the direction and amplitude, respectively. *Right panel:* Origin of CMB polarization (Hu and White, 1997). Thomson scattering of radiation with a quadrupole anisotropy off a free electron, thereby inducing linear polarization.

series of physical processes such as the Doppler effect or the Sachs-Wolfe effect. A comprehensive description of the primary anisotropies is provided in Hu and White (1997).

As the CMB photons travel from the surface of last scattering to our instruments, they experience distinct cosmological phenomena such as reionization or the gravitational potential of the cosmic large-scale structures, which alter the photons' energy. To exploit the primordial information encoded in the CMB, these effects (for e.g. Sunyaev-Zel'dovich effect or gravitational lensing), must be accounted for. Moreover, they are a powerful probe of the late-time Universe. Challinor and Peiris (2009) review in depth such *secondary* anisotropies.

1.4 CMB polarization

We review the origins and mathematical description of the CMB polarization, and subsequently outline their significance and the state-of-the-art pertaining to the observations of the polarized CMB.

1.4.1 Origins via Thomson scattering

In addition to the temperature anisotropies, the CMB radiation is also polarized, with the associated power spectrum being around two orders of magnitude lower than its temperature counterpart. This polarization was first detected in 2002 by the Degree Angular Scale Interferometer (DASI) (Kovac et al., 2002). The CMB polarization anisotropies resulted from the Thomson scattering of the anisotropic photon distribution off free electrons during the recombination and reionization epochs (Hu and White, 1997; Kamionkowski and Kovetz, 2016), with the left panel of Fig. 1.2 depicting the map from the final Planck release (Planck Collaboration et al., 2018a). Polarization is indicated via rods whose orientation and length represent the direction and amplitude, respectively, of the polarized field. The coloured background is the map of temperature anisotropies, smoothed with a Gaussian kernel.

If we consider a Cartesian frame, where a photon is travelling with a wavevector parallel to the $\hat{x}$ axis, then its transverse electric field would cause the electron to oscillate along the $\hat{y}$ and $\hat{z}$ axes. Since the photons in the primordial plasma are incident from all directions, their respective contributions average to zero, such that there is no net polarization. Hence, Thomson scattering of isotropic incoming radiation off an electron would not select any particular polarization direction. However, if the temperature anisotropy has a quadrupolar pattern, Thomson scattering may induce a net linear polarization, as illustrated in the right panel of Fig. 1.2, where the two perpendicular components have different temperature. By symmetry, Thomson scattering does not generate any circular polarization. The relatively low power of CMB polarization is due to the suppression of the quadrupole moments by Compton scattering.

1.4.2 Stokes parameters

Linear polarization can be described formally using the $(\mathcal{Q},\mathcal{U})$ Stokes parameters, and the spherical harmonic coefficients $a^{\mathcal{E}}_{\ell m}$ and $a^{\mathcal{B}}_{\ell m}$, analogous to the temperature anisotropies:

$$\big(\mathcal{Q}\pm i\mathcal{U}\big)(\hat{\boldsymbol{n}}) = -\sum_{\ell,m}\big(a^{\mathcal{E}}_{\ell m}\pm i a^{\mathcal{B}}_{\ell m}\big)\,{}_{\pm2}Y_{\ell m}(\hat{\boldsymbol{n}}), \tag{1.16}$$

where the above spin-weighted harmonics are due to the fact that $(\mathcal{Q},\mathcal{U})$ transform as a spin-2 field under rotation by an angle ψ of the plane of the detector, tangent to the sphere (Hu and White, 1997; Seljak and Zaldarriaga, 1997; Kamionkowski, Kosowsky, and Stebbins, 1997b; Zaldarriaga and Seljak, 1997), as follows:

$$\big(\mathcal{Q}\pm i\mathcal{U}\big)(\hat{\boldsymbol{n}}) \to e^{\mp 2i\psi}\big(\mathcal{Q}\pm i\mathcal{U}\big)(\hat{\boldsymbol{n}}). \tag{1.17}$$

The $\mathcal{I}$ Stokes parameter is usually taken to be the intensity of temperature anisotropies, while the fourth Stokes parameter $\mathcal{V}$ describing circular polarization is not considered to be of relevance since it cannot be generated via Thomson scattering.

1.4.3 $\mathcal{E}/\mathcal{B}$ decomposition of CMB polarization

In analogy with mathematical properties of electromagnetic fields, we can introduce two new quantities, a curl-free component $\mathcal{E}$ and a divergence-free component $\mathcal{B}$ of the polarization vector, as rotationally invariant alternatives to $\mathcal{Q}$ and $\mathcal{U}$. Analogous to Eq. (1.14), the power spectra of the $\mathcal{E}$ and $\mathcal{B}$ modes are defined as:

$$\langle a^{\mathcal{E}}_{\ell m} a^{*\mathcal{E}}_{\ell' m'}\rangle = \delta_{\ell\ell'}\delta_{mm'}C^{EE}_{\ell} \tag{1.18}$$

$$\langle a^{\mathcal{B}}_{\ell m} a^{*\mathcal{B}}_{\ell' m'}\rangle = \delta_{\ell\ell'}\delta_{mm'}C^{BB}_{\ell}. \tag{1.19}$$

$\mathcal{E}$ is a scalar field with even parity, while $\mathcal{B}$ is pseudo-scalar field with odd parity, i.e. $a^{\mathcal{E}}_{\ell m}\to(-1)^{\ell}a^{\mathcal{E}}_{\ell m}$ and $a^{\mathcal{B}}_{\ell m}\to(-1)^{\ell+1}a^{\mathcal{B}}_{\ell m}$, respectively, under a reverse parity ($\hat{\boldsymbol{n}}\to-\hat{\boldsymbol{n}}$) transformation. As such, the cross power spectrum C^{EB}_{ℓ} vanishes, $\langle a^{\mathcal{E}}_{\ell m} a^{*\mathcal{B}}_{\ell' m'}\rangle = 0$. This also leads to distinct polarization patterns for the $\mathcal{E}$ and $\mathcal{B}$ components, as depicted in Fig. 1.3. $\mathcal{E}$ modes have tangential and radial

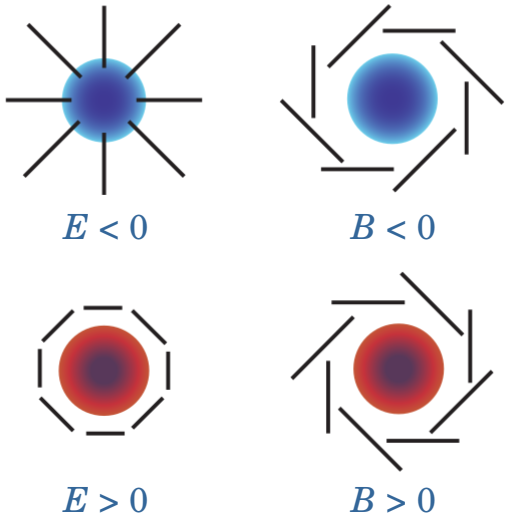


FIGURE 1.3: Characteristic polarization patterns of $\mathcal{E}$ and $\mathcal{B}$ modes. The polarization directions of the $\mathcal{E}$ field are tangential (radial) around hot (cold) spots, indicated by red (blue) regions, whilst $\mathcal{B}$ modes have swirling patterns with different orientations around hot and cold spots.

polarization directions around hot and cold spots, respectively. $\mathcal{B}$ modes, in contrast, have a characteristic swirling pattern, with opposite directions around hot and cold spots (Kamionkowski and Kovetz, 2016). An in-depth derivation of the harmonics of the scalar fields deduced from the (Q, U) Stokes parameters is provided in Zaldarriaga and Seljak (1997).

The separation of CMB polarization into the $\mathcal{E}$ and $\mathcal{B}$ modes is essential for cosmological inference from observations of the polarization anisotropies. Unfortunately, this $\mathcal{E}/\mathcal{B}$ decomposition, as encapsulated in Eq. (1.16), is non-local when the sky is partially observed, such that the relatively larger $\mathcal{E}$-mode power corrupts the observed $\mathcal{B}$ modes. This is commonly referred to as the $\mathcal{E}/\mathcal{B}$ *leakage* or *coupling* issue.[4] Since neither current nor forthcoming CMB experiments will provide full sky coverage, ensuring no cross contamination between these two components is a fundamental requirement of any data analysis pipeline aiming to detect the primordial gravitational waves.

While a unique decomposition of the polarization field into $\mathcal{E}$ and $\mathcal{B}$ modes is not possible in a finite patch of the sky, the signal vector space can be uniquely decomposed into three orthogonal subspaces: *pure* $\mathcal{E}$, *pure* $\mathcal{B}$ and *ambiguous* subspaces (Lewis, Challinor, and Turok, 2002; Bunn et al., 2003; Smith, 2006a; Zhao and Baskaran, 2010; Kim and Naselsky, 2010; Bunn, 2011). Mathematically, we can define the $\mathcal{E}$- and $\mathcal{B}$-mode spherical harmonic vectors as ${}_s\mathbf{Y}^{\mathcal{E},\mathcal{B}}_{\ell m} \equiv \mathbf{D}^{\mathcal{E},\mathcal{B}}_s Y_{\ell m}$, where $\mathbf{D}^{\mathcal{E},\mathcal{B}}_s$ are the respective differential operators described in Appendix B.3. They are the spin-2 analogues of divergence and curl, respectively. Here, $s = 2$ since we are dealing with spin-2 fields. We can thus define a polarization field f as a $\mathcal{B}$ mode if $(\mathbf{D}^{\mathcal{E}}_s)^\dagger \cdot f = 0$. Polarization fields living on the manifold Ω form a normed vector space with the inner product:

$$(f, g) \equiv \int_\Omega \mathrm{d}\Omega \; f^\dagger \cdot g, \tag{1.20}$$

[4]Note that the leakage between the polarized and unpolarized observational channels due to instrumental effects, such as asymmetric beams, pair differencing and polarization angle calibration, is a completely distinct effect.

such that two fields f and g are orthogonal if $(f, g) = 0$. A pure $\mathcal{B}$ mode, therefore, is defined as the field orthogonal to all $\mathcal{E}$ modes. The differential operators satisfy the relations: $\mathbf{D}_s^{\mathcal{E}\dagger} \cdot \mathbf{D}_s^{\mathcal{B}} = \mathbf{D}_s^{\mathcal{B}\dagger} \cdot \mathbf{D}_s^{\mathcal{E}} = 0$, such that $\mathcal{B}$ modes can always be generated via the application of $\mathbf{D}_s^{\mathcal{B}}$ to a polarization field. We can subsequently show that $\mathcal{E}$ and $\mathcal{B}$ modes, defined in this way, are orthogonal in the full sky without boundary:

$$(\mathbf{D}_s^{\mathcal{E}}\varphi, \mathbf{D}_s^{\mathcal{B}}\psi) = \int_\Omega \mathrm{d}\Omega\, (\mathbf{D}_s^{\mathcal{E}}\varphi)^\dagger \cdot \mathbf{D}_s^{\mathcal{B}}\psi = 0, \tag{1.21}$$

for arbitrary scalar fields φ and ψ, using integration by parts. For spin-2 fields on an incomplete sky, the above equation holds only if the Dirichlet and Newmann boundary conditions are satisfied simultaneously, i.e. the field and its first derivative vanishes on the boundary $\partial\Omega$ of the observed region, respectively:

$$\psi|_{\partial\Omega} = \hat{\boldsymbol{n}} \cdot \nabla\psi|_{\partial\Omega} = 0, \tag{1.22}$$

where $\hat{\boldsymbol{n}}$ is the normal to the boundary. This allows us to construct a complete orthonormal basis for all square-integrable polarization fields in a given sky region Ω (cf. Section III in Bunn et al., 2003, for complementary information) in terms of pure $\mathcal{E}$ (curl-free), pure $\mathcal{B}$ (divergence-free) and ambiguous (both curl-free and divergence-free) modes, with the latter component lying in the subspace orthogonal to both pure subspaces. As such, we can decompose unambiguously any polarization field into three components, thereby projecting out the ambiguous modes resulting from the partially observed or masked sky. Further mathematical description of $\mathcal{E}/\mathcal{B}$ decomposition is provided in Chapter 5.4.

1.4.4 Physical interpretations of $\mathcal{E}$ and $\mathcal{B}$ modes

The $\mathcal{E}/\mathcal{B}$ decomposition, as described in Section 1.4.3, allows us to distinguish the physical origins of the geometrical patterns illustrated in Fig. 1.3. Scalar (or density) perturbations can only generate symmetric polarization patterns, i.e. accounting only for the $\mathcal{E}$ modes, whilst gravitational waves can induce both $\mathcal{E}$ and $\mathcal{B}$ modes, thereby leading to a non-vanishing $\mathcal{B}$-mode power spectrum. As such, $\mathcal{B}$ modes are solely a signature of tensor perturbations. It is important to note, however, that gravitational lensing distorts $\mathcal{E}$ modes such that they behave like $\mathcal{B}$ modes, on the small angular scales (Zaldarriaga and Seljak, 1998; Seljak and Hirata, 2004). These lensed $\mathcal{B}$ modes, which manifest on the small angular scales, are distinct from the primordial ones on the large scales.

While the gravitational waves also induce temperature and $\mathcal{E}$ modes, the contribution of the tensor perturbations to their respective power spectra is negligible as it is overwhelmed by the contribution from scalar perturbations. As a result, the primordial $\mathcal{B}$-mode power spectrum is the key quantity to probe the extremely early Universe (e.g. Guzzetti et al., 2016). In particular, the amplitude of the $\mathcal{B}$-mode signal, emanating from the gravity waves, determines the characteristics of the inflationary phase, thereby encoding the physics of the earliest observable period of our Universe.

The primordial power spectra of the scalar[5] and tensor perturbations are parameterized as follows:

$$P_{\rm s}(k) = A_{\rm s}\left(\frac{k}{k_0}\right)^{n_{\rm s}-1} \tag{1.23}$$

$$P_{\rm t}(k) = A_{\rm t}\left(\frac{k}{k_0}\right)^{n_{\rm t}}, \tag{1.24}$$

where $A_{\rm s}$ ($A_{\rm t}$) and $n_{\rm s}$ ($n_{\rm t}$) correspond to the amplitude and spectral index, respectively, of the scalar (tensor) modes, and k_0 is an arbitrary scale at which the perturbation amplitude is evaluated and is usually referred to as the *pivot* scale. The primordial tensor spectrum is commonly expressed via the *tensor-to-scalar* ratio denoted by r. This quantity characterizes the ratio of the tensor to scalar mode power generated by cosmic inflation at the pivot scale:

$$r \equiv \frac{A_{\rm t}(k_0)}{A_{\rm s}(k_0)}. \tag{1.25}$$

The above power spectrum parameters can be related to inflationary parameters. Hence, r is a direct measure of the energy scale of inflation (e.g. Baumann, 2011):

$$V^{\frac{1}{4}} \sim \left(\frac{r}{0.01}\right)^{\frac{1}{4}} 10^{16}\ \text{GeV}, \tag{1.26}$$

where values of $r \geq 0.01$ would correspond to inflation occurring at the *Grand Unified Theory* (GUT) scale. The extremely high energy scale in the prefactor implies that measurements of the primordial fluctuations may shed light on high energy physics in a regime far beyond the standard model of particle physics. Current joint experimental constraints on r, from Planck and BICEP2/Keck array, have set an upper bound of $r < 0.06$ at 95% confidence level (BICEP2 Collaboration et al., 2018). In a nutshell, the measurement of r would allow us to test and validate distinct inflationary scenarios with different theoretically predicted values of r.

1.4.5 Detection of CMB polarization

The first detection of the $\mathcal{E}$ modes by the Degree Angular Scale Interferometer (DASI, Kovac et al., 2002) marked the onset of the CMB polarization era. Consequently, this provided significant encouragement for further investment in the instrumental and data analysis aspects to cope with such challenging observations. Several CMB experiments were designed to estimate the $\mathcal{E}$-mode power spectrum. A few years ago, the ACTPol collaboration accurately reconstructed the power spectra of the temperature and $\mathcal{E}$ modes, and their cross spectra C_ℓ^{TE} on small scales (Naess et al., 2014).

The main interest of CMB research subsequently switched to the detection of $\mathcal{B}$ modes, with a series of ground-based, spatial or balloon experiments being designed. The first direct detections of $\mathcal{B}$ modes, corresponding to the lensed $\mathcal{B}$ modes on small angular scales, were recently made by the POLARBEAR experiment (Ade et al., 2014) and the South Pole Telescope (SPT, Hanson et al., 2013),

[5]This is the dimensionless power spectrum for curvature fluctuations, where $n_{\rm s} = 1$ corresponds to a scale-invariant (Harrison-Zel'dovich) spectrum (Harrison, 1970; Zeldovich, 1972).

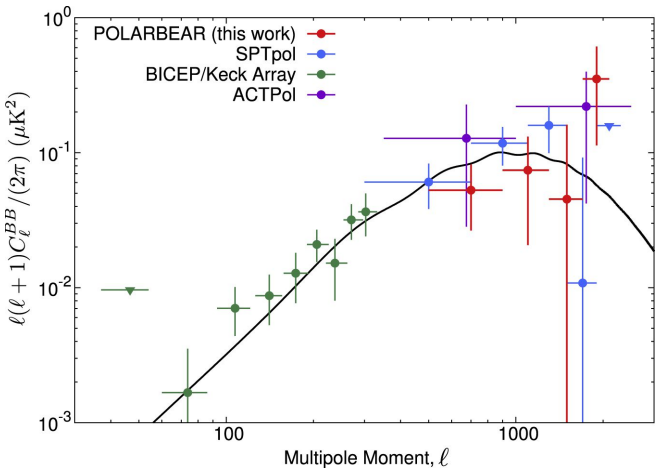


FIGURE 1.4: $\mathcal{B}$-mode polarization power spectrum measurements by several CMB experiments, as labelled on the figure (The POLARBEAR Collaboration et al., 2017). The black curve indicates the theoretical ΛCDM spectrum obtained with the cosmological parameters from the Planck Collaboration et al. (2016b) release.

followed by measurements from the Atacama Cosmology Telescope (ACTPol, Louis et al., 2017), as illustrated in Fig. 1.4. As such, the focus shifted to the detection of $\mathcal{B}$ modes, sourced by primordial gravitational waves, on large angular scales. In 2014, the BICEP2 collaboration claimed the groundbreaking detection of this primordial signal (BICEP2 Collaboration et al., 2014), but this discovery was controversial due to potential foreground contaminations (e.g. Flauger, Hill, and Spergel, 2014; Mortonson and Seljak, 2014). The Planck data then played a crucial role, with a detailed study of the polarized dust over the sky (Planck Collaboration et al., 2016e) and a subsequent joint Planck-BICEP2 analysis revealing that the excess of power found by the BICEP2 team was primarily due to polarized dust contamination, with no statistically significant evidence for primordial $\mathcal{B}$ modes (BICEP2/Keck Collaboration et al., 2015).

With the first conclusive detection of the primordial $\mathcal{B}$ modes as yet unaccomplished, various CMB experiments with unprecedented sensitivity, tailored for the detection of this extremely weak signal, are currently under construction. Some examples include the CMB-S4 (Abazajian et al., 2016) and Simons Observatory (Ade et al., 2019).

Chapter 2

The Wiener filter

This chapter draws from Kodi Ramanah, Lavaux, and Wandelt (2017), Kodi Ramanah, Lavaux, and Wandelt (2018) and Kodi Ramanah, Lavaux, and Wandelt (2019).

2.1 The problem

One of the most frequently encountered and ubiquitous problems in astrophysics and cosmology (and many other fields in science) is signal reconstruction from noisy data. Solutions to this problem have therefore been researched extensively for the last two centuries (e.g. Gauss, 1809; Jaynes, 1957; Kalman, 1960). The Wiener filter (Wiener, 1949) has emerged as a standard tool for the analysis of large data sets for the inference of high-dimensional signals, such as the large-scale structure and CMB analyses. Hence, it has been employed for a wide range of applications in cosmology and astrophysics, as described in Section 2.3 below.

The Wiener filter originates from the following statistical problem. We assume our observed data set $\boldsymbol{d}$ to be a linear combination of the signal $\boldsymbol{s}$ with covariance $\mathbf{S}$ and noise $\boldsymbol{n}$ with covariance $\mathbf{N}$, as follows:

$$\boldsymbol{d} = \boldsymbol{\mathcal{R}}\boldsymbol{s} + \boldsymbol{n}, \tag{2.1}$$

where the signal and noise covariances are given by $\mathbf{S} = \langle \boldsymbol{s}\boldsymbol{s}^{\dagger} \rangle$ and $\mathbf{N} = \langle \boldsymbol{n}\boldsymbol{n}^{\dagger} \rangle$, respectively. $\boldsymbol{\mathcal{R}}$ is the complete response operator, with the inclusion of harmonic transforms, beam and mask effects, that models the instrument response to incoming signal. It effectively corresponds to the overall model of how the instrument converts, on average, an incoming signal $\boldsymbol{s}$ to the observed data $\boldsymbol{d}$, with the residual being the noise $\boldsymbol{n}$, while encoding the relevant physics.

The Wiener filter solution, $\boldsymbol{s}_{\mathrm{WF}}$, is the optimal linear filter when the signal and noise are both Gaussian random fields. For a particular realization of the data, $\boldsymbol{s}_{\mathrm{WF}}$ therefore maximizes the posterior probability distribution $\propto \exp(-\chi^2/2)$, or equivalently minimizes:

$$\chi^2 = (\boldsymbol{d} - \boldsymbol{\mathcal{R}}\boldsymbol{s})^{\dagger}\mathbf{N}^{-1}(\boldsymbol{d} - \boldsymbol{\mathcal{R}}\boldsymbol{s}) + \boldsymbol{s}^{\dagger}\mathbf{S}^{-1}\boldsymbol{s}, \tag{2.2}$$

leading to the Wiener filter equation,

$$\boldsymbol{s}_{\mathrm{WF}} = (\mathbf{S}^{-1} + \boldsymbol{\mathcal{R}}^{\dagger}\mathbf{N}^{-1}\boldsymbol{\mathcal{R}})^{-1}\boldsymbol{\mathcal{R}}^{\dagger}\mathbf{N}^{-1}\boldsymbol{d}. \tag{2.3}$$

$\boldsymbol{s}_{\mathrm{WF}}$ is the least-square optimal solution: the Wiener filter minimizes the mean-square deviations $\langle \boldsymbol{\varepsilon}^{\dagger}\boldsymbol{\varepsilon} \rangle$ of the reconstruction errors $\boldsymbol{\varepsilon} = \boldsymbol{s}_{\mathrm{WF}} - \boldsymbol{s}$, averaged over all signal and noise realizations.

2.2 Computational challenges

Computing the Wiener filter solution for large and complex data sets from modern experiments is numerically challenging. As is evident from Eq. (2.3), the direct numerical implementation of the Wiener filter requires inversion of dense matrices. This task is rendered intractable by the size of modern data sets from state-of-the-art experiments, and therefore represents a computational bottleneck. While it would be extremely convenient if there existed a common basis set, easily accessible by fast transforms, where both $\mathbf{S}$ and $\mathbf{N}$ are sparse, this is often not the case as for instance, the signal and noise covariances may be sparse in Fourier or pixel space, respectively. As such, the first matrix inversion above is dense in all bases and lives in a high-dimensional space. This space has typically the size of the number of elements in $\boldsymbol{d}$, which for Planck maps is $\mathcal{O}(10^9)$, when accounting for polarization components and the nine frequency bands. Due to the size of the covariance matrices scaling as the square of the number of data samples, the storage and processing of dense systems become numerically intractable.

Previous approaches relied on the assumption of a homogeneous and isotropic noise distribution to find approximate solutions to the Wiener filter equation (e.g. Hirata et al., 2004; Komatsu, Spergel, and Wandelt, 2005; Mangilli and Verde, 2009). Traditional approaches of computing the Wiener filter rely on costly and highly non-trivial numerical schemes like Krylov space methods, such as preconditioned conjugate gradient (PCG) methods (Eriksen et al., 2004; Wandelt, Larson, and Lakshminarayanan, 2004; Smith, Zahn, and Doré, 2007; Seljebotn et al., 2014; Seljebotn et al., 2019; Puglisi et al., 2018; Papež, Grigori, and Stompor, 2018), requiring a preconditioner to approximate the matrix inversion involved. A basic description of the PCG method is provided in Section 2.4. These complex techniques, however, suffer from various numerical limitations when dealing with high-dimensional data sets. They require intricate and costly numerical schemes such as preconditioning the linear system with suitable matrices and subsequently demands significant investment in software development. Moreover, this approach involves a further inherent stumbling block as finding efficient preconditioners is a complicated task in itself (Oh, Spergel, and Hinshaw, 1999) since matrices are often extremely ill-conditioned in typical CMB problems (Eriksen et al., 2004). The inclusion of polarization data in the analysis further exacerbates this predicament (Larson et al., 2007).

Standard schemes based on the PCG method involved the use of a combination of block and diagonal preconditioners on large and small angular scales, respectively (Eriksen et al., 2004), or were based on a recursive algorithm, where the conjugate gradient solution on a coarse grid is adopted as the preconditioner on a finer grid (Smith, Zahn, and Doré, 2007). Although both of these approaches yielded satisfactory performance with the analysis of WMAP data (e.g. Eriksen et al., 2008), they were found to be too computationally intensive for the high-resolution and sensitivity analysis of Planck data. Seljebotn et al. (2014) developed a multi-level solver for Gaussian constrained realizations of the CMB, which is fast once a suitable preconditioner is found but requires careful tuning and costly precomputations in terms of both computing power and memory requirements. This approach is therefore less attractive when we have to solve many different systems, each requiring a specific preconditioner. A pseudo-inverse based preconditioner was recently proposed by Seljebotn et al. (2019) for CMB component separation. The complexity of this method is illustrated by

the variety of preconditioners that exist in the literature. Such PCG methods do have some merits, however, as they are conducive to fast convergence, provided an adequate preconditioner tailored to the specific problem is available.

Various alternatives to the PCG methods have been proposed in the literature. A recent work by Horowitz, Seljak, and Aslanyan (2018) was based on recasting the Wiener filtering problem as an optimization scheme and provides an alternative promising approach for dealing with complex noise models. Münchmeyer and Smith (2019) recently proposed another interesting approach where a neural network was trained to Wiener filter CMB maps.

Elsner and Wandelt (2013) developed the messenger algorithm which bypasses the need for a preconditioner, with this approach being efficient and robust. In Section 2.5, we describe the conceptual framework of both the standard and a new formulation that is dual to the messenger algorithm. The evaluation of the Wiener filter from Eq. (2.3) via both the standard and dual messenger algorithms constitutes the crux of Chapter 3.

2.3 Some astrophysical and cosmological applications

Some common applications of Wiener filtering in the analysis of CMB data include map-making (e.g. Bunn et al., 1994; Tegmark, 1997a), optimal power spectrum estimation (e.g. Tegmark, 1997b; Bond, Jaffe, and Knox, 1998; Oh, Spergel, and Hinshaw, 1999; Elsner and Wandelt, 2012a), likelihood analysis (e.g. Hinshaw et al., 2007; Dunkley et al., 2009; Elsner and Wandelt, 2012c), treatment of foregrounds (e.g. Bouchet, Prunet, and Sethi, 1999), reconstruction of lensing potential, de-lensing and template-matching (e.g. Hirata and Seljak, 2003; Hirata et al., 2004; Seljak and Hirata, 2004; Hanson et al., 2013; Manzotti et al., 2017; Millea, Anderes, and Wandelt, 2017) and investigation of primordial non-Gaussianity (e.g. Komatsu, Spergel, and Wandelt, 2005; Elsner and Wandelt, 2009; Elsner and Wandelt, 2010). Moreover, we encounter the Wiener filter in Bayesian inference analyses involving the large-scale structures or the CMB (e.g. Wandelt, Larson, and Lakshminarayanan, 2004; Eriksen et al., 2004; O'Dwyer et al., 2004; Jewell, Levin, and Anderson, 2004; Jasche et al., 2010b; Jasche and Wandelt, 2013b; Jasche and Lavaux, 2015; Anderes, Wandelt, and Lavaux, 2015; Alsing et al., 2016; Jasche and Lavaux, 2017). It has also been employed for the reconstruction of 21-cm signal from contaminated data (Gleser, Nusser, and Benson, 2008) and velocity field reconstructions (Lavaux, 2016).

2.4 Preconditioned conjugate gradient algorithm

In general, the PCG approach consists of solving the following set of linear equations:

$$\boldsymbol{\mathcal{A}}\boldsymbol{x} = \mathbf{y}, \tag{2.4}$$

where $\boldsymbol{\mathcal{A}}$ is usually a very large matrix. We wish to avoid computing the inverse of this dense matrix and to this end, in the traditional PCG scheme, we make use of a sparse matrix known as

the preconditioner $\boldsymbol{\mathcal{M}}$, which is the approximate inverse of $\boldsymbol{\mathcal{A}}$, i.e. $\boldsymbol{\mathcal{M}} \approx \boldsymbol{\mathcal{A}}^{-1}$, as follows:

$$\boldsymbol{\mathcal{M}}\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{\mathcal{M}}\mathbf{y}. \tag{2.5}$$

The Wiener filter Eq. (2.3) can be rewritten as

$$(\mathbb{1} + \mathbf{S}^{1/2}\mathbf{N}^{-1}\mathbf{S}^{1/2})\mathbf{S}^{-1/2}\boldsymbol{s}_{\rm WF} = \mathbf{S}^{1/2}\mathbf{N}^{-1}\boldsymbol{d}, \tag{2.6}$$

such that $\boldsymbol{\mathcal{A}} = \mathbb{1} + \mathbf{S}^{1/2}\mathbf{N}^{-1}\mathbf{S}^{1/2}$, $\boldsymbol{x} = \mathbf{S}^{-1/2}\boldsymbol{s}_{\rm WF}$ and $\mathbf{y} = \mathbf{S}^{1/2}\mathbf{N}^{-1}\boldsymbol{d}$, in accordance with Eq. (2.4). For the specific problem considered in Chapter 3, we make use of a suitable preconditioner which is diagonal in Fourier space, as follows:

$$\mathcal{M}^{-1}_{\ell\ell} = \mathcal{A}_{\ell\ell} = \left(\sum_k n_k\right) C_\ell \frac{1}{L^4}, \tag{2.7}$$

where n_k are the eigenvalues of the inverse noise covariance matrix, $\mathbf{N}^{-1}$. We may then implement this preconditioner in a PCG algorithm (e.g. Golub and Van Loan, 1996) which proceeds as follows: We make an initial guess $\boldsymbol{x}_0$ and iterate until $|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i| \,/\, |\boldsymbol{x}_i| < \epsilon$, resulting in an $\boldsymbol{x}_{\rm WF}$ that corresponds to the Wiener filter solution $\boldsymbol{s}_{\rm WF}$. If convergence is not yet achieved, we re-initialize all parameters every 200 iterations and resume iterations, in order to facilitate convergence. An in-depth review of the PCG algorithm is provided in Shewchuk (1994). Again, we stress the fact that finding a suitable preconditioner is the key factor when making use of the PCG method (e.g. Oh, Spergel, and Hinshaw, 1999) and this consequently is the major stumbling block for state-of-the-art CMB data analysis.

2.5 Messenger methods

In this section, we describe the conceptual underpinnings of the standard messenger and the new dual messenger methods as an elegant solution to the Wiener filtering problem.

2.5.1 The standard messenger algorithm

Elsner and Wandelt (2013) proposed a high precision, iterative algorithm for the solution to the full Wiener filter equation, while being numerically efficient and straightforward to implement. Conceptually, the key idea is to introduce a stochastic auxiliary field $\boldsymbol{t}$, the so-called "messenger" field, with covariance $\mathbf{T}$, where $\mathbf{T}$ is proportional to the identity matrix. Taking advantage of the useful property of the identity matrix being invariant under any orthogonal basis transformation, the messenger field acts as an intermediate in transformations between different preferred orthogonal bases, in which signal and noise covariance matrices are expressed conveniently, i.e. are sparse. As a result, although directly applying combinations like $(\mathbf{S} + \mathbf{N})^{-1}$ may not be possible, we can always apply expressions like $(\mathbf{S} + \mathbf{T})^{-1}$ and $(\mathbf{N} + \mathbf{T})^{-1}$, irrespective of the basis chosen to render $\mathbf{S}$ and $\mathbf{N}$ sparse. Under such a scheme, the information from the data is transmitted to the signal via the messenger field that can be transformed efficiently from one basis representation to another, thereby obviating the requirement to apply the inverse Wiener covariance matrix to data.

With the introduction of the messenger field $\boldsymbol{t}$, the modified χ^2, where the posterior probability distribution of $\boldsymbol{s}$ is proportional to $\exp\left(-\chi^2/2\right)$, is as follows:

$$\chi^2_T = (\boldsymbol{d}-\boldsymbol{t})^\dagger \bar{\mathbf{N}}^{-1}(\boldsymbol{d}-\boldsymbol{t}) + (\boldsymbol{t}-\boldsymbol{s})^\dagger \mathbf{T}^{-1}(\boldsymbol{t}-\boldsymbol{s}) + \boldsymbol{s}^\dagger \mathbf{S}^{-1}\boldsymbol{s}, \tag{2.8}$$

where we defined $\bar{\mathbf{N}} \equiv \mathbf{N} - \mathbf{T}$, and we choose the covariance matrix of the auxiliary field $\boldsymbol{t}$ according to $\mathbf{T} = \alpha\mathbb{1}$, where $\alpha \equiv \min(\mathrm{diag}(\mathbf{N}))$. Minimizing with respect to $\boldsymbol{s}$ and $\boldsymbol{t}$ leads to the following two equations:

$$\left[\bar{\mathbf{N}}^{-1} + (\lambda\mathbf{T})^{-1}\right]\boldsymbol{t} = \bar{\mathbf{N}}^{-1}\boldsymbol{d} + (\lambda\mathbf{T})^{-1}\boldsymbol{s} \tag{2.9}$$

$$\left[\mathbf{S}^{-1} + (\lambda\mathbf{T})^{-1}\right]\boldsymbol{s} = (\lambda\mathbf{T})^{-1}\boldsymbol{t}, \tag{2.10}$$

where we also introduced a scalar parameter λ whose purpose is to accelerate convergence. In the limit of $\lambda = 1$, the above system of Eqs. (2.9) and (2.10) reduces to the usual Wiener filter Eq. (2.3), as shown in Appendix A.1. Note that here we set the response operator to identity matrix, i.e. $\boldsymbol{\mathcal{R}} = \mathbb{1}$, for simplicity. The messenger algorithm basically involves the following steps, as outlined in Algorithm 1. We initialize the vectors $\boldsymbol{s}$ and $\boldsymbol{t}$ with zeros, and choose an initial high value of λ. We first solve Eq. (2.9) for the messenger field $\boldsymbol{t}$ in the basis defined by $\mathbf{N}$, the noise covariance matrix. Then, we change to a basis where the signal covariance matrix $\mathbf{S}$ has a sparse representation, such as Fourier space. Next, we solve for $\boldsymbol{s}$ using Eq. (2.10) and the resulting $\boldsymbol{t}$ from the previous step. Finally, we transform the result back to the original basis. The signal reconstruction converges to the Wiener filter solution, i.e. $\boldsymbol{s} \to \boldsymbol{s}_{\rm WF}$, as $\lambda \to 1$.

Algorithm 1 Messenger algorithm

procedure MESSENGER($\boldsymbol{d}$, $\mathbf{N}$, $\mathbf{S}$, $N_{\rm pix}$, L)
  $\boldsymbol{s}_0 = \mathrm{zeros}(N_{\rm pix}, N_{\rm pix})$ ▷ Initialize $\boldsymbol{s}$ with zeros
  $\boldsymbol{t}_0 = \boldsymbol{d}$ ▷ Initialize $\boldsymbol{t}$ via an initial guess
  $\alpha = \min(\mathrm{diag}(\mathbf{N}))$ ▷ Compute the covariance of messenger field $\boldsymbol{t}$, such that $\mathbf{T} = \alpha\mathbb{1}$
  $\bar{\mathbf{N}} = \mathbf{N} - \mathbf{T}$ ▷ Compute the covariance $\bar{\mathbf{N}}$
  while $\lambda = 10^4 \to 1$ **do**
    repeat
      ▷ Transform to Fourier ($\mathcal{F}$) space, $\hat{\boldsymbol{s}}_{\vec{\ell}} = \mathcal{F}\boldsymbol{s}_{\vec{x}} = \left(\frac{L}{N}\right)^2 \sum_{\vec{x}} \omega^{-\vec{x}\cdot\vec{k}} \boldsymbol{s}_{\vec{x}}$, where $\omega = \exp(\frac{i2\pi}{N})$
      $\hat{\boldsymbol{s}}_{i+1,\vec{\ell}} = \left[\mathbf{S}^{-1} + \sigma(\lambda\mathbf{T})^{-1}\right]^{-1}_{\vec{\ell}} \mathcal{F}\left[(\lambda\mathbf{T})^{-1}_{\vec{x}} \boldsymbol{t}_{i,\vec{x}}\right]_{\vec{\ell}}$
      ▷ $\sigma = \frac{N_{\rm pix}^2}{L^4}$ is a numerical factor due to $\mathcal{F}$
      $\boldsymbol{s}_{i+1,\vec{x}} = \mathcal{F}^{-1}(\hat{\boldsymbol{s}}_{i+1,\vec{\ell}})$ ▷ Transform to pixel space
      $\boldsymbol{t}_{i+1,\vec{x}} = \left[\bar{\mathbf{N}}^{-1} + (\lambda\mathbf{T})^{-1}\right]^{-1}_{\vec{x}} \left[\bar{\mathbf{N}}^{-1}\boldsymbol{d} + (\lambda\mathbf{T})^{-1}\boldsymbol{s}_{i+1}\right]_{\vec{x}}$
      $i \leftarrow i + 1$
    until $\|\boldsymbol{s}_i - \boldsymbol{s}_{i-1}\| \,/\, \|\boldsymbol{s}_i\| < \epsilon$
    $\lambda \leftarrow \lambda \times \eta$ ▷ Cooling scheme for λ
  end while
  $\boldsymbol{s} \to \boldsymbol{s}_{\rm WF}$ ▷ as $\lambda = 1$
  return $\boldsymbol{s}_{\rm WF}$
end procedure

If ϵ_i is the residual at the i^{th} step, then at the following $(i+1)^{\text{th}}$ step, the corresponding residual is

$$\epsilon_{i+1} = \left[\mathbf{S}(\mathbf{S}+\lambda\mathbf{T})^{-1}\right]\left[\bar{\mathbf{N}}(\bar{\mathbf{N}}+\lambda\mathbf{T})^{-1}\right]\epsilon_i \tag{2.11}$$

$$= \left[\mathbf{S}(\mathbf{S}+\lambda\mathbf{T})^{-1}\right]\left\{(\mathbf{N}-\mathbf{T})\left[\mathbf{N}+(\lambda-1)\mathbf{T}\right]^{-1}\right\}\epsilon_i. \tag{2.12}$$

For the case of homogeneous noise, i.e. $\mathbf{N} \propto \mathbb{1}$, the system is solved exactly in a single step. A key observation is that $|\epsilon_{i+1}| < |\epsilon_i|$ for all i since the terms in brackets are less than unity, resulting in unconditional convergence of the signal reconstruction $\boldsymbol{s}$ to the Wiener filter solution $\boldsymbol{s}_{\text{WF}}$. A careful inspection of Eq. (2.12) yields the following insight: Convergence is fast for low noise pixels and modes with low signal prior variance while the converse is also true, i.e. the system converges slowly for high prior variance and high noise pixels. In the latter regimes, a high value of λ would speed up convergence. Elsner and Wandelt (2013) have shown that it is possible to find a cooling scheme for λ, where $\lambda \gg 1$ initially, to smoothly bring the algorithm to the final solution ($\lambda = 1$). Here, we reduce λ by a constant factor $(1/\eta)$, where $1/2 \leq \eta \leq 1$. The rationale behind the cooling scheme adopted in this work is laid out in Appendix A.1.

The standard messenger method has enjoyed considerable success, thereby establishing its credentials as a reliable method for fast Wiener filtering, without having recourse to a preconditioner. Essentially, it speeds up the computation of the Wiener filter when the noise is strongly inhomogeneous and/or the data is masked. Some applications of the standard messenger algorithm are listed in Section 2.5.3.

2.5.2 The dual messenger algorithm

In the above messenger algorithm, we made use of a relation to trivially split the noise into two components: one with a trivial covariance matrix and the other as a fluctuating component over the sky. An alternative approach is to introduce the auxiliary field at the level of the signal in a complementary formalism to the standard messenger framework. Due to the two schemes being complementary to each other, the new algorithm is referred to as the dual messenger algorithm. The corresponding log-posterior to be optimized then becomes

$$\chi^2_U = (\boldsymbol{d}-\boldsymbol{s})^\dagger\mathbf{N}^{-1}(\boldsymbol{d}-\boldsymbol{s}) + (\boldsymbol{s}-\boldsymbol{u})^\dagger\mathbf{U}^{-1}(\boldsymbol{s}-\boldsymbol{u}) + \boldsymbol{u}^\dagger\bar{\mathbf{S}}^{-1}\boldsymbol{u}, \tag{2.13}$$

where, analogous to the standard approach, $\mathbf{U} = \nu\mathbb{1}$ with $\nu \equiv \min(\text{diag}(\mathbf{S}))$, and the covariance of the auxiliary field, $\bar{\mathbf{S}} \equiv \mathbf{S} - \mathbf{U}$. We derive the corresponding equations that must be satisfied by $\boldsymbol{s}$ and $\boldsymbol{u}$ at the minimum of χ^2_U:

$$\left(\mathbf{N}^{-1}+\mathbf{U}^{-1}\right)\boldsymbol{s} = \mathbf{N}^{-1}\boldsymbol{d} + \mathbf{U}^{-1}\boldsymbol{u} \tag{2.14}$$

$$\left(\mathbf{U}^{-1}+\bar{\mathbf{S}}^{-1}\right)\boldsymbol{u} = \mathbf{U}^{-1}\boldsymbol{s}. \tag{2.15}$$

The dual messenger algorithm has interesting convergence properties. The amount of reduction of the residual at each iteration is given by

$$\epsilon_{i+1} = [\mathbf{N}(\mathbf{N}+\mathbf{U})^{-1}][\bar{\mathbf{S}}(\bar{\mathbf{S}}+\mathbf{U})^{-1}]\epsilon_i \tag{2.16}$$
$$= [\mathbf{N}(\mathbf{N}+\mathbf{U})^{-1}][(\mathbf{S}-\mathbf{U})\mathbf{S}^{-1}]\epsilon_i. \tag{2.17}$$

This provides the basis for the following mechanism: We artificially truncate the spectrum $\mathbf{S}$ to some lower initial value of $\ell_{\rm iter}$ that corresponds to a covariance μ and bring $\ell_{\rm iter}$ slowly to $\ell_{\rm max}$ corresponding to our final covariance ν. So, essentially, we vary the covariance $\mathbf{U}$ via a cooling scheme to bring $\mu \to \nu$, where, in the limit $\mu = \nu$, the above system of Eqs. (2.14) and (2.15) reduces to the usual Wiener filter Eq. (2.3), as shown in Appendix A.2. This results in a redefinition of $\bar{\mathbf{S}}$ using the Heaviside function as $\bar{\mathbf{S}} = \Theta(\mathbf{S}-\mathbf{U})$, as described quantitatively in Appendix A.2 (cf. Eqs. (A.17) and (A.18)). For $\ell \lesssim \ell_{\rm iter}$, the ratio given by Eq. (2.17) is always convergent. The algorithm consists of the following steps. We initialize the vectors $\boldsymbol{s}$ and $\boldsymbol{u}$ with zeros. We begin iterations with an initial value of $\ell_{\rm iter}$ and correspondingly μ, and iterate until $\mu \to \nu$ at $\ell_{\rm iter} = \ell_{\rm max}$, i.e. $\boldsymbol{s} \to \boldsymbol{s}_{\rm WF}$, in accordance with a chosen convergence criterion for each value of μ. In analogy with the messenger technique, we need to adopt a cooling scheme for μ. The idea is to reduce μ by a constant factor $(1/\beta)$, where $0 < \beta < 1$, which matches the convergence speed of one iteration with the modification in μ. The rationale behind the cooling scheme tailored for the dual messenger algorithm is illustrated in Appendix A.2.

Algorithm 2 Dual messenger algorithm

procedure DUAL MESSENGER($\boldsymbol{d}$, $\mathbf{N}$, $\mathbf{S}$, $N_{\rm pix}$, L)
  $\boldsymbol{s}_0 = \text{zeros}(N_{\rm pix}, N_{\rm pix})$ ▷ Initialize $\boldsymbol{s}$ with zeros
  $\boldsymbol{t}_0 = \boldsymbol{d}$ ▷ Initialize $\boldsymbol{t}$ via an initial guess
  $\alpha = \min(\text{diag}(\mathbf{N}))$ ▷ Compute the covariance of auxiliary field $\boldsymbol{t}$, such that $\mathbf{T} = \alpha\mathbb{1}$
  $\bar{\mathbf{N}} = \mathbf{N} - \mathbf{T}$ ▷ Compute the covariance $\bar{\mathbf{N}}$
  while $\xi = (\alpha+\mu) \to \alpha$ **do**
    $\mathbf{U} = (\sigma\mu)\mathbb{1}$ ▷ Compute covariance $\mathbf{U}$
    ▷ As in Algorithm 1, factor of σ due to $\mathcal{F}$
    $\bar{\mathbf{S}} = \Theta(\mathbf{S}-\mathbf{U})$ ▷ Compute covariance $\bar{\mathbf{S}}$
    repeat
      $\hat{\boldsymbol{t}}_{i+1,\vec{\ell}} = \left[(\bar{\mathbf{S}}^{-1} + \sigma\boldsymbol{\xi}^{-1})^{-1}\sigma\boldsymbol{\xi}^{-1}\right]_{\vec{\ell}} \mathcal{F}(\boldsymbol{t}_{i,\vec{x}})$ ▷ Moving to Fourier space, $\mathcal{F}$
      $\boldsymbol{s}_{i+1,\vec{x}} = \mathcal{F}^{-1}(\hat{\boldsymbol{t}}_{i+1,\vec{\ell}})$ ▷ Transform to pixel space
      $\boldsymbol{t}_{i+1,\vec{x}} = \left(\mathbf{N}^{-1} + \boldsymbol{\xi}^{-1}\right)^{-1}_{\vec{x}} \left(\mathbf{N}^{-1}\boldsymbol{d} + \boldsymbol{\xi}^{-1}\boldsymbol{s}_{i+1}\right)_{\vec{x}}$
      $i \leftarrow i+1$
    until $\|\boldsymbol{s}_i - \boldsymbol{s}_{i-1}\| \,/\, \|\boldsymbol{s}_i\| < \epsilon$
    $\xi \leftarrow \xi \times \beta$ ▷ Cooling scheme for ξ
    $\mu \leftarrow (\xi - \alpha)/\sigma$ ▷ Compute resulting μ
  end while
  $\boldsymbol{s} \to \boldsymbol{s}_{\rm WF}$ ▷ as $\xi = \alpha$, $\mu = 0$
  return $\boldsymbol{s}_{\rm WF}$
end procedure

However, numerically, the above algorithm does not result in the correct final solution due to the continuous mode of the signal, i.e. the zero eigenvalue in the signal covariance $\mathbf{S}$. We therefore

require $\mu \to \nu = 0$ to obtain the proper solution at the end, which cannot be accommodated by the dual messenger scheme above. To remedy this numerical predicament, we introduce an extra degree of freedom, α, in the system, where $\alpha \equiv \min(\mathrm{diag}(\mathbf{N}))$, thereby incorporating aspects of the standard messenger method into the dual messenger scheme. This leads to the following χ^2:

$$\chi^2_{\tilde{\xi}} = (\boldsymbol{d}-\boldsymbol{t})^\dagger \bar{\mathbf{N}}^{-1}(\boldsymbol{d}-\boldsymbol{t}) + (\boldsymbol{t}-\boldsymbol{s})^\dagger \boldsymbol{\xi}^{-1}(\boldsymbol{t}-\boldsymbol{s}) + \boldsymbol{s}^\dagger \bar{\mathbf{S}}^{-1}\boldsymbol{s}, \tag{2.18}$$

where $\boldsymbol{\xi} = (\alpha+\mu)\mathbb{1} = \xi\mathbb{1} = \mathbf{T}+\mathbf{U}$, with μ, $\mathbf{T}$, $\bar{\mathbf{S}}$ and $\bar{\mathbf{N}}$ inheriting their previous definitions. The corresponding set of equations to be solved iteratively is then:

$$\left(\bar{\mathbf{N}}^{-1} + \boldsymbol{\xi}^{-1}\right)\boldsymbol{t} = \bar{\mathbf{N}}^{-1}\boldsymbol{d} + \boldsymbol{\xi}^{-1}\boldsymbol{s} \tag{2.19}$$

$$\left(\boldsymbol{\xi}^{-1} + \bar{\mathbf{S}}^{-1}\right)\boldsymbol{s} = \boldsymbol{\xi}^{-1}\boldsymbol{t}. \tag{2.20}$$

If $\alpha = 0$, we recover the usual dual messenger scheme, while setting $\mu = 0$ yields the standard messenger algorithm. The definition of $\boldsymbol{\xi}$ implies that the cooling scheme described above still applies to this hybrid method. As outlined in Algorithm 2, we proceed in similar steps as described above for the previous scheme, except that here we reduce the norm of $\boldsymbol{\xi}$ by the factor $(1/\beta)$ and iterate until $\xi \to \alpha$, at which point $\mu = 0$ and we obtain the proper solution as desired.

2.5.3 Applications of the messenger methods

Elsner and Wandelt (2012b) and Elsner and Wandelt (2013) applied the messenger algorithm on CMB data from WMAP satellite and found the final map to be accurate to about 1 part in 10^5, compared to standard conjugate gradient solvers. Elsner and Wandelt (2012b) also demonstrated the generation of constrained realizations using the messenger method. Even with the inclusion of polarization data in the analysis, the messenger algorithm maintained its efficiency. Mangilli et al. (2013) implemented the messenger algorithm to produce Wiener-filtered simulations of non-Gaussian CMB maps with the lensing-integrated Sachs Wolf bispectrum signal. The messenger method was also employed in Planck data analysis, specifically for inverse covariance filtering of CMB maps at high angular resolutions (Planck Collaboration et al., 2014a; Planck Collaboration et al., 2014b).

Jasche and Lavaux (2015) implemented a Gibbs sampling adaptation of the messenger algorithm in a simple, easy to implement but efficient algorithm for Bayesian large-scale structure inference, specifically aiming at the joint inference of cosmological density fields and power spectra for linear data models. Anderes, Wandelt, and Lavaux (2015) adopted a similar approach in their Bayesian hierarchical modelling of the CMB gravitational lensing. Alsing et al. (2016) also implemented the Gibbs-messenger sampling adaptation developed by Jasche and Lavaux (2015) for Bayesian hierarchical modelling of cosmic shear power spectrum inference and eventually for cosmological parameter inference (Alsing, Heavens, and Jaffe, 2017). The works of Jasche and Lavaux (2015), Anderes, Wandelt, and Lavaux (2015) and Alsing et al. (2016) show that the messenger algorithm can be successfully adapted for high-resolution conditional Gaussian sampling using Markov Chain Monte Carlo techniques, demonstrating the flexibility of the method.

Messenger techniques are becoming increasingly popular and are being further developed as a viable solution to complex and realistic problems. Recently, Huffenberger and Næss (2018) adapted the standard messenger method for map-making applications. Using mock Advanced ACTPol-data, they illustrated the superior quality of maps obtained relative to a traditional PCG approach, thereby showcasing messenger field map-making as a potentially powerful CMB data analysis tool. Huffenberger (2018) also applied the messenger method to problems with multiple, uncorrelated noise sources.

The dual messenger algorithm has also been implemented in the field of optical and information engineering. For instance, it has been adapted for the removal of atmospheric haze from images (Fu et al., 2018), demonstrating the versatility of the tool developed. This class of messenger methods can therefore be tailored to solve a range of Wiener filtering problems and is not limited to astrophysical and cosmological applications.

Chapter 3

Application to CMB temperature anisotropies

The work presented in this chapter is based on Kodi Ramanah, Lavaux, and Wandelt (2017).

3.1 Generation of artificial CMB data set

In this section, we describe the generation of a mock CMB data set to investigate the efficacy of our dual messenger technique. We generated artificial CMB data, in a realistic scenario, by drawing Gaussian random fields on a 2D flat sky with grid resolution, $N_{\rm pixels} = 512^2$, and angular extent, $L = 10.0$ degrees. This true simulated map is contaminated by Gaussian white noise with covariance matrix, $\mathbf{N} = (64.0\,\mu\mathrm{K}^2)\,\mathbb{1}$, via a linear model as given in Eq. (2.1). We made use of CAMB[1] (Lewis, Challinor, and Lasenby, 2000) to generate the input angular power spectrum from which the CMB signals are drawn. We assume a standard ΛCDM cosmology with the set of cosmological parameters ($\Omega_m = 0.32$, $\Omega_\Lambda = 0.69$, $\Omega_b = 0.05$, $h = 0.67$, $\sigma_8 = 0.83$, $n_s = 0.97$) from Planck (Planck Collaboration et al., 2016b). A portion of the flat sky is masked by imposing a central square patch of angular extent 5.0 degrees with a noise covariance of $10^6\mathbf{N}$. The three different methods, messenger, dual messenger and PCG, are applied to this set-up to investigate their effectiveness and efficiency of signal reconstruction. We implement the same "weak" Cauchy convergence criterion, $\|\boldsymbol{s}_{i+1} - \boldsymbol{s}_i\| \,/\, \|\boldsymbol{s}_i\| < \epsilon$, where $\epsilon = 10^{-6}$, in all the algorithms, to ensure a fair comparison. As a reference, we make use of the PCG method with the more stringent $\epsilon = 10^{-9}$ to provide results against which the other methods can be compared. Currently regarded as the standard Wiener filtering technique by the scientific community, the PCG method is the natural choice for providing a reference solution in the absence of an exact analytic solution. By imposing a more stringent convergence criterion, we ensure that the PCG solution is close to the ground truth.

The remainder of this chapter is organized as follows. We investigate the performance of our dual messenger algorithm in terms of convergence, computation time and stability, and draw comparisons to the standard messenger scheme and the popular PCG method in Section 3.2. Finally, in Section 3.3, we summarize the main aspects of our findings and discuss the areas of applications where the potential of our new algorithm can be fully exploited. In Appendices A.1 and A.2, we illustrate the rationale behind the schemes implemented in this work for fast convergence.

[1] http://camb.info

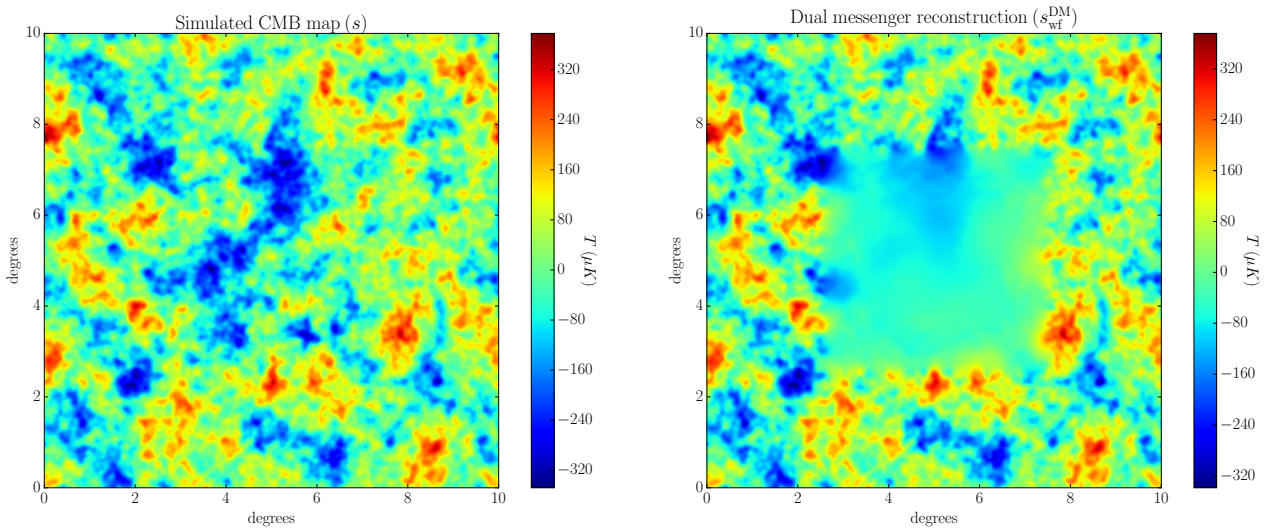


FIGURE 3.1: The simulated CMB map and the reconstruction via the dual messenger algorithm. The left-hand panel shows the simulated map, which is subsequently contaminated by white noise, with the central square patch masked by extremely high noise covariance. The right-hand panel illustrates the corresponding reconstructed map obtained via the dual messenger technique. The messenger and PCG reconstructions are not shown as they look similar to the dual messenger reconstructed map. The residual maps displayed in Fig. 3.3, however, help to discern the differences between the various reconstructions.

3.2 Comparison of messenger and conjugate gradient methods

We now perform a quantitative assessment of the efficacy of reconstruction of the standard and dual messenger algorithms and a conventional PCG[2] method.

3.2.1 Reconstruction of CMB maps and power spectra

The true CMB simulated signal and the reconstructed map obtained using the dual messenger algorithm are displayed in Fig. 3.1. The characteristic feature of Wiener filtering, i.e. the extrapolation of signal into the mask, can be distinctly observed from the reconstructed map. At an initial cursory glance, the reconstructed maps from all three methods appear rather similar, and hence, only the dual messenger reconstruction is displayed in Fig. 3.1. This observation holds for the low noise regions but under scrutiny, the reconstruction in masked regions shows some slight differences. The residual maps depicted in Fig. 3.3, generated by computing the difference, over the full sky, between our reference map and the corresponding reconstructed map obtained using each method, help to discern these differences. The messenger technique provides the most accurate reconstruction in the masked region as illustrated by its residual map, with less than 3% residuals. The dual messenger and PCG reconstructions yield around 4% and 9% residuals, respectively, with the residuals obtained via the former algorithm being concentrated on the edges of the mask.

The effectiveness of reconstruction of the dual messenger technique on both small and large scales is also manifest from Fig. 3.2 which shows the reconstructed power spectra from the different algorithms. The corresponding power spectra recovered via the messenger, dual messenger and PCG methods are all in good agreement, on all scales, with the reference one, computed using the

[2]The corresponding preconditioner adopted is described in Chapter 2.4.

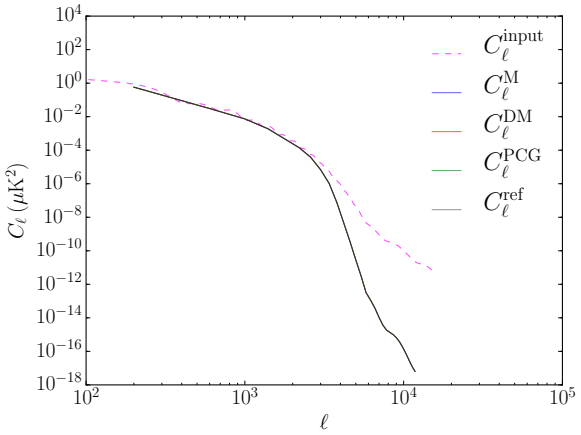


FIGURE 3.2: Reconstructed power spectra computed using the different algorithms. The dashed line indicates the input angular power spectrum, $C_\ell^{\rm input}$, from which the CMB signals are drawn. The power spectra recovered by the three algorithms are all in good agreement with the reference power spectrum computed using PCG method with $\epsilon = 10^{-9}$ on all scales. The deviations on the small scales from the input power spectrum are due to the characteristic feature of a Wiener filtered signal, where the power on small scales, in the low signal to noise regime, is suppressed.

PCG scheme with the more stringent convergence criterion of $\epsilon = 10^{-9}$. The characteristic feature of Wiener filtering where the recovered power on the small scales in the low signal to noise regime is suppressed is also observed from Fig. 3.2. Standard Wiener filtering algorithms usually encounter difficulties in dealing with masked regions having infinite noise. By masking 25% of the simulated map with noise of the order 10^6 higher than the unmasked region, we are investigating the worst case performance aspect of the dual messenger algorithm. Hence, we do not expect the algorithm to encounter any difficulties in Wiener filtering maps with high levels of inhomogeneous noise.

The dual messenger algorithm is therefore relevant for current and future high-resolution CMB experiments such as South Pole Telescope, Advanced ACTPol, Simons Observatory and CMB-S4. The application of the algorithm can be extended to problems involving the polarization of the CMB. The formalism remains unchanged for spin field reconstruction, although the numerical implementation is less trivial due to the correlation between the temperature and polarization components of the signal covariance. The Wiener filtering of polarised CMB data with more complex noise models will be subjected to future investigation to further showcase the efficiency of the dual messenger algorithm in treating complex CMB problems.

3.2.2 Numerical convergence and stability

We now investigate the convergence properties of the dual messenger algorithm and also provide a more in-depth study of the performance of the standard messenger scheme. We quote the usual statistics for convergence and the change in χ^2 of the posterior probability density between successive iterations for each method, so that unbiased comparisons of their efficiency and effectiveness can be drawn.

Fig. 3.4 shows the residual error at each iteration for the different methods, as a function of the number of iterations. For the messenger algorithm, we relax the convergence criterion ϵ for high values of λ and reduce ϵ to 10^{-6} as $\lambda \to 1$ in a three-step procedure, as depicted in Fig. 3.4 in

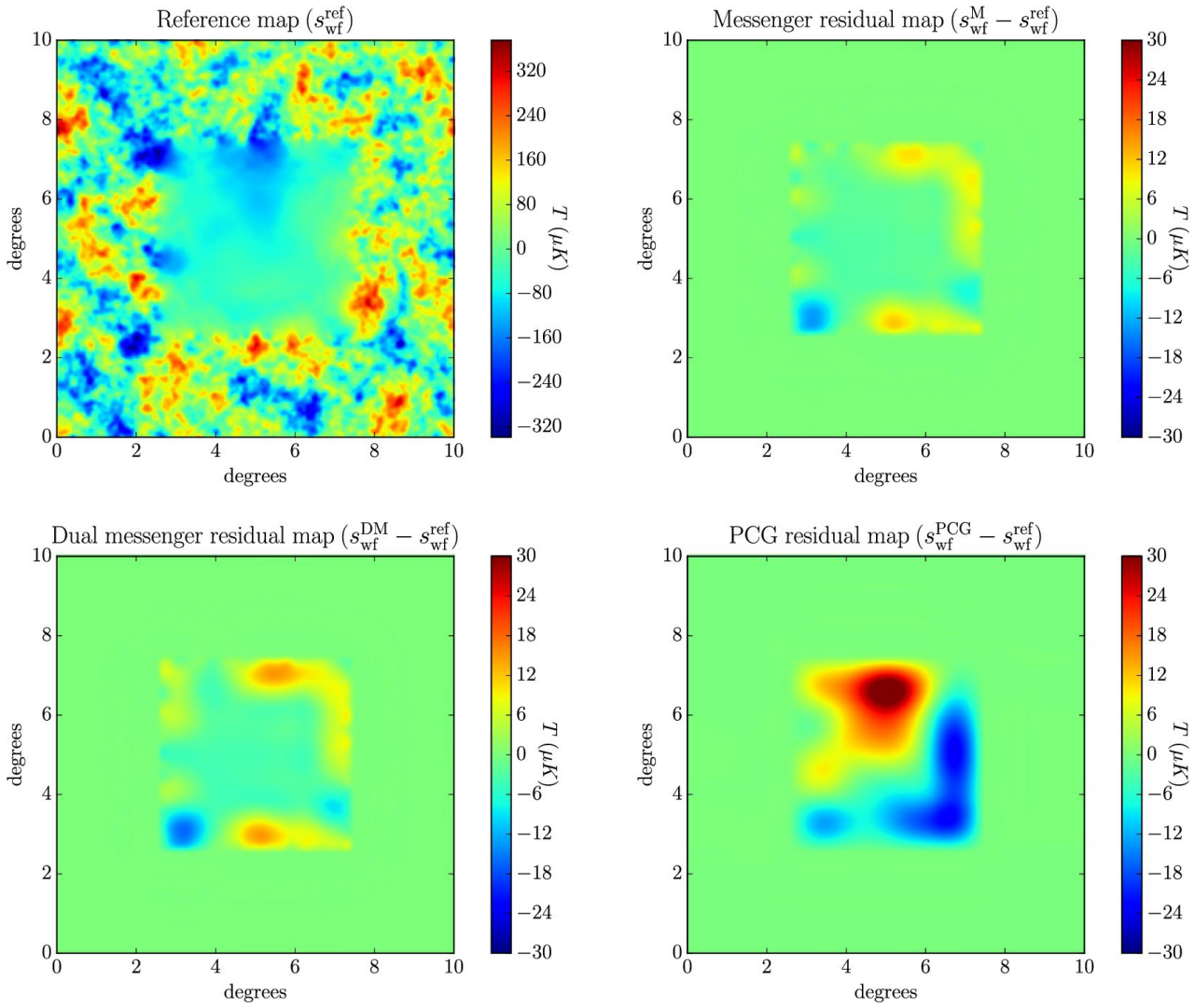
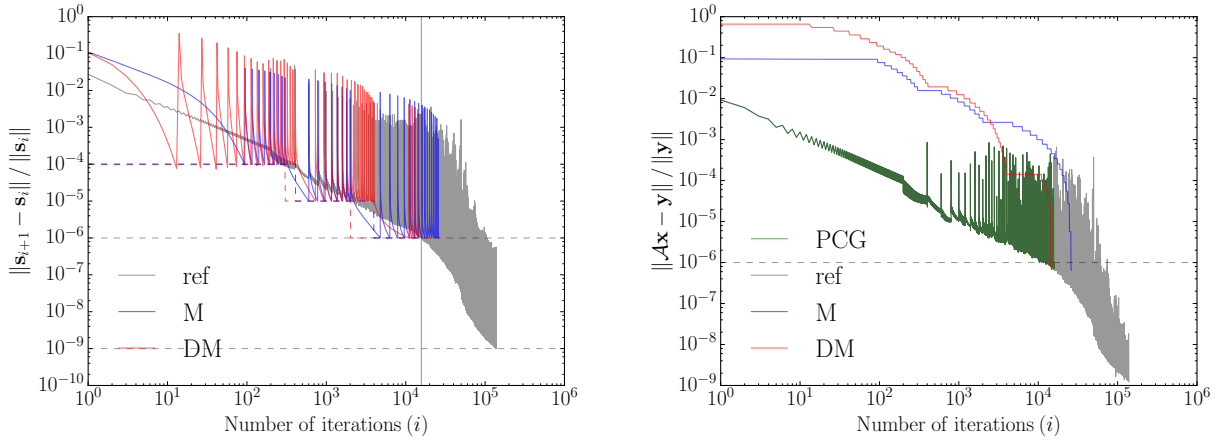


FIGURE 3.3: The reference map and residual maps from the three algorithms. The top left-hand panel depicts the reference map computed using the PCG scheme with $\epsilon = 10^{-9}$, while the other panels illustrate the residual maps yielded by the three methods, generated by computing the difference between the reference map and the corresponding reconstructed maps over the full sky. The messenger approach produces the least amount of residuals, around 3%, while the dual messenger and PCG schemes result in approximately 4% and 9% residuals, respectively. For the messenger reconstructions, the residuals lie mostly on the edges of the mask.

dashed red lines. Here, we adopt a cooling scheme where λ is reduced by a constant factor of $4/3$, i.e, $\eta = 3/4$, until $\lambda = 1$. This consequently ensures that the χ^2 decreases rapidly, thereby bringing us closer to the final solution with a smaller number of iterations, resulting in faster convergence.

For the dual messenger algorithm, we implemented a cooling scheme for ξ with $\beta = 3/4$, which results in fast convergence while providing accurate results. So, we pick an initial value of μ corresponding to an initial value of $\ell_{\rm iter}$, and therefore ξ, and iterate until convergence, as dictated by a given Cauchy criterion. We repeat this iterative procedure, in accordance with the aforementioned cooling scheme, until $\xi = \alpha$, at which point we have the Wiener filter solution. Again, we impose less stringent convergence criterion at higher values of ξ, as shown in Fig. 3.4 in dashed blue lines.

The left panel of Fig. 3.4 essentially illustrates the convergence of the different methods. The



(A) Cauchy criterion

(B) $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| / \|\boldsymbol{y}\|$

FIGURE 3.4: Variation of the residual errors, given by the Cauchy criterion (*left panel*) and $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| / \|\boldsymbol{y}\|$ (*right panel*), with number of iterations for the messenger and PCG methods. The Cauchy convergence criteria ϵ_i imposed for the different regimes of the cooling schemes for λ and μ are also displayed, along with the corresponding thresholds ϵ used for the two PCG methods, in dashed lines. The vertical green line denotes the convergence point of the PCG scheme, with its convergence behaviour already represented by the reference PCG method. The dual messenger algorithm requires the smallest number of iterations to converge to the final solution than the other two methods. The corresponding residual errors, characterized by $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| / \|\boldsymbol{y}\|$, for all methods drop below the convergence thresholds adopted, demonstrating the consistency of our computations dictated by the Cauchy criteria.

dual messenger algorithm requires slightly fewer iterations to converge to the final solution than the PCG approach, while the messenger technique requires nearly twice as many iterations as its dual counterpart for convergence. In terms of wall-clock times, the dual messenger has the definite upper hand, as it runs to completion in around 258 seconds whereas the messenger and PCG methods have corresponding wall-clock times of roughly 435 seconds and 663 seconds, i.e. reconstruction via the dual messenger algorithm is nearly three and two times faster than the PCG and messenger methods, respectively. All computations were run on a single core of an Intel Core i5-4690 CPU (3.50 GHz). Both messenger algorithms possess the same algorithmic complexity and memory requirements. They require two Fourier transforms, $\mathcal{O}(N_{\rm pixels} \log N_{\rm pixels})$, and two scalar multiplications corresponding to algebraic operations of $\mathcal{O}(N_{\rm pixels})$, per iteration. In terms of memory requirements, two vectors of size $N_{\rm pixels}$ must be temporarily stored in memory. In comparison, the PCG method requires three Fourier transforms and ten scalar multiplications per iteration, and temporary storage of eight vectors of dimension $N_{\rm pixels}$ in memory.

We also performed a consistency check by verifying the variation of the residual error given by $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| / \|\boldsymbol{y}\|$. This is usually adopted as a convergence criterion for PCG computations. The variation of this residual error is illustrated in the right panel of Fig. 3.4. The corresponding residual errors obtained via the different algorithms all drop below their respective Cauchy convergence thresholds implemented. The oscillatory behaviour of the PCG solution, displayed in the left and right panels of Fig. 3.4, is mainly due to the re-initialization step after every two hundred iterations in the algorithm, as described in Section 2.4. However, there are also some oscillations in the residual errors due to the PCG method being susceptible to instabilities sourced by numerical noise. The

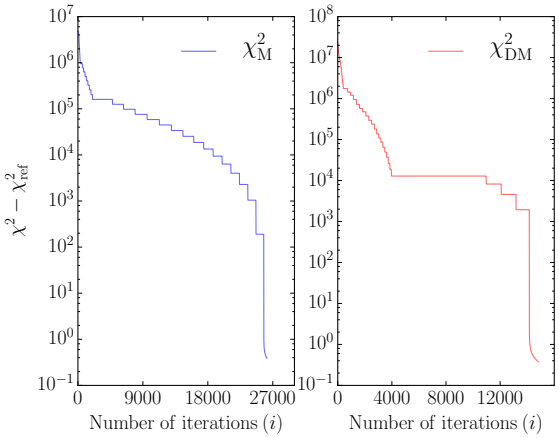


FIGURE 3.5: Variation of χ^2 with number of iterations for the two messenger methods. In the left panel, we find that the $\chi^2_{\rm M}$ of the messenger algorithm drops rapidly with the cooling scheme for λ adopted, as expected from the earlier discussion. In the right panel, the $\chi^2_{\rm DM}$ of its dual counterpart displays a similar behaviour. In both cases, the final solution matches the $\chi^2_{\rm ref}$ of the PCG method with $\epsilon = 10^{-9}$.

oscillations in the residual errors in Fig. 3.4 of the two messenger solutions are however due to their respective cooling schemes, resulting in transitions in the systems of equations with the varying covariances of the auxiliary fields (cf. Eqs. (A.5) and (A.22)), with the peaks produced coinciding with these transitions. It is important to note that the residual errors always drop sharply after the peaks, thereby demonstrating the unconditional stability of the messenger algorithms. From numerical experiments, the messenger techniques have proven to be far more stable than the PCG method for nearly degenerate systems.

Due to the Wiener filter being the maximum a posteriori solution, the χ^2 of the intermediate solution can be regarded as a useful convergence diagnostic. The χ^2 variation as a function of number of iterations is displayed in Fig. 3.5, with the left and right panels correspondingly showing the convergence of the messenger and dual messenger algorithms. The cooling scheme for λ implemented in the standard messenger technique causes the $\chi^2_{\rm M}$ to drop rapidly with each change in λ. Intuitively, this decrease of λ via a series of such steps seems reasonable since the χ^2 decreases sharply with each change in λ, and then reaches a plateau for a given λ until the latter decreases further, as evidenced in Fig. 2 of Elsner and Wandelt (2012b). The $\chi^2_{\rm DM}$ of the dual messenger algorithm has a similar behaviour. The χ^2 in both cases finally attains the $\chi^2_{\rm ref}$ of the reference PCG method with $\epsilon = 10^{-9}$, with $\Delta\chi^2_{\rm M}/\chi^2_{\rm ref} = 9.2 \times 10^{-6}$ and $\Delta\chi^2_{\rm DM}/\chi^2_{\rm ref} = 9.0 \times 10^{-6}$, where $\chi^2_{\rm ref} = 4.2 \times 10^4$.

Another important convergence diagnostic is the variation of the relative error characterized by $C_\ell(\boldsymbol{s}_{\rm WF} - \boldsymbol{s}^{\rm ref}_{\rm WF})/C_\ell(\boldsymbol{s}^{\rm ref}_{\rm WF})$, computed over the full sky, as a function of scale, ℓ, depicted in Figures 3.6, 3.7 and 3.8 for the messenger, dual messenger and PCG algorithms, respectively. The relative errors on the small scales are of the order 10^{-12}, 10^{-12} and 10^{-15}, and conversely, on the large scales, 10^{-3}, 10^{-3} and 10^{-2}, correspondingly, for the messenger, dual messenger and PCG methods. The messenger technique displays smooth and nearly uniform convergence on small and intermediate scales, with the relative error dropping below 10^{-6}, while remaining below 10^{-3} for the largest scales. For the dual messenger scheme, the corresponding convergence rate highlights the hierarchical fashion in which the solution is computed, while yielding similar final relative error across

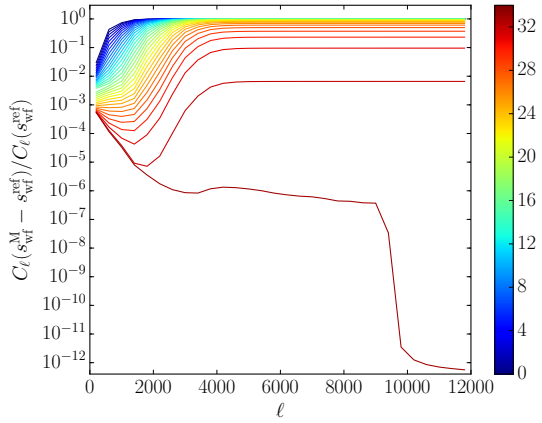


FIGURE 3.6: Convergence rate by frequency bin for the messenger method. This illustrates the relative error as a function of scale, ℓ. Each line in the figure corresponds to the relative error for a specific value of the scalar parameter λ. The colour bar indicates the steps involved in the cooling scheme. The messenger algorithm converges smoothly and in nearly uniform fashion on small and intermediate scales, with the relative error dropping till below 10^{-6}. However, for larger scales, the relative error is reduced by lower extent, but stays below 10^{-3} for the largest scales. For $\lambda = 1$, the behaviour is similar to that displayed by PCG (cf. Fig. 3.8).

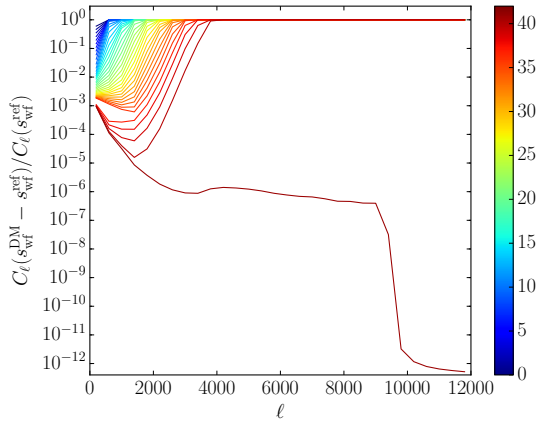


FIGURE 3.7: Convergence rate by frequency bin for the dual messenger method. Same as Fig. 3.6, except that here each line in the figure corresponds to a given value of ξ, displaying the hierarchical nature of this scheme. The final relative error is sufficiently low on all scales. A quantitative explanation of the convergence behaviour is given in Appendix A.2.

all scales as the messenger algorithm. The PCG method has the lowest relative error on the smallest scales, although this may be biased by the fact that the reference method is also a PCG, but remains inferior to both messenger methods on the largest scales. This is consistent with the significant residuals resulting from the PCG reconstruction, as observed in the previous section (cf. Fig. 3.3). We stress that all relative errors above are computed with respect to the reference PCG method with $\epsilon = 10^{-9}$.

We also carry out a series of additional runs with various values of the convergence criterion ϵ to investigate the wall-clock times required for convergence in each case, thereby providing a more in-depth picture of the performance of the different algorithms. As illustrated in Fig. 3.9, the dual messenger technique converges faster than the other methods, except for the extreme values of ϵ. It is also interesting to note that we can further reduce its execution time by lowering the factor β for

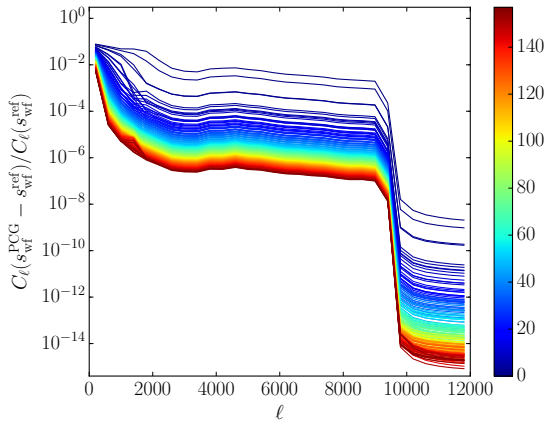


FIGURE 3.8: Convergence rate by frequency bin for the PCG method. Same as Fig. 3.6. The relative error as a function of scale for every $100^{\rm th}$ iteration is illustrated. The PCG method has a lower relative error than the dual messenger algorithm on the small scales, but the behaviour on the largest scales remains inferior to that of both messenger methods. The relative spacing of the lines indicates that convergence slows down as the PCG method progresses in iterations, and this is in agreement with its variation of the residual error with number of iterations, as displayed by the green line in Fig. 3.4.

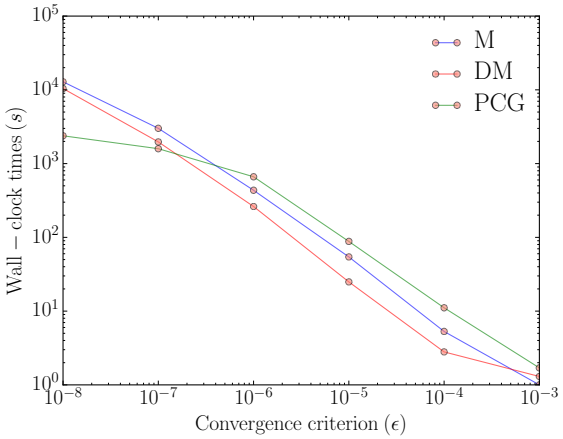


FIGURE 3.9: Execution times required for convergence as a function of convergence criterion ϵ for the different methods. This demonstrates the efficient performance of the two messenger algorithms with respect to the PCG method.

the cooling scheme without degrading the accuracy of results significantly. For instance, choosing $\beta = 1/2$ reduces the number of iterations required, and therefore computation time, by around 25% for the case $\epsilon = 10^{-6}$.

3.3 Conclusions

We presented a new formulation that is dual to the recently developed messenger algorithm, where, unlike in the standard approach, the auxiliary field is introduced at the level of the signal and is consequently associated with the signal reconstruction instead of the noise. This new iterative solver provides another pathway to solve the ill-posed problem of Wiener filtering, frequently encountered in several applications in cosmology and astrophysics.

We tested our new method on a simulated CMB data set and its performance was evaluated in terms of effectiveness of reconstruction, convergence properties, processing time and stability. The dual messenger scheme is shown to match the accuracy of reconstruction of the standard messenger and PCG methods on all scales. Regarding the convergence of the algorithm, it is shown to perform smoothly over all scales, with the relative error being sufficiently low even on the largest scales. For the specific problem under consideration, the dual messenger algorithm has a definite edge over the standard messenger scheme and the PCG approach in terms of computation time. The efficiency and effectiveness of this new technique in calculating the Wiener filter solution of general data sets has therefore been demonstrated. We also provided further insight into the mechanism of the standard messenger method so that it can be optimized for data analysis by the scientific community.

The dual messenger algorithm, like its predecessor, does not require an ingenious choice of preconditioner and is straightforward to implement, robust and flexible. It is also capable of taking into account inhomogeneous noise distributions and arbitrary mask geometries. A key aspect of this technique is that it computes the Wiener filter solution in a hierarchical manner due to the thresholding of the signal covariance matrix inherent in the algorithm. Given the success of the standard messenger method, this new dual messenger technique is highly promising and may therefore be adapted for high precision large-scale data analysis methods in cosmology and astrophysics.

Other performance improvements to this algorithm can be obtained by adapting the working resolution such that the Nyquist frequency is always slightly higher than the current ℓ_{iter} considered in the dual messenger method. This would consequently reduce the number of operations required for Fourier transform. Another possibility is a generalization of both the messenger and dual messenger algorithms in a combined approach that may further optimize the performance and efficiency compared to traditional techniques of solving the Wiener filter problem.

Chapter 4

Optimal and fast $\mathcal{E}/\mathcal{B}$ separation of CMB polarization

The work presented in this chapter is based on Kodi Ramanah, Lavaux, and Wandelt (2018).

4.1 Introduction

The polarization of the CMB radiation provides a window to probe the physics of the early Universe (e.g. Hu and White, 1997; Hu and Dodelson, 2002; Hu, 2003) and consequently lies at the frontiers of research in modern cosmology. Polarization maps can provide more stringent constraints on cosmological parameters for cosmic variance limited experiments (e.g. Galli et al., 2014), although their fascination lies primarily in the potential detection of the primordial $\mathcal{B}$-mode signals (e.g. Guzzetti et al., 2016; Kamionkowski and Kovetz, 2016), with the associated physical significance highlighted in Part I. The mathematical formalism of CMB polarization has been laid out around two decades ago by Kamionkowski, Kosowsky, and Stebbins (1997b) and Zaldarriaga and Seljak (1997), with the link to primordial gravitational waves extensively investigated (e.g. Kamionkowski, Kosowsky, and Stebbins, 1997a; Seljak and Zaldarriaga, 1997).

Cosmological inference from current and next-generation CMB experiments tailored for mapping the polarized sky therefore requires sophisticated tools to optimize the scientific returns. An underlying issue is to extract the gradient and curl components, or the $\mathcal{E}$ and $\mathcal{B}$ modes, of the polarization signal from the data. A potential solution is provided by the Wiener filter (Wiener, 1949), a powerful signal reconstruction tool that incorporates statistical information about the signal and noise properties. It has widespread applications in cosmology and astrophysics, especially in the post-processing of observational data, as outlined in Chapter 2.3. The numerical issues encountered in computing the exact Wiener filter solution are described in Chapter 2.2. Traditional methods involve the use of the preconditioned conjugate gradient (PCG) scheme (e.g. Eriksen et al., 2004; Wandelt, Larson, and Lakshminarayanan, 2004; Smith, Zahn, and Doré, 2007), with the panoply of PCG variants proposed in the literature outlined in Chapter 2.4.

The messenger method, first proposed by Elsner and Wandelt (2013), is a preconditioner-free Wiener filtering technique. In Chapter 3, we presented the dual messenger algorithm, an enhanced variant of the standard messenger approach, as a general-purpose tool for Wiener filtering, which surpasses its predecessor in execution time, with the focus being on the formalism and convergence properties of the algorithm (Kodi Ramanah, Lavaux, and Wandelt, 2017) (hereafter KLW17). We

demonstrated the efficiency, effectiveness and unconditional stability of the dual messenger scheme with respect to the PCG approach when analyzing CMB temperature maps.

The extension of the dual messenger algorithm for the analysis of polarized CMB data sets is the next key step. Moreover, Bunn and Wandelt (2017) have recently shown that the application of the Wiener filter to CMB polarization data may produce pure $\mathcal{E}$ and $\mathcal{B}$ maps, free from cross-contamination, at a much lower computational cost than the standard methods (e.g. Lewis, Challinor, and Turok, 2002; Lewis, 2003; Bunn et al., 2003; Bunn, 2011). This provides further motivation for an efficient polarized Wiener filtering tool.

The inclusion of polarization data significantly increases the condition number of the covariance matrices involved (Larson et al., 2007) such that finding appropriate preconditioners becomes highly non-trivial (Oh, Spergel, and Hinshaw, 1999). As a result, the PCG method may require expensive pre-computations (e.g. Seljebotn et al., 2014) and may be prone to numerical instabilities, as also illustrated in this work. The messenger techniques circumvent this ill-conditioning predicament with relative ease since they do not require any preconditioning, as demonstrated by Elsner and Wandelt (2013). They implemented the standard messenger method for the Wiener filtering of polarized maps from WMAP, but they excluded $\mathcal{B}$ modes from the analysis, thereby reducing the dimension of the signal covariance.

In this work, we adapt our recently proposed dual messenger algorithm for spin field reconstruction. We consider an artificially generated polarized CMB data set with correlated noise and distinct temperature and polarization masks, while incorporating the $\mathcal{B}$ modes in the polarization signal. We demonstrate the ease of implementation, efficiency and unconditional stability of the algorithm. As a comparison, we also implement a PCG method and illustrate the difficulties encountered in converging to a plausible solution.

This chapter is structured as follows. In Section 4.2, we provide a brief description of the dual messenger algorithm and outline its numerical implementation for polarized signal reconstruction and generation of constrained realizations. We also describe a generalized procedure for dealing with masks in Section 4.2.3. We then showcase the capabilities of our scheme in Section 4.3 and present a modified Jacobi relaxation scheme to further refine the solution in Section 4.3.5. Finally, we summarize our main findings in Section 4.4. We provide the preconditioner adopted for the PCG implementation in Appendix B.1.

4.2 Spin field reconstruction with a dual messenger field

The essence of the messenger methods lies in the introduction of an auxiliary field that acts as a mediator between the different bases where the signal and noise covariances, $\mathbf{S}$ and $\mathbf{N}$, can be conveniently expressed as sparse matrices. This essentially splits the Wiener filter equation into a set of algebraic equations that must be solved iteratively, obviating the need for matrix inversions or preconditioners. As pointed out in KLW17, the formalism of the dual messenger algorithm remains invariant with the inclusion of polarization data in the analysis. We briefly review the general system of equations that yields the two key equations to be implemented. A complementary and more in-depth description of the dual messenger algorithm is provided in KLW17. With the

introduction of the messenger fields $\boldsymbol{t}$ and $\boldsymbol{u}$ at the level of the noise and signal, respectively, the modified χ^2 is as follows:

$$\chi^2_{T,U} = (\boldsymbol{d}-\boldsymbol{t})^\dagger \bar{\mathbf{N}}^{-1}(\boldsymbol{d}-\boldsymbol{t}) + (\boldsymbol{t}-\boldsymbol{u})^\dagger \mathbf{T}^{-1}(\boldsymbol{t}-\boldsymbol{u}) + (\boldsymbol{u}-\boldsymbol{s})^\dagger \mathbf{U}^{-1}(\boldsymbol{u}-\boldsymbol{s}) + \boldsymbol{s}^\dagger \bar{\mathbf{S}}^{-1}\boldsymbol{s}, \tag{4.1}$$

where we defined $\bar{\mathbf{N}} \equiv \mathbf{N} - \mathbf{T}$, and we choose the covariance matrix of the auxiliary field $\boldsymbol{t}$ according to $\mathbf{T} = \alpha\mathbb{1}$, where $\alpha \equiv \min(\mathrm{diag}(\mathbf{N}))$, and in analogous fashion, $\mathbf{U} = \nu\mathbb{1}$ with $\nu \equiv \min(\mathrm{diag}(\mathbf{S}))$ for the second auxiliary field $\boldsymbol{u}$. The covariance of the signal $\boldsymbol{s}$ is then given by $\bar{\mathbf{S}} \equiv \mathbf{S} - \mathbf{U}$. Physically, $\boldsymbol{t}$ corresponds to a homogeneous component of the noise covariance, while $\boldsymbol{u}$ is the analogous component associated with the signal covariance.

Minimizing the χ^2 with respect to $\boldsymbol{s}$, $\boldsymbol{t}$ and $\boldsymbol{u}$ leads to the following three equations:

$$\boldsymbol{s} = (\mathbf{U}^{-1} + \bar{\mathbf{S}}^{-1})^{-1}\mathbf{U}^{-1}\boldsymbol{u} \tag{4.2}$$
$$\boldsymbol{u} = (\mathbf{U}^{-1} + \mathbf{T}^{-1})^{-1}(\mathbf{T}^{-1}\boldsymbol{t} + \mathbf{U}^{-1}\boldsymbol{s}) \tag{4.3}$$
$$\boldsymbol{t} = (\bar{\mathbf{N}}^{-1} + \mathbf{T}^{-1})^{-1}(\mathbf{T}^{-1}\boldsymbol{u} + \bar{\mathbf{N}}^{-1}\boldsymbol{d}). \tag{4.4}$$

We can reduce the above set of three equations to two equations, as in KLW17, as we need only one messenger field, but here, we contract the equations in an alternative way, to provide a more numerically convenient form of the equations adapted to deal with ill-conditioned systems, typical of CMB polarization problems, while improving convergence. By plugging Eq. (4.2) in Eq. (4.3) and defining $\boldsymbol{\xi} = \mathbf{U} + \mathbf{T}$, we obtain the following set of two equations to be solved iteratively:

$$\boldsymbol{u} = (\bar{\mathbf{S}} + \mathbf{U})(\bar{\mathbf{S}} + \boldsymbol{\xi})^{-1}\boldsymbol{t} \tag{4.5}$$
$$\boldsymbol{t} = (\bar{\mathbf{N}}^{-1} + \mathbf{T}^{-1})^{-1}(\mathbf{T}^{-1}\boldsymbol{u} + \bar{\mathbf{N}}^{-1}\boldsymbol{d}), \tag{4.6}$$

where Eq. (4.5) is simply a Wiener filter of $\boldsymbol{t}$, assuming a modified signal covariance $(\bar{\mathbf{S}} + \mathbf{U})$. The mechanism adopted to improve convergence is as follows: We artificially truncate the signal covariance $\mathbf{S}$ to some lower initial value of $\ell_{\rm iter}$ that corresponds to a covariance μ. By implementing a cooling scheme for $\boldsymbol{\xi}$, we subsequently vary $\mathbf{U}$ to bring $\mu \to \nu$, where, in the limit $\mu = \nu$, we have $\boldsymbol{u} = \boldsymbol{s}$ and the above system of Eqs. (4.5) and (4.6) reduces to the usual Wiener filter Eq. (2.3). This results in a redefinition of $\bar{\mathbf{S}}$ using the Heaviside function as $\bar{\mathbf{S}} = \Theta(\mathcal{S} - \mathbf{U})$, where $\mathcal{S}$ corresponds to the eigenvalues of $\mathbf{S}$. This leads to a hierarchical framework, where we obtain the solution on the largest scales initially, and gradually the algorithm resolves the fine structures on the small scales.

Due to the continuous mode of the signal, i.e. the zero eigenvalue of $\mathbf{S}$, we therefore require $\mu \to \nu = 0$ to finally obtain the desired Wiener filter solution. The cooling scheme for $\boldsymbol{\xi}$ involves reducing $\boldsymbol{\xi}$ by a constant factor and iterating until $\boldsymbol{\xi} \to \mathbf{T}$, at which point $\mu = 0$, as desired. Moreover, $\mathbf{U}$ does not need to be strictly proportional to the identity matrix, and this useful property allows us to solve the temperature and polarization signals at different rates. The rationale behind the above approach is described quantitatively in KLW17.

Note that the above equations, for simplicity, are written in a single basis, but all the operators are written in their respective bases, with all basis transformations made explicit, in Algorithm 3

below. In terms of the numerical implementation for the joint temperature and polarization analysis, the formalism of the signal and data vectors, $\boldsymbol{s}$ and $\boldsymbol{d}$, signal and noise covariances, $\mathbf{S}$ and $\mathbf{N}$, must be generalized as described below.

4.2.1 Numerical implementation

The CMB signal can be described as a $3N_{\rm pix}$ dimensional vector of harmonic coefficients, $\boldsymbol{s}_\ell = (a_\ell^{\mathcal{T}}, a_\ell^{\mathcal{E}}, a_\ell^{\mathcal{B}})$, for each ℓ, where $\mathcal{T}$, $\mathcal{E}$ and $\mathcal{B}$ imply temperature, electric/gradient, and magnetic/curl, respectively, for discretized sky maps of $N_{\rm pix}$ pixels. The data, $\boldsymbol{d} = (\boldsymbol{d}_\mathcal{I}, \boldsymbol{d}_\mathcal{Q}, \boldsymbol{d}_\mathcal{U})$, are pixelized maps of the Stokes parameters, $\mathcal{I}$, $\mathcal{Q}$ and $\mathcal{U}$, where $\mathcal{I}$ here corresponds to the temperature anisotropy.

Since we assume that the CMB anisotropies are isotropic and Gaussian, the signal covariance $\mathbf{S}$ is diagonal in Fourier space, in the flat-sky approximation. $\mathbf{S}$ becomes a block-diagonal matrix, with a 3×3 sub-matrix, for all multipole moments ℓ, as follows:

$$\mathbf{S}_\ell = \begin{pmatrix} C_\ell^{TT} & C_\ell^{TE} & 0 \\ C_\ell^{TE} & C_\ell^{EE} & 0 \\ 0 & 0 & C_\ell^{BB} \end{pmatrix}, \tag{4.7}$$

with the vanishing cross-spectra, C_ℓ^{TB} and C_ℓ^{EB}, set to zero.

Here, we consider correlated noise, where the noise covariance matrix is a 3×3 block-diagonal matrix, with non-zero elements. The noise covariance matrix has the following block-diagonal structure, for every pixel i:

$$\mathbf{N}_i = \begin{pmatrix} \langle \mathcal{I}\mathcal{I} \rangle & \langle \mathcal{I}\mathcal{Q} \rangle & \langle \mathcal{I}\mathcal{U} \rangle \\ \langle \mathcal{Q}\mathcal{I} \rangle & \langle \mathcal{Q}\mathcal{Q} \rangle & \langle \mathcal{Q}\mathcal{U} \rangle \\ \langle \mathcal{U}\mathcal{I} \rangle & \langle \mathcal{U}\mathcal{Q} \rangle & \langle \mathcal{U}\mathcal{U} \rangle \end{pmatrix}. \tag{4.8}$$

The numerical implementation of the dual messenger algorithm adapted for polarized signal reconstruction is outlined in Algorithm 3. In particular, the computations of $\bar{\mathbf{S}}$ and $\bar{\mathbf{N}}$, via their respective diagonalized forms $\bar{\mathcal{S}}$ and $\bar{\mathcal{N}}$, are clearly illustrated. Since the algorithm requires $\bar{\mathcal{N}}^{-1}$, we circumvent the corner case resulting from $\bar{\mathcal{N}} = \mathbf{0}$, i.e. when $\mathcal{N} = \mathbf{T}$, by imposing the following constraint on $\boldsymbol{t}$ in the correct vector subspace: $\mathfrak{D}\boldsymbol{t}|_{\mathcal{N}=\mathbf{T}} = \mathfrak{D}\boldsymbol{d}|_{\mathcal{N}=\mathbf{T}}$, following the notation set in Algorithm 3.

In Section 4.3.5, we provide a brief description of how the algorithm can be embedded in a modified Jacobi relaxation mechanism to further refine the solution. We implemented the mask by setting the noise covariance for the masked pixels to a numerically high value $\sim \mathcal{O}(10^{10})$. While the formalism above, as illustrated in Algorithm 3, is still valid for dealing with a common temperature and polarization mask, where the inverse noise covariance must be set to zero, it is nevertheless not completely adequate for dealing with different masks. The procedure for solving the messenger Eq. (4.6) must be consequently modified as described in Section 4.2.3. We will implement this exact masking procedure in a forthcoming publication where complex noise covariances will be considered for the analysis of real data sets.

A brief note concerning Fourier transforms: Defining ϕ_ℓ as the angle between the vector $\boldsymbol{\ell}$ and the ℓ_x axis, we use the following convention for the Fourier transforms of the corresponding maps

Algorithm 3 Dual messenger algorithm for polarization

procedure DUAL MESSENGER($\boldsymbol{d}$, $\mathbf{N}$, $\mathbf{S}$, $N_{\rm pix}$, L)
  $\boldsymbol{s}_0 = \text{zeros}(N_{\rm pix}, N_{\rm pix}, 3)$ ▷ Initialize $\boldsymbol{s}$ with zeros
  $\boldsymbol{t}_0 = \boldsymbol{d}$ ▷ Initialize $\boldsymbol{t}$ via an initial guess
  $\mathbf{N} = \mathfrak{D}^\dagger \mathcal{N} \mathfrak{D}$ ▷ Diagonalize $\mathbf{N}$ via basis transformation
  $\alpha = \min(\mathcal{N})$ ▷ Compute the covariance of messenger field $\boldsymbol{t}$, such that $\mathbf{T} = \alpha \mathbb{1}$
  $\bar{\mathcal{N}} = \mathcal{N} - \mathbf{T}$ ▷ Compute the covariance $\bar{\mathcal{N}}$
  $\mathbf{S} = \mathfrak{R}^\dagger \mathcal{S} \mathfrak{R}$ ▷ Diagonalize $\mathbf{S}$ via basis transformation
  $\boldsymbol{\xi} = \boldsymbol{\mu} + \sigma^{-1}\alpha \mathbb{1}$ ▷ Compute $\boldsymbol{\xi}$, $\boldsymbol{\mu} = (\mu_{\mathcal{T}}, \mu_{\mathcal{E}}, \mu_{\mathcal{B}})$
  ▷ $\sigma = N_{\rm pix}^2 / L^4$ is a numerical factor, $\mathcal{F}\mathcal{F}^\dagger = \sigma \mathbb{1}$
  while $\boldsymbol{\xi} \to \sigma^{-1}\alpha\mathbb{1}$ **do**
    $\mathbf{U}_\ell = \boldsymbol{\mu}$ ▷ Assign/update covariance $\mathbf{U}$
    $\bar{\mathcal{S}}_\ell = \Theta(\mathcal{S} - \mathbf{U})_\ell$ ▷ Compute covariance $\bar{\mathcal{S}}$
    repeat
      ▷ Transform to Fourier space, cf. Eqs. (4.9) - (4.13)
      $\hat{\boldsymbol{s}}_{i+1,\ell} = \mathfrak{R}^\dagger \left(\bar{\mathcal{S}} + \mathbf{U}\right)_\ell \left(\bar{\mathcal{S}} + \boldsymbol{\xi}\right)_\ell^{-1} \mathfrak{R} \left(\sigma^{-1} \mathcal{F}^\dagger \boldsymbol{t}_{i,x}\right)_\ell$
      $\boldsymbol{s}_{i+1,x} = \mathcal{F}(\hat{\boldsymbol{s}}_{i+1,\ell})$ ▷ Transform to pixel space
      $\boldsymbol{t}_{i+1,x} = \mathfrak{D}^\dagger \left(\bar{\mathcal{N}}^{-1} + \mathbf{T}^{-1}\right)_x^{-1} \left(\mathfrak{D}\mathbf{T}^{-1} \boldsymbol{s}_{i+1} + \bar{\mathcal{N}}^{-1} \mathfrak{D} \boldsymbol{d}\right)_x$
      $i \leftarrow i + 1$
    until $\|\boldsymbol{s}_i - \boldsymbol{s}_{i-1}\| \,/\, \|\boldsymbol{s}_i\| < \epsilon$
    $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi} \times \beta$ ▷ Cooling scheme for $\boldsymbol{\xi}$
    $\boldsymbol{\mu} \leftarrow \boldsymbol{\xi} - \sigma^{-1}\alpha\mathbb{1}$ ▷ Compute resulting $\boldsymbol{\mu}$
  end while
  $\boldsymbol{s} \to \boldsymbol{s}_{\rm WF}$ ▷ as $\boldsymbol{\xi} \to \sigma^{-1}\alpha\mathbb{1}, \boldsymbol{\mu} \to \boldsymbol{\nu} = \mathbf{0}$
  return $\boldsymbol{s}_{\rm WF}$
end procedure

of the Stokes parameters:

$$\widehat{\mathcal{I}}_\ell = \mathcal{F}^\dagger \mathcal{I}_x = \left(\frac{1}{L^2}\right) \sum_x \omega^{-x\cdot k} \mathcal{I}_x \tag{4.9}$$

$$\widehat{\mathcal{Q}}_\ell = \mathcal{F}^\dagger \mathcal{Q}_x = \left(\frac{1}{L^2}\right) \sum_x \left(a_\ell^{\mathcal{E}} \cos 2\phi_\ell - a_\ell^{\mathcal{B}} \sin 2\phi_\ell\right) \omega^{-x\cdot k} \tag{4.10}$$

$$\widehat{\mathcal{U}}_\ell = \mathcal{F}^\dagger \mathcal{U}_x = \left(\frac{1}{L^2}\right) \sum_x \left(a_\ell^{\mathcal{E}} \sin 2\phi_\ell + a_\ell^{\mathcal{B}} \cos 2\phi_\ell\right) \omega^{-x\cdot k} \tag{4.11}$$

for $\omega = \exp(i2\pi/N_{\rm pix})$ and an observed sky patch of angular extent L, such that, for the harmonic coefficients, we have:

$$a_\ell^{\mathcal{E}} = \widehat{\mathcal{Q}}_\ell \cos 2\phi_\ell + \widehat{\mathcal{U}}_\ell \sin 2\phi_\ell \tag{4.12}$$

$$a_\ell^{\mathcal{B}} = -\widehat{\mathcal{Q}}_\ell \sin 2\phi_\ell + \widehat{\mathcal{U}}_\ell \cos 2\phi_\ell, \tag{4.13}$$

and trivially, $a_\ell^{\mathcal{T}} = \widehat{\mathcal{I}}_\ell$. The corresponding inverse Fourier transform, $\mathcal{F}$, satisfies $\mathcal{F}\mathcal{F}^\dagger = \sigma\mathbb{1}$, where $\sigma = N_{\rm pix}^2/L^4$. From a linear algebraic standpoint, we have two operators $\mathcal{F}^\dagger$ and $\mathcal{F}$ that act on the vector spaces $(\mathcal{I}, \mathcal{Q}, \mathcal{U})$ and $(a_\ell^{\mathcal{T}}, a_\ell^{\mathcal{E}}, a_\ell^{\mathcal{B}})$, respectively, while still satisfying the orthogonality condition, $\mathcal{F}\mathcal{F}^\dagger = \sigma\mathbb{1}$.

4.2.2 Constrained realizations

It is well-known that the application of a Wiener filter leads to a reduced signal covariance,

$$\langle \boldsymbol{s}_{\rm WF} \boldsymbol{s}_{\rm WF}^{\dagger} \rangle = \mathbf{S}(\mathbf{S}+\mathbf{N})^{-1}\mathbf{S}, \tag{4.14}$$

by suppressing the power related to the noise. We therefore need to add a fluctuation vector $\boldsymbol{f}$ to the Wiener-filtered maps to obtain signals that are consistent with the observed data, i.e. having the correct covariance properties, according to

$$\langle \boldsymbol{f}\boldsymbol{f}^{\dagger} \rangle = (\mathbf{S}^{-1}+\mathbf{N}^{-1})^{-1}. \tag{4.15}$$

The dual messenger algorithm presented above can be augmented to generate full-sky, noiseless maps by making some minor adjustments. We simulate a fake signal $\hat{\boldsymbol{s}}$ with the (prior) signal covariance $\mathbf{S}$ assumed for the Wiener filter, which is subsequently contaminated with noise with covariance $\mathbf{N}$ to generate a fake data set $\hat{\boldsymbol{d}}$. We obtain constrained realizations (e.g. Hoffman and Ribak, 1991) via

$$\begin{aligned} \boldsymbol{s}_{\rm CR} &= \boldsymbol{s}_{\rm WF} + \boldsymbol{f} \\ &= \widetilde{\mathcal{W}}\, \boldsymbol{d} + (\hat{\boldsymbol{s}} - \widetilde{\mathcal{W}}\, \hat{\boldsymbol{d}}) \\ &= \widetilde{\mathcal{W}}\, (\boldsymbol{d} - \hat{\boldsymbol{d}}) + \hat{\boldsymbol{s}}, \end{aligned} \tag{4.16}$$

using the Wiener-filtered map, $\boldsymbol{s}_{\rm WF} = \widetilde{\mathcal{W}}\, \boldsymbol{d}$, and the fluctuation map, $\boldsymbol{f} = \hat{\boldsymbol{s}} - \widetilde{\mathcal{W}}\, \hat{\boldsymbol{d}}$. The input to the dual messenger algorithm is therefore $(\boldsymbol{d} - \hat{\boldsymbol{d}})$, such that only one execution of the algorithm is sufficient to provide a constrained realization.

Now, the algorithm yields a solution that is a random realization of a fluctuation map with the correct signal properties. For a single $\boldsymbol{d}$, the Wiener filter samples are plausible signal realizations that optimally take into account the constraints on the signal from the data and are hence known as constrained realizations.

4.2.3 Incomplete sky coverage

To account for different temperature and polarization masks, the procedure for solving the messenger Eq. (4.6) must be modified, although the covariance $\mathbf{T}$ of the messenger field $\boldsymbol{t}$ is still computed as given in Algorithm 3. The noise covariance $\mathbf{N}$ can be written as $\mathbf{N} = \mathbf{\Sigma C \Sigma} = \mathbf{\Sigma P}^{\dagger} \mathbf{\Delta P \Sigma}$, where $\mathbf{\Sigma}$ is a diagonal matrix with the eigenvalues $\{\sigma^i_{\mathcal{I}}, \sigma^i_{\mathcal{Q}}, \sigma^i_{\mathcal{U}}\}$ corresponding to the noise amplitudes for the $i^{\rm th}$ pixel, with the orthonormal decomposition of $\mathbf{C}$ yielding a diagonal matrix $\mathbf{\Delta}$. We then obtain $\bar{\mathbf{N}} = \mathbf{N} - \mathbf{T}$ as follows:

$$\begin{aligned} \bar{\mathbf{N}} &= \mathbf{\Sigma P}^{\dagger} \mathbf{\Delta P \Sigma} - \mathbf{T} \\ &= \mathbf{\Sigma P}^{\dagger} (\mathbf{\Delta} - \alpha \mathbf{P \Sigma}^{-2} \mathbf{P}^{\dagger}) \mathbf{P \Sigma}, \end{aligned} \tag{4.17}$$

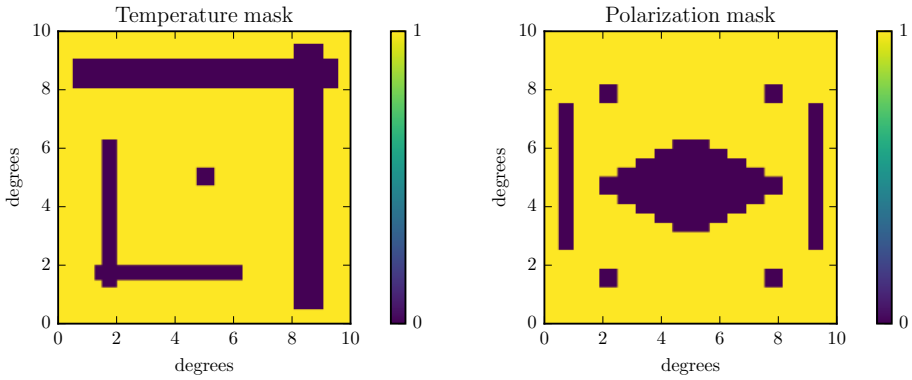


FIGURE 4.1: The temperature and polarization masks implemented in the data analysis, corresponding to sky fractions of $f_{\rm sky}^{\mathcal{T}} = 0.78$ and $f_{\rm sky}^{\mathcal{P}} = 0.82$, respectively.

where, as a reminder, $\mathbf{T} = \alpha\mathbb{1}$, with α being the smallest eigenvalue of $\mathbf{N}$. To solve the messenger Eq. (4.6), we require the inverse $\bar{\mathbf{N}}^{-1}$,

$$\bar{\mathbf{N}}^{-1} = \mathbf{\Sigma}^{-1}\mathbf{P}^{\dagger}(\mathbf{\Delta} - \alpha\mathbf{P}\mathbf{\Sigma}^{-2}\mathbf{P}^{\dagger})^{-1}\mathbf{P}\mathbf{\Sigma}^{-1}, \tag{4.18}$$

such that $\bar{\mathbf{N}}^{-1}$ has a block-diagonal structure in pixel space. We obtain the solution for the messenger field by simply evaluating Eq. (4.6) in pixel space,

$$\boldsymbol{t}_x = \left(\bar{\mathbf{N}}^{-1} + \mathbf{T}^{-1}\right)_x^{-1}\left(\mathbf{T}_x^{-1}\mathcal{F}\boldsymbol{s}_\ell + \bar{\mathbf{N}}_x^{-1}\boldsymbol{d}_x\right). \tag{4.19}$$

We implement the temperature and polarization masks by increasing the noise variance to infinity for masked pixels, or numerically by setting the inverse noise covariance to zero. This is achieved by setting $\mathbf{\Sigma}^{-1}|_{\rm mask} = 0$, subsequently ensuring that data from masked regions do not contaminate the messenger field.

4.3 Application to CMB polarization

In this section, we demonstrate the application of the dual messenger algorithm to a mock CMB polarization data set and highlight the contrast in performance relative to a conventional PCG method.

4.3.1 Polarized map simulation

To simulate joint temperature and polarization maps in the flat-sky approximation, we make use of a Cholesky decomposition to generate realizations of $a_\ell^{\mathcal{T}}$, $a_\ell^{\mathcal{E}}$ and $a_\ell^{\mathcal{B}}$ signals with the correct covariance properties (cf. Eq. (4.7)), taking into account the correlation between CMB temperature anisotropy and polarization. We made use of CAMB[1] (Lewis, Challinor, and Lasenby, 2000) to generate the input angular power spectra, C_ℓ^{TT}, C_ℓ^{EE}, C_ℓ^{BB} and C_ℓ^{TE}, from which the corresponding CMB signals are drawn. We assume a standard ΛCDM cosmology with the set of cosmological parameters ($\Omega_{\rm m} = 0.32$, $\Omega_\Lambda = 0.69$, $\Omega_{\rm b} = 0.05$, $h = 0.67$, $\sigma_8 = 0.83$, $n_{\rm s} = 0.97$) from Planck (Planck

[1] http://camb.info

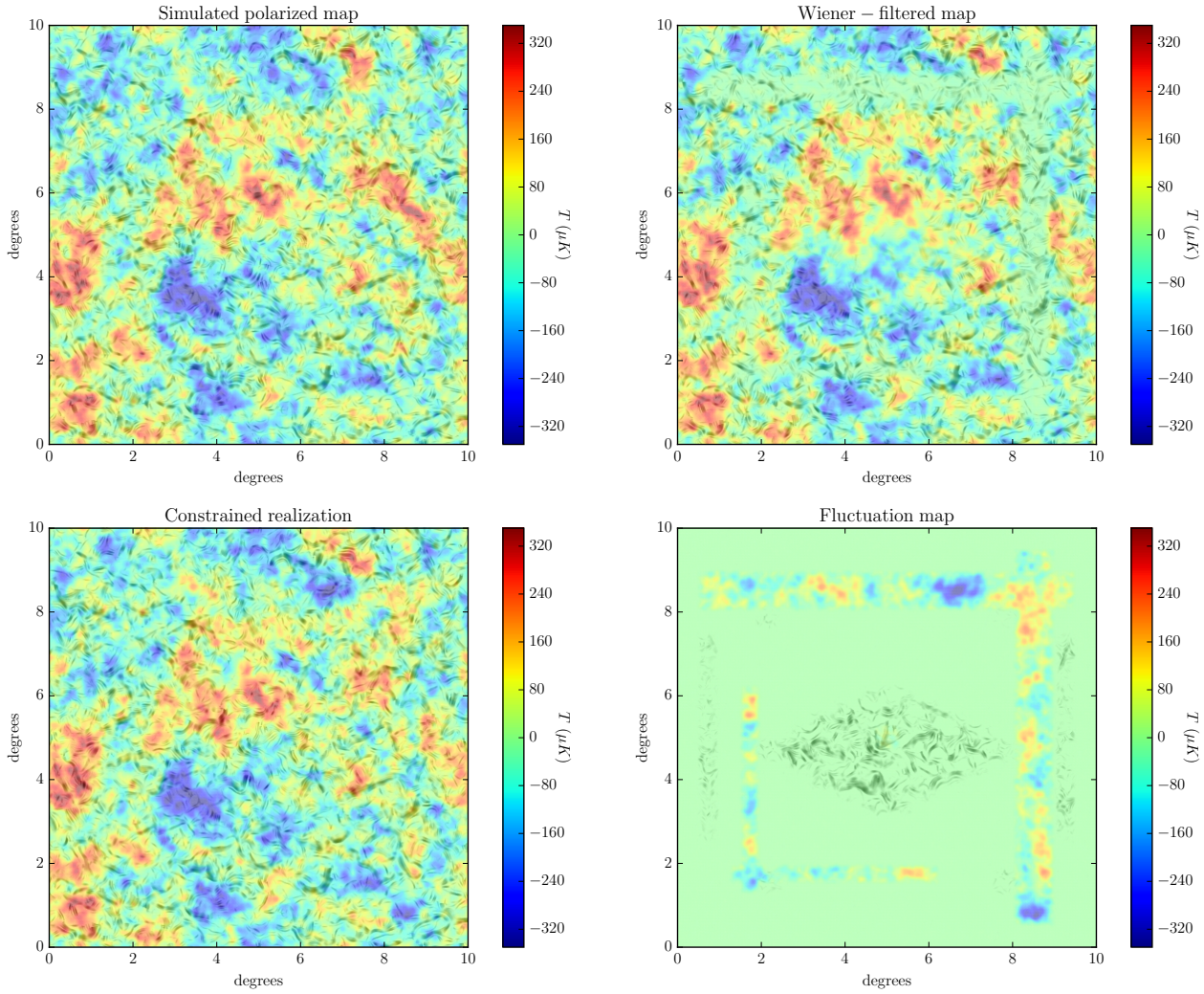


FIGURE 4.2: *Top panel:* The simulated and Wiener-filtered polarization maps. The images above depict temperature as colour and polarization overlaid as a pattern of stripes. The alignment of the stripes indicates the direction of polarization while the level of transparency corresponds to the polarization intensity, with the darker regions implying stronger polarization. The Wiener-filtered map is the maximum *a posteriori* reconstruction of the signal from the data. *Bottom panel:* The fluctuation map and constrained realization. To compensate for the power loss due to noise and masked sky, the Wiener-filtered map is augmented with a fluctuation map, thereby yielding a full-sky, noiseless sample, i.e. constrained realization, with the correct signal properties. The temperature and polarization masks applied are visible in the fluctuation map.

Collaboration et al., 2016b). We can then construct the input $\mathcal{Q}$ and $\mathcal{U}$ maps by transforming realizations of $\mathcal{E}$ and $\mathcal{B}$ signals (cf. Section IIc in Bunn et al., 2003) over flat-sky patches with angular extent, $L = 10.0$ degrees, and grid resolution, $N_{\rm pix} = 1024^2$. The input Stokes parameters' maps are subsequently contaminated with correlated noise, according to the noise covariance given by Eq. (4.8), with a noise amplitude of 4.0 μK, typical of high-sensitivity CMB experiments tailored for the detection of $\mathcal{B}$ modes. The temperature and polarization masks implemented, corresponding to sky fractions of $f_{\rm sky}^{\mathcal{T}} = 0.78$ and $f_{\rm sky}^{\mathcal{P}} = 0.82$, respectively, are depicted in Fig. 4.1.

4.3.2 CMB polarization analysis

We showcase the application of the dual messenger algorithm in polarization data analysis, while drawing a comparison to the corresponding solution provided by a PCG method. For the PCG computation, we make use of the preconditioner provided in Appendix B.1.

As per standard data analysis pipelines, we must filter out the noise and reconstruct a clean map via a Wiener filtering algorithm. We implement the dual messenger scheme described in the above sections to compute the Wiener-filtered $\mathcal{I}$, $\mathcal{Q}$ and $\mathcal{U}$ maps of the Stokes parameters. The algorithm loops through the iterations until the fractional difference between successive iterations has reached a sufficiently low value. Here, we implement Cauchy's "weak" criterion for convergence, $\|\boldsymbol{s}_{i+1} - \boldsymbol{s}_i\| \,/\, \|\boldsymbol{s}_i\| < \epsilon$, where $\epsilon = 10^{-6}$. We adopt the same cooling scheme as in KLW17, whereby we reduce $\boldsymbol{\xi}$ by a constant factor, i.e. $\boldsymbol{\xi} \to \boldsymbol{\xi}\beta$, where $\beta = 3/4$, until $\boldsymbol{\xi} \to \sigma^{-1}\alpha\mathbb{1}$. Since we are not solving the desired system of equations initially, we can also implement a cooling scheme for the threshold ϵ. This speeds up the computation significantly by around a factor of three. We relax the convergence criterion by a factor of η for each μ, thereby reducing ϵ from 10^{-4} to 10^{-6}, where we choose $\eta = 1.1$.

The simulated and corresponding Wiener-filtered maps are displayed in the top row of Fig. 4.2. The polarization intensity is given by $(\mathcal{Q}^2 + \mathcal{U}^2)^{1/2}$, while the direction of polarization corresponds to $\arctan(\mathcal{Q}/\mathcal{U})/2$. The Wiener-filtered map, as the maximum *a posteriori* reconstruction, represents the CMB signal content of the data, with the reconstruction of the large-scale modes in the masked areas, based on the information content of the observed sky regions, being a natural consequence of Wiener filtering. As described in Section 4.16, the algorithm generates a fluctuation map to compensate for the suppressed power on the small scales due to the noise and masked regions of the sky. The resulting constrained realization map, after combining the Wiener-filtered and fluctuation maps, has the correct statistical properties consistent with the simulated input map, as illustrated in Fig. 4.2.

The power spectra of the dual messenger reconstruction for temperature and polarization are provided in the left panel of Fig. 4.3, thereby showing that the Wiener-filtered maps can be augmented to constrained realizations, resulting in unbiased power spectra. In the high signal-to-noise regime, the power spectrum of the full-sky, noiseless map is determined by the data, while in the low signal-to-noise regime, by the assumed power spectrum. The main issue with the PCG reconstruction, depicted in the right panel, is that it fails to raise the power associated with $\mathcal{B}$ modes, with the $\mathcal{E}$-mode power spectrum also displaying some artefacts, especially on the large and intermediate scales. The low quality of the PCG reconstruction is evident in the behaviour of the convergence diagnostics for the PCG solution as discussed below.

To illustrate the convergence behaviour of the algorithms, we provide the variations of the residual errors given by $\|\boldsymbol{s}_{i+1} - \boldsymbol{s}_i\| \,/\, \|\boldsymbol{s}_i\|$ and $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| \,/\, \|\boldsymbol{y}\|$, for a linear system of equations given by $\mathcal{A}\boldsymbol{x} = \boldsymbol{y}$, in Fig. 4.4. The residual error given by the latter criterion better characterizes the accuracy of the final solution. We find that, for the dual messenger scheme, this residual error always decreases as the iterations proceed, demonstrating the unconditional stability of the algorithm. The oscillations in the residual error provided by the Cauchy criterion, as explained in KLW17, are due

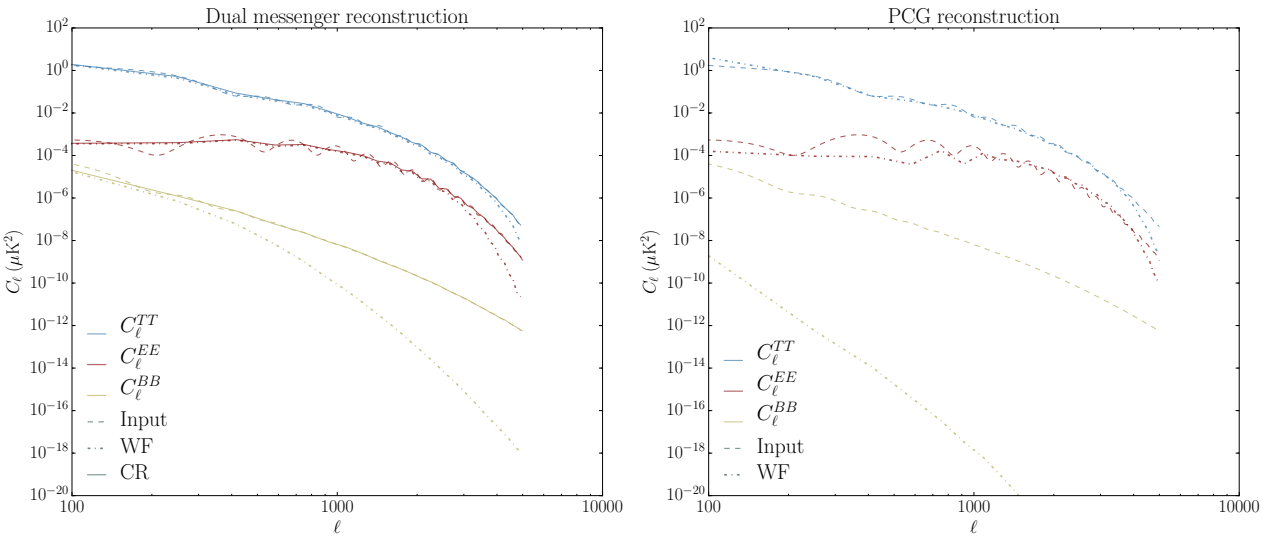


FIGURE 4.3: Reconstructed (binned) power spectra computed using the dual messenger and PCG algorithms. The dashed lines indicate the input angular power spectra from which the polarized CMB signals are drawn, while the reconstructions from the Wiener-filtered (WF) and constrained realization (CR) maps are depicted by the corresponding dotted and solid lines. *Left panel:* We demonstrate that a constrained realization obtained via the dual messenger algorithm, as a combination of the Wiener-filtered and fluctuation maps, yields unbiased power spectra compared to the corresponding input power spectra. *Right panel:* The PCG reconstruction strikingly fails to find enough $\mathcal{B}$ modes, while the $\mathcal{E}$-mode power spectrum also displays some artefacts, especially on large and intermediate scales, demonstrating the unreliability of the power spectra reconstruction. Remarkably, the dual messenger reconstruction, even with a more lenient $\epsilon = 10^{-4}$, is visually undistinguishable from that displayed in the left panel with $\epsilon = 10^{-6}$, further highlighting the stark contrast in performance between these two methods.

to the cooling scheme implemented. The peaks result from the transitions in the systems of equations with the varying covariance of the auxiliary field, with the residual error always dropping sharply after each peak. However, the oscillations in the residual errors for the PCG implementation indicate its vulnerability to instabilities sourced by numerical noise. The residual error in the final PCG solution does not drop below 10^{-3}, three orders of magnitude higher than that achieved by the dual messenger solution, implying that the accuracy of the PCG solution remains nevertheless significantly inadequate.

The corresponding variation in the χ^2 is also displayed in the right panel of Fig. 4.4. The $\chi^2_{\rm DM}$ of the dual messenger solution drops rapidly in accordance with the cooling scheme implemented and the final value matches $\langle \chi^2_{\rm d.o.f} \rangle$, the expectation value of the χ^2, given by the number of degrees of freedom (d.o.f), for the final solution. This is not the case, however, for the PCG solution, with $\chi^2_{\rm PCG}$ failing to attain the expected level. The convergence diagnostics discussed above demonstrate the effectiveness and quality of polarized signal reconstruction via the dual messenger algorithm.

The solution may be further refined via the modified Jacobi relaxation scheme described in Section 4.3.5. This Jacobi adaptation was found to be stable with $\epsilon = 10^{-5}$ and lower, but the extra computational cost required, for this specific problem, is not justified due to the inherently high quality of the dual messenger solution. This may nevertheless be of general interest for other Wiener filtering applications, to yield adequate solutions for a significantly reduced number of iterations.

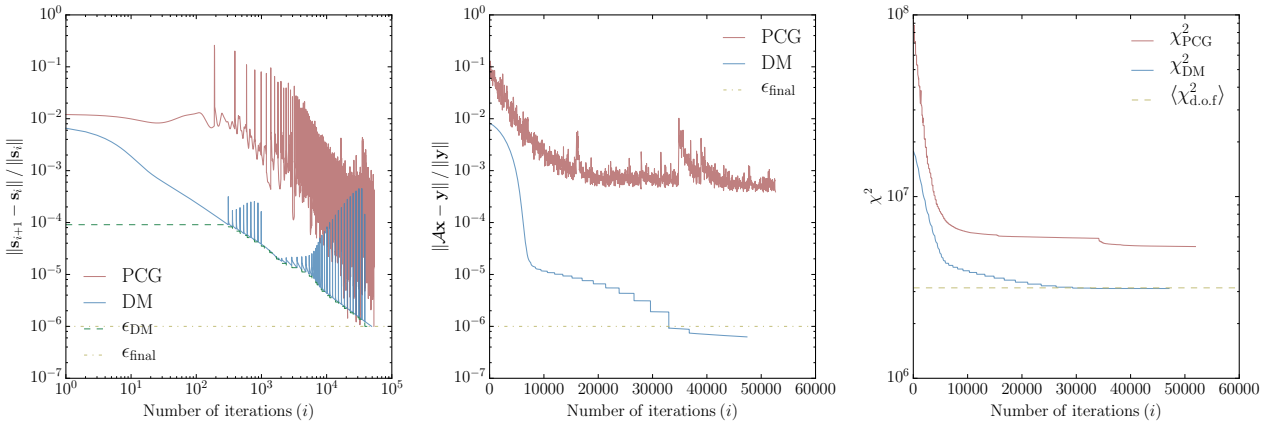


FIGURE 4.4: Convergence diagnostics corresponding to the dual messenger (DM) and PCG reconstructions. *Left panel:* Variation of the residual error, given by the Cauchy criterion, with number of iterations. The cooling scheme for the convergence threshold ϵ, indicated by the dashed lines, allows for faster convergence to the final tolerance desired. *Middle panel:* The corresponding variation of the residual error, given by $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| / \|\boldsymbol{y}\|$, with number of iterations. The monotonic decrease in this residual error demonstrates the unconditional stability of the dual messenger algorithm. In contrast, for the PCG counterpart, this residual error does not drop below 10^{-3}, three orders of magnitude above that attained by the dual messenger solution, implying that the accuracy of the PCG solution is significantly lower. The PCG implementation is also susceptible to numerical instabilities, as indicated by the oscillations in the residual error. *Right panel:* Variation of χ^2 with number of iterations. $\chi^2_{\rm DM}$ drops rapidly and finally matches the expectation value of the χ^2 of the final solution, $\langle\chi^2_{\rm d.o.f}\rangle$, given by the number of degrees of freedom, indicated by the dashed line. $\chi^2_{\rm PCG}$, however, fails to achieve this expected level, again highlighting the unreliability and poor quality of the PCG solution.

While the choice of another preconditioner may yield an improved solution for the PCG implementation, in practice, it is highly non-trivial to construct an effective preconditioner, especially when dealing with an ill-conditioned system, as considered in this work. Moreover, as illustrated above, the PCG method remains susceptible to numerical instabilities.

4.3.3 Separation of $\mathcal{E}$ and $\mathcal{B}$ modes

We briefly discuss the so-called $\mathcal{E}$-$\mathcal{B}$ coupling problem, whereby the $\mathcal{E}$-mode power leaks into the much smaller $\mathcal{B}$-mode power, that plagues well-known approximate methods such as the pseudo-C_ℓ methods (e.g. Bond, Jaffe, and Knox, 1998). This aliasing of power is due to the fact that the spherical harmonics are not orthogonal on a sky with masked regions (e.g. Zaldarriaga, 2001; Bunn, 2002a; Bunn et al., 2003). However, exact methods such as Gibbs sampling circumvent this predicament.

The standard procedure of obtaining a complete sky sample involves two steps: A Wiener filter is first used to filter out the noise and reconstruct a clean map. Second, the power loss due to noise and incomplete sky coverage is compensated by a random fluctuation term. The combination of the Wiener-filtered and fluctuation maps subsequently yields a full-sky, noiseless sample that is consistent with observations, i.e. constrained realization (cf. Section 4.2.2). The corresponding $\mathcal{E}$ and $\mathcal{B}$ maps resulting from the dual messenger algorithm are depicted in Fig. 4.5. The $\mathcal{E}$-$\mathcal{B}$ coupling issue no longer arises since we now have a full-sky sample. Since the Wiener filter depends on the choice of an input power spectrum, a Gibbs sampling scheme, where samples of power spectrum would

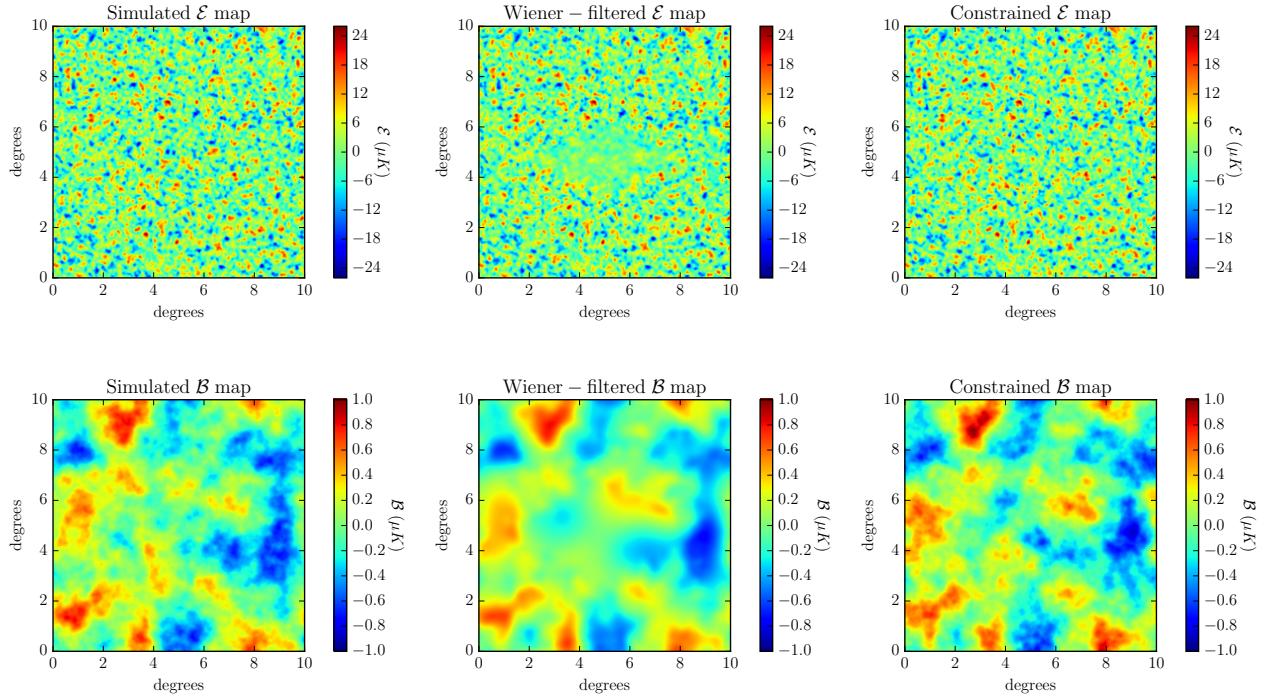


FIGURE 4.5: The simulated, Wiener-filtered and constrained $\mathcal{E}$ and $\mathcal{B}$ maps. In the top row, the Wiener filter characteristically extends the signal into the masked regions; this works especially well in the small excized areas of the map, usually required for dealing with point foreground sources. Due to the non-vanishing cross spectrum C_ℓ^{TE}, the polarization reconstruction in masked areas is rather efficient, unless the temperature and polarization masks overlap. In the bottom row, the low-amplitude $\mathcal{B}$ modes are inevitably smoothed out due to the suppression of the small-scale power. The constrained maps represent plausible realizations of de-noised full-sky $\mathcal{E}$ and $\mathcal{B}$ maps.

be drawn conditional on the data itself, would simultaneously yield the posterior probability distributions of constrained realizations and their power spectra. This Bayesian framework therefore allows for a statistically optimal separation of $\mathcal{E}$ and $\mathcal{B}$ modes in terms of power spectra.

The dual messenger algorithm can thus be incorporated in such a Gibbs sampling scheme, for instance, as described by Larson et al. (2007), for optimal power spectrum inference from high-resolution polarized CMB data sets. Larson et al. (2007) implemented a PCG method, while considering uncorrelated noise only, but the preconditioner was limited by the signal-to-noise ratio of the data and was not sufficiently efficient for joint analysis of temperature and polarization data. We have demonstrated that the dual messenger technique is not hindered by such limitations and maintains its efficiency in performing the two key steps described above, even for high-resolution maps, while accounting for correlated noise, as quantitatively substantiated in the next section.

A recent work by Bunn and Wandelt (2017), whereby they demonstrate that pure $\mathcal{E}$ and $\mathcal{B}$ maps, free from any cross-contamination, may be obtained via a Wiener filtering approach, provides further motivation for the polarized Wiener filter. This approach provides real-space maps of the $\mathcal{E}$ and $\mathcal{B}$ modes while conventional methods are limited to power spectrum estimation (e.g. Challinor and Chon, 2005; Smith, 2006b; Smith, 2006a; Smith and Zaldarriaga, 2007; Grain, Tristram, and Stompor, 2009) or produce only the derivatives of the polarization maps (e.g. Kim and Naselsky, 2010; Zhao and Baskaran, 2010; Kim, 2011; Bowyer, Jaffe, and Novikov, 2011). Other wavelet-based reconstruction methods (e.g. Cao and Fang, 2009; Rogers et al., 2016; Leistedt et al., 2017) must

be carefully adapted for the specific problem being investigated. The dual messenger algorithm can therefore be optimized to yield pure $\mathcal{E}$ and $\mathcal{B}$ maps via the framework proposed by Bunn and Wandelt (2017).

4.3.4 Computational performance

TABLE 4.1: This table provides the computational performance diagnostics for a series of convergence criteria (ϵ).

ϵ	i	$t(\mathrm{mins})$	χ^2_{final}	$\frac{\|\mathcal{A}x_{\mathrm{final}}-y\|}{\|y\|}$
10^{-4}	304	10	3.1300×10^6	3.1×10^{-5}
10^{-5}	3589	31	3.1207×10^6	3.9×10^{-6}
10^{-6}	49076	309	3.1196×10^6	6.3×10^{-7}

The execution times and number of iterations required by the dual messenger algorithm, for $N_{\mathrm{pix}} = 1024^2$, to run to completion on a single core of a standard workstation, Intel Core i5-4690 CPU (3.50 GHz), for a series of convergence criteria ϵ, are provided in Table 4.1. The χ^2 and residual errors corresponding to the final solutions are also displayed. While the polarization analysis above was carried out with $\epsilon = 10^{-6}$, such a stringent convergence criterion is not required by most Wiener filtering applications, so execution time is drastically reduced. Code parallelization is another key option to speed up the execution for high-resolution data sets.

It is worth pointing out that even with more lenient criteria, the algorithm provides decent results. For instance, executing the above computation with $\epsilon = 10^{-4}$ requires only 304 iterations for convergence, corresponding to few minutes of computation time, and already correctly reconstructs the power spectra on all scales, visually indistinguishable from that displayed in the left panel of Fig. 3.2. The point at which the threshold $|\Delta\chi^2| \lesssim \sigma_{\chi^2}$ is attained, where $|\Delta\chi^2| = |\chi^2 - \langle\chi^2_{\mathrm{d.o.f}}\rangle|$ and $\sigma_{\chi^2} \sim \sqrt{2n_{\mathrm{d.o.f.}}}$ for $n_{\mathrm{d.o.f.}}$ degrees of freedom, is another performance indicator. The χ^2 for the $\epsilon = 10^{-4}$ run drops below this level in 246 iterations, again highlighting the reliability of the solution achievable with such a low number of iterations. This remarkable performance of the dual messenger technique is especially significant for applications involving exact inference, such as Gibbs sampling.

There is a minor caveat, nonetheless, concerning the current iteration scheme with the relaxed convergence threshold. When dealing with masked regions and very low noise amplitude per pixel, this may result in a marginal lack of convergence on the largest scales for the temperature map, where more iterations are required to improve the reconstruction. The $\mathcal{E}$ and $\mathcal{B}$ maps and their associated power spectra are, however, unaffected by this issue of convergence. We therefore plan to devise an enhanced iteration scheme, such as an adaptive dual messenger algorithm, to improve the treatment of the temperature mask when dealing with real data sets on the sphere.

The algorithmic complexity of the polarized version of the dual messenger method now reflects the fact that the number of Fourier transforms required per iteration is increased by a factor of three. The algorithm now requires six Fourier transforms, $\mathcal{O}(N_{\mathrm{pix}} \log N_{\mathrm{pix}})$, and two scalar multiplications

corresponding to algebraic operations of $\mathcal{O}(3N_{\rm pix})$, per iteration. In terms of memory requirements, temporary storage of two vectors of dimension $3N_{\rm pix}$ is required. In contrast, the PCG algorithm requires nine Fourier transforms and ten scalar multiplications, per iteration, while eight vectors of size $3N_{\rm pix}$ must be temporarily stored in memory.

4.3.5 Refinement via Jacobi relaxation

We provide a brief review of the iterative Jacobi technique and describe how it can be implemented with the dual messenger algorithm for a refinement of the solution.

Jacobi relaxation is a well-known iterative technique to solve linear systems of equations (Jacobi, 1845; Saad, 2003). The conventional Jacobi method cannot be applied to the ill-conditioned system investigated here. We therefore augment the dual messenger algorithm via a modified Jacobi relaxation scheme.

The standard Jacobi iterative method for the solution at two consecutive iterations, denoted by n, is as follows:

$$\begin{aligned} \boldsymbol{s}_{\rm WF}^{n+1} &= \boldsymbol{s}_{\rm WF}^{n} + \widetilde{\mathcal{W}}\,(\boldsymbol{d} - \mathcal{W}^{+}\boldsymbol{s}_{\rm WF}^{n}) \\ &= \boldsymbol{s}_{\rm WF}^{n} + \widetilde{\mathcal{W}}\,\boldsymbol{d} - \widetilde{\mathcal{W}}\,\hat{\boldsymbol{d}}, \end{aligned} \tag{4.20}$$

where the $\mathcal{W}^{+}$ operator is given by $\mathcal{W}^{+} = \mathbf{N}^{-1}(\mathbf{S}^{-1} + \mathbf{N}^{-1})$ and $\hat{\boldsymbol{d}} = \mathcal{W}^{+}\boldsymbol{s}_{\rm WF}^{n}$. We split the conventional Jacobi operation into two parts; a usual dual messenger ($\widetilde{\mathcal{W}}$) operation on the data $\boldsymbol{d}$ and another $\widetilde{\mathcal{W}}$ operation on the Jacobi correction term. For masked regions, where formally no inverse of $\mathbf{N}$ exists, $\mathcal{W}^{+}$ is the pseudo-inverse, i.e. $\mathcal{W}\mathcal{W}^{+} = \Pi$, where Π is a projector.

We first compute the usual dual messenger solution, keeping track of the number of iterations required for each signal covariance truncation μ, then drive the second operation using this prescription for the number of iterations to ensure that the $\widetilde{\mathcal{W}}$ operator is invariant. For numerical stability, we also modify the second Eq. (4.6), after substituting $\boldsymbol{d} \to \hat{\boldsymbol{d}}$, as follows:

$$\boldsymbol{t}_{i+1,x} = \mathfrak{D}^{\dagger}\left[\mathbf{T}\mathfrak{D}z^{n} + (\mathcal{N}^{-1} + \mathbf{T}^{-1})\mathfrak{D}\mathbf{T}^{-1}\boldsymbol{s}_{i+1}\right]_{x}^{-1}, \tag{4.21}$$

where

$$z_{x}^{n} = \mathcal{F}^{\dagger\,-1}\mathfrak{R}^{\dagger}\mathcal{S}_{\ell}^{-1/2}\left(\mathbb{1} + \mathcal{S}_{\ell}^{1/2}\mathfrak{R}\mathcal{F}^{\dagger}\mathcal{N}_{x}^{-1}\mathcal{F}\mathfrak{R}^{\dagger}\mathcal{S}_{\ell}^{1/2}\right)\left(\mathcal{S}^{-1/2}\mathfrak{R}\hat{s}_{\rm WF}^{n}\right)_{\ell}, \tag{4.22}$$

including all basis transformations, consistent with the notation employed in Algorithm 3. The above rewriting ensures that the equation is free from any singularities.

4.4 Conclusions

We demonstrated that our recently proposed dual messenger algorithm maintains its efficiency for the data analysis of joint temperature and polarization maps with correlated noise and reduced sky coverage. This preconditioner-free method deals with ill-conditioned systems, typically encountered in the data analysis of polarized CMB maps, with relative ease. This is particularly important as conventional conjugate gradient solvers fail to deal efficiently with this increase in the condition

number of the covariance matrices, as illustrated by the failure of the PCG method implemented in this work to converge to a sensible solution.

The dual messenger algorithm can be conveniently adapted to generate constrained Gaussian realizations of the CMB sky, which is an essential component of present-day CMB data analyses. The reconstruction of $\mathcal{E}$ and $\mathcal{B}$ maps from the resulting full-sky, noiseless sample avoids the leakage issue that arises due to partial sky coverage. The algorithm can also be extended to perform pure $\mathcal{E}/\mathcal{B}$ decomposition via the approach proposed by Bunn and Wandelt (2017), as illustrated in the following chapter.

The implementation of the dual messenger algorithm is straightforward, while being numerically robust and flexible. We also described how this method can be optimized in a modified Jacobi relaxation scheme to further refine the solution. In Chapter 5, we demonstrate the extension of our dual messenger method for applications on the sphere and describe how the formalism can be naturally augmented via a further level of sophistication to account for more complex and realistic noise models, such as modulated and (spatially) correlated noise, resulting from the scanning strategy of the instrument.

We intend to showcase the application on real data sets to further validate this algorithm as a viable signal reconstruction tool for modern CMB data sets. The dual messenger technique can be further refined to reduce execution time. A potentially significant performance upgrade, especially on the sphere, is provided by the hierarchical framework of the dual messenger algorithm: We adapt the working resolution progressively such that the Nyquist frequency is always slightly above the current ℓ_{iter} considered in the algorithm so as to reduce the cost of harmonic transforms.

We are hopeful that the development of the dual messenger method will ultimately render exact global Bayesian analyses of high-resolution and high-sensitivity CMB observations numerically tractable and computationally efficient. This high-performance algorithm is particularly adapted to cope with the complex numerical challenges posed by modern data sets and is therefore relevant for current and future high-resolution CMB missions such as Planck, South Pole Telescope, Advanced ACTPol, POLARBEAR, QUBIC, Simons Observatory and CMB-S4. Whilst we have demonstrated the application of this algorithm in a cosmological context, it remains nevertheless relevant for spin field reconstruction in a general framework.

Chapter 5

Pure $\mathcal{E}/\mathcal{B}$ decomposition of CMB maps with anisotropic correlated noise

The work presented in this chapter is based on Kodi Ramanah, Lavaux, and Wandelt (2019).

5.1 Introduction

Cosmological inference from observations of the CMB polarization, as mentioned in Chapter 4, necessitates the separation of the contributions of the gradient and curl (or $\mathcal{E}$ and $\mathcal{B}$) components of the polarization signal to the data. These scalar $\mathcal{E}$ and pseudo-scalar $\mathcal{B}$ modes correspond to the spin-2 analogues of curl-free and divergence-free vector fields (cf. Chapter 1.4.3), respectively, with a polarization map being represented as the sum of both components (Zaldarriaga and Seljak, 1997; Seljak and Zaldarriaga, 1997; Kamionkowski, Kosowsky, and Stebbins, 1997b; Kamionkowski, Kosowsky, and Stebbins, 1997a).

The decomposition of the $\mathcal{E}$ and $\mathcal{B}$ modes on a partial sky is highly non-trivial due to the induced leakage between the two modes. Masked regions produce a discontinuity at the edges of the map and this results in a mixing of $\mathcal{E}$ and $\mathcal{B}$ modes, yielding ambiguous modes. Such modes can be sourced by either $\mathcal{E}$ or $\mathcal{B}$ components and cannot be uniquely assigned (Zaldarriaga, 2001; Lewis, Challinor, and Turok, 2002; Lewis, 2003; Bunn, 2002a; Bunn, 2002b; Bunn et al., 2003; Bunn, 2011). This is a highly prevalent issue due to most observations being made on an incomplete sky or full-sky maps being subjected to additional masking to reduce foreground contamination. Since the $\mathcal{E}$-mode power spectrum is much larger than that of $\mathcal{B}$ modes, the ambiguous modes significantly increase the variance of the $\mathcal{B}$ modes, resulting in a spurious measurement of the $\mathcal{B}$-mode power spectrum. The detection of the inflationary gravitational waves is especially challenging due to their relatively small amplitude and hence, efficient methods for pure $\mathcal{E}/\mathcal{B}$ decomposition are essential for extracting the cosmological information from CMB polarization data.

Several approaches for pure $\mathcal{E}/\mathcal{B}$ decomposition are described in the literature. While some techniques yield real-space maps of the derivatives of the polarization maps (e.g. Kim and Naselsky, 2010; Zhao and Baskaran, 2010; Kim, 2011; Bowyer, Jaffe, and Novikov, 2011), others are limited to power spectrum estimation via the construction of an eigenbasis for the pure-ambiguous decomposition (e.g. Challinor and Chon, 2005; Smith, 2006b; Smith, 2006a; Smith and Zaldarriaga, 2007; Grain, Tristram, and Stompor, 2009; Alonso et al., 2019). Nevertheless, such approaches do not result in pure $\mathcal{E}$ and $\mathcal{B}$ maps in the real space and are computationally intensive. Ferté et al.

(2013) provides a quantitative comparison of the efficiency of the above techniques for power spectrum reconstruction. Wavelet-based techniques have also been proposed, but they must be carefully adapted for the problem under investigation (Cao and Fang, 2009; Rogers et al., 2016; Leistedt et al., 2017).

Bunn and Wandelt (2017) (hereafter BW) have recently shown that the $\mathcal{E}/\mathcal{B}$ decomposition can be approached from a Wiener filtering (Wiener, 1949) viewpoint, resulting in faster implementation as compared to the above methods, while providing real-space maps of the $\mathcal{E}$ and $\mathcal{B}$ modes. Another key advantage of such an approach is that it can be naturally extended to treat more interesting cases such as providing $\mathcal{E}$ maps free from the temperature anisotropy contributions by accounting for temperature and polarization correlations.

In this chapter, we present an augmented version of our dual messenger algorithm (Kodi Ramanah, Lavaux, and Wandelt, 2018) (hereafter KLW18) for pure $\mathcal{E}/\mathcal{B}$ decomposition on the sphere, based on the principle of the Wiener filter. We adapt the algorithm to encode the BW prescription for reconstruction of pure $\mathcal{E}/\mathcal{B}$ maps and naturally extend the dual messenger framework to account for complex and realistic noise models. We demonstrate the application of this enhanced algorithm, designated as DANTE (DuAl messeNger filTEr), on a simulated CMB polarization data set which emulates the features of the actual Planck data.

The remainder of this chapter is structured as follows. In Section 5.2, we provide a brief description of the dual messenger formalism adapted for applications on the sphere and illustrate a Jacobi relaxation method to account for the non-orthogonality of spherical harmonic transforms. We describe how it can be augmented to deal with non-trivial noise covariance in Section 5.3. We subsequently illustrate how the algorithm can encode the BW prescription for pure $\mathcal{E}/\mathcal{B}$ reconstruction, followed by an outline of the numerical implementation in Section 5.4. We then present a new optimization scheme for the estimation of noise covariance from Monte Carlo simulations and showcase the capabilities of DANTE in reconstructing pure $\mathcal{E}$ and $\mathcal{B}$ maps from simulated Planck data in Section 5.5. Finally, we summarize our main findings in Section 5.6. In Appendix B.2, we provide a brief description of spherical harmonic transforms, as employed in this work. Appendix B.4 outlines the main steps in the derivation of the essential dual messenger equations to account for anisotropic correlated noise.

5.2 Dual messenger field on the sphere

Conceptually, the essence of the messenger methods lies in the introduction of an auxiliary field to mediate between the different bases where the signal and noise covariances, $\mathbf{S}$ and $\mathbf{N}$, can be most conveniently expressed as sparse matrices. The addition of this messenger field allows the Wiener filter equation to be rewritten as a set of algebraic equations that must be solved iteratively, circumventing the requirement of matrix inversions or preconditioners.

With respect to the standard messenger technique, where a messenger field $\boldsymbol{t}$ is introduced at the level of the noise, the dual messenger algorithm incorporates an extra messenger field $\boldsymbol{u}$, at the

level of the signal, with corresponding covariances $\mathbf{T}$ and $\mathbf{U}$, yielding the following augmented χ^2:

$$\chi^2_{\boldsymbol{T},\boldsymbol{U}} = (\boldsymbol{d}-\boldsymbol{t})^\dagger \bar{\mathbf{N}}^{-1}(\boldsymbol{d}-\boldsymbol{t}) + (\boldsymbol{t}-\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u})^\dagger \mathbf{T}^{-1}(\boldsymbol{t}-\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u}) + (\boldsymbol{u}-\boldsymbol{s})^\dagger \mathbf{U}^{-1}(\boldsymbol{u}-\boldsymbol{s}) + \boldsymbol{s}^\dagger \bar{\mathbf{S}}^{-1}\boldsymbol{s}, \tag{5.1}$$

where $\bar{\mathbf{N}} = \mathbf{N} - \mathbf{T}$ with $\mathbf{T} = \alpha\mathbb{1}$, where $\alpha \equiv \min(\mathrm{diag}(\mathbf{N}))$, and $\bar{\mathbf{S}} = \mathbf{S} - \mathbf{U}$ with $\mathbf{U} = \nu\mathbb{1}$, where $\nu \equiv \min(\mathrm{diag}(\mathbf{S}))$. When dealing with polarization fields, μ and ν are actually 3×3 matrices, corresponding to the temperature, $\mathcal{E}$ and $\mathcal{B}$ components. $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Y}}^\dagger$ correspond to the basis operators (synthesis and analysis operators, respectively) for the spherical harmonic transforms, as described in Appendix B.2, while $\mathbf{B}$ indicates convolution with an instrument beam. In terms of physical significance, $\boldsymbol{t}$ corresponds to a homogeneous portion of the noise covariance while its counterpart $\boldsymbol{u}$ is associated with the signal covariance. Optimizing $\chi^2_{\boldsymbol{T},\boldsymbol{U}}$ yields the following system of equations to be solved iteratively:

$$\boldsymbol{u} = (\bar{\mathbf{S}}+\mathbf{U})\left[\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbf{B}(\bar{\mathbf{S}}+\mathbf{U})+\mathbf{T}\right]^{-1}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{t} \tag{5.2}$$

$$\boldsymbol{t} = (\bar{\mathbf{N}}^{-1}+\mathbf{T}^{-1})^{-1}(\mathbf{T}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u}+\bar{\mathbf{N}}^{-1}\boldsymbol{d}). \tag{5.3}$$

Note that this is the reduced system of equations, with one of the messenger fields made implicit. The general system of equations is described in more depth in KLW18. To improve convergence, we implement a similar scheme as in KLW18. We artificially truncate the signal covariance $\mathbf{S}$ to some lower initial value of ℓ_{iter}, corresponding to a covariance μ. We subsequently vary $\mathbf{U}$ to bring $\mu \to \nu$, such that in the limit $\mu = \nu$, we have $\boldsymbol{u} = \boldsymbol{s}$ and we recover the usual Wiener filter Eq. (2.3) from the above system of Eqs. (5.2) and (5.3). This is formally valid as long as μ and ν are block matrices over harmonic space. We may therefore exploit this degree of freedom to solve the temperature and polarization signals at different rates. The above cooling scheme leads to a redefinition of $\bar{\mathbf{S}}$ using the Heaviside function as $\bar{\mathcal{S}} = \Theta(\mathcal{S}-\mathbf{U})$, where $\mathcal{S}$ corresponds to the eigenvalues of $\mathbf{S}$, i.e. $\mathbf{S} = \mathfrak{R}^\dagger\mathcal{S}\mathfrak{R}$ and $\bar{\mathbf{S}} = \mathfrak{R}^\dagger\bar{\mathcal{S}}\mathfrak{R}$.

To implement such a hierarchical scheme, we vary $\mathbf{U}$ via a cooling scheme for $\boldsymbol{\xi}$, where $\boldsymbol{\xi} = \mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbf{B}\mathbf{U} + \mathbf{T}$. To obtain the appropriate Wiener filter solution, we need to reduce $\mu \to \nu = 0$, due to the continuous mode of the signal, i.e. the zero eigenvalue of $\mathbf{S}$. The cooling scheme for $\boldsymbol{\xi}$ entails gradually reducing $\boldsymbol{\xi}$ by a constant factor and iterating until $\boldsymbol{\xi} \to \mathbf{T}$, thereby finally attaining $\mu = 0$, as required. A quantitative description of the rationale underlying the above cooling scheme is presented in our previous work (Kodi Ramanah, Lavaux, and Wandelt, 2017).

DANTE is implemented in Python and it makes use of the HEALPIX[1] (Górski et al., 2005) library, in particular the Python wrapper HEALPY, to perform the spherical harmonic transforms (SHTs). HEALPIX employs an equal area projection scheme, where the SHTs are quasi-orthogonal, i.e. $\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbb{1} \approx (N_{\mathrm{pix}}/4\pi)\mathbb{1} \equiv \beta\mathbb{1}$, where N_{pix} denotes the number of pixels. We account for this non-orthogonality of SHTs via efficient Jacobi relaxation schemes, as described in future sections. If an equidistant cylindrical projection on a grid is adopted for the SHTs (e.g. Muciaccia, Natoli, and Vittorio, 1997; Huffenberger and Wandelt, 2010; McEwen and Wiaux, 2011), the equations presented in this work are significantly simplified, as a result of $\beta = 1$. DANTE employs the

[1] http://healpix.jpl.nasa.gov

NUMBA[2] (Lam, Pitrou, and Seibert, 2015) compiler for Python arrays and numerical functions to yield high-performance functions for all the required matricial manipulations to boost execution speed. NUMBA generates optimized native code using the LLVM compiler (Lattner and Adve, 2004) infrastructure and is used to parallelize the array operations.

5.2.1 Non-orthogonality of spherical harmonic transforms

Unlike in the case of discrete Fourier transforms, the spherical harmonic synthesis and analysis operators, i.e. $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{Y}}^\dagger$, respectively, are not orthogonal and differ by more than a transposition and a scale factor. The quality of the approximation, $\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbb{1} \approx (N_{\rm pix}/4\pi)\mathbb{1} \equiv \beta\mathbb{1}$, depends on the $\ell_{\rm max}$, $N_{\rm pix}$ and spherical grid considered.

To account for the non-orthogonality of spherical harmonic transforms, we incorporate an internal Jacobi relaxation method (Jacobi, 1845; Saad, 2003) in DANTE to refine the solution in harmonic space (cf. equation (5.2)). To obtain $\mathbf{U}$, we need the eigenvalues of $\mathbf{S}$, i.e. $\mathbf{U} = \min(\mathcal{S})\mathbb{1}$, where $\mathfrak{R}^\dagger\mathcal{S}\mathfrak{R} = \mathbf{S}$. Equation (5.2) can be formulated as $\boldsymbol{s} = \boldsymbol{\mathcal{A}}\boldsymbol{b}$, where $\boldsymbol{\mathcal{A}}$ is given by:

$$\boldsymbol{\mathcal{A}} = \mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U})\left[\mathfrak{R}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbf{B}\mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) + \alpha\mathbb{1}\right]^{-1}\mathfrak{R}, \tag{5.4}$$

after including the basis transformations, and $\boldsymbol{b} = \mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{t}$.

An approximation to $\boldsymbol{\mathcal{A}}$ can be obtained as follows:

$$\widetilde{\boldsymbol{\mathcal{A}}} = \mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U})\left[\beta\mathfrak{R}\mathbf{B}^\dagger\mathbf{B}\mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) + \alpha\mathbb{1}\right]^{-1}\mathfrak{R}, \tag{5.5}$$

after using the approximate orthogonality relation $\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbb{1} \approx \beta\mathbb{1}$. The application of the operator $\boldsymbol{\mathcal{A}}$ is not well-defined but we nevertheless can apply its inverse $\boldsymbol{\mathcal{A}}^{-1}$ to a vector, by applying the relevant operators sequentially,

$$\boldsymbol{\mathcal{A}}^{-1} = \mathfrak{R}^\dagger\left[\mathfrak{R}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbf{B}\mathfrak{R}^\dagger + \alpha(\bar{\mathcal{S}} + \mathbf{U})^{-1}\right]\mathfrak{R}. \tag{5.6}$$

We therefore make use of $\boldsymbol{\mathcal{A}}^{-1}$ and $\widetilde{\boldsymbol{\mathcal{A}}}$, to obtain the solution for $\boldsymbol{s}$ via the following Jacobi iterations:

$$\boldsymbol{s}_{n+1} = \boldsymbol{s}_n + \widetilde{\boldsymbol{\mathcal{A}}}(\boldsymbol{b} - \boldsymbol{\mathcal{A}}^{-1}\boldsymbol{s}_n), \tag{5.7}$$

where n denotes the number of Jacobi iterations.

The term $(\bar{\mathcal{S}} + \mathbf{U})^{-1}$ poses a numerical predicament for the final truncation in the signal covariance, where $\bar{\mathcal{S}} = \mathcal{S} - \mathbf{U}$ and $\mathbf{U} = \mathbf{0}$, and we subsequently require the inversion of $\mathcal{S}$. We circumvent the corner case due to the zero eigenvalues of the continuous modes in $\mathcal{S}$ by imposing the following constraint on the subspace $\mathcal{V}$ where $\mathcal{S} = \mathbf{0}$: $(\bar{\mathcal{S}} + \mathbf{U})^{-1}|_{\mathcal{V}} = \mathcal{S}^+$. $\mathcal{S}^+$ is the pseudo-inverse, i.e. $\mathcal{S}\mathcal{S}^+ = \Pi$, where Π is a projector.

[2] https://numba.pydata.org

5.2.2 Treatment of masked regions

CMB data analysis inevitably requires the treatment of masks, with many practical applications requiring that parts of the sky be masked out. For full-sky observations, this is mainly to avoid contamination from the galactic foreground emissions, thereby preventing spurious power spectra measurements. In the case of ground-based or sub-orbital CMB experiments with partial sky coverage, missing data are accounted for using masks.

We provide an outline of the general procedure for solving the messenger Eq. (5.3) when dealing with temperature and polarization masks. Here, we assume correlated noise, such that the noise covariance $\mathbf{N}$ has the following block-diagonal form for every pixel i:

$$\mathbf{N}_i = \begin{pmatrix} \langle \mathcal{I}\mathcal{I} \rangle & \langle \mathcal{I}\mathcal{Q} \rangle & \langle \mathcal{I}\mathcal{U} \rangle \\ \langle \mathcal{Q}\mathcal{I} \rangle & \langle \mathcal{Q}\mathcal{Q} \rangle & \langle \mathcal{Q}\mathcal{U} \rangle \\ \langle \mathcal{U}\mathcal{I} \rangle & \langle \mathcal{U}\mathcal{Q} \rangle & \langle \mathcal{U}\mathcal{U} \rangle \end{pmatrix}, \tag{5.8}$$

where $\mathcal{I}$, $\mathcal{Q}$ and $\mathcal{U}$ are the Stokes parameters. More complex noise models will be described in Section 5.3.

We compute the covariance $\mathbf{T}$ of the messenger field $\boldsymbol{t}$ as follows: $\mathbf{T} = \min(\mathcal{N})\mathbb{1} = \alpha\mathbb{1}$, where $\mathfrak{D}^\dagger \mathcal{N} \mathfrak{D} = \mathbf{N}$. The noise covariance $\mathbf{N}$ can be written as $\mathbf{N} = \mathbf{\Sigma C \Sigma} = \mathbf{\Sigma P}^\dagger \mathbf{\Delta P \Sigma}$, following the orthonormal decomposition of $\mathbf{C}$, where $\mathbf{\Sigma}$ is a diagonal matrix with the eigenvalues $\{\sigma^i_{\mathcal{I}}, \sigma^i_{\mathcal{Q}}, \sigma^i_{\mathcal{U}}\}$ corresponding to the noise amplitudes for the i^{th} pixel, with the orthonormal decomposition of $\mathbf{C}$ resulting in the diagonal matrix $\mathbf{\Delta}$. We then obtain $\bar{\mathbf{N}} = \mathbf{N} - \mathbf{T}$ as follows:

$$\begin{aligned} \bar{\mathbf{N}} &= \mathbf{\Sigma P}^\dagger \mathbf{\Delta P \Sigma} - \mathbf{T} \\ &= \mathbf{\Sigma P}^\dagger (\mathbf{\Delta} - \alpha \mathbf{P \Sigma}^{-2} \mathbf{P}^\dagger) \mathbf{P \Sigma}, \end{aligned} \tag{5.9}$$

where α, as stated above, is the smallest eigenvalue of $\mathbf{N}$. To solve the messenger Eq. (5.3), we require the inverse $\bar{\mathbf{N}}^{-1}$,

$$\bar{\mathbf{N}}^{-1} = \mathbf{\Sigma}^{-1} \mathbf{P}^\dagger (\mathbf{\Delta} - \alpha \mathbf{P \Sigma}^{-2} \mathbf{P}^\dagger)^{-1} \mathbf{P \Sigma}^{-1}, \tag{5.10}$$

such that $\bar{\mathbf{N}}^{-1}$ has a block-diagonal structure in pixel space. We obtain the solution for the messenger field by simply evaluating Eq. (5.3) in pixel space,

$$\boldsymbol{t}_x = \left(\bar{\mathbf{N}}^{-1} + \mathbf{T}^{-1} \right)^{-1}_x \left(\mathbf{T}_x^{-1} \boldsymbol{\mathcal{Y}} \mathbf{B} \boldsymbol{u}_\ell + \bar{\mathbf{N}}_x^{-1} \boldsymbol{d}_x \right). \tag{5.11}$$

We implement the temperature and polarization masks by increasing the noise variance to infinity for masked pixels, or numerically by setting the inverse noise covariance to zero. This is achieved by setting $\mathbf{\Sigma}^{-1}|_{\text{mask}} = \mathbf{0}$, subsequently ensuring that data from masked regions do not contaminate the messenger field.

5.3 Dual messenger generalizations

The dual messenger approach can be extended to a broader class of problems, accounting for highly non-trivial noise covariance, as described in the following sections. In practice, the structure of the noise covariance, in pixel space, is influenced by the noise properties of the CMB time-ordered data and the scanning pattern of the telescope.

5.3.1 Correlated modulated noise

The first possible generalization for the noise model is the case of correlated modulated noise. The noise covariance, in pixel space, can be written as:

$$\mathbf{N} = \boldsymbol{\mathcal{Y}}\mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger}, \tag{5.12}$$

where $\mathbf{F}$ is a smoothing kernel which is diagonal in the same basis as $\mathbf{S}$, i.e. harmonic space, while $\mathbf{D}$ is the noise variance that can be easily diagonalized in some other basis, e.g. pixel space. The desired Wiener filter from Eq. (2.3) is then

$$s_{\mathrm{WF}} = \mathbf{S}(\mathbf{S} + \mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{F})^{-1}\boldsymbol{\mathcal{Y}}^{-1}d. \tag{5.13}$$

It turns out that this can be solved directly by the dual messenger scheme described above without any modifications. We transform the data via a simple pre-whitening step, as follows:

$$\tilde{d} = (\boldsymbol{\mathcal{Y}}\mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}d. \tag{5.14}$$

The Wiener filter for this model can then be computed via the following steps:

$$\begin{aligned}\tilde{s}_{\mathrm{WF}} &= (\mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}\mathbf{S}(\boldsymbol{\mathcal{Y}}\mathbf{F})^{-1}\left[(\mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}\mathbf{S}(\boldsymbol{\mathcal{Y}}\mathbf{F})^{-1} + \mathbf{D}\right]^{-1}\tilde{d} \\ &= (\mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}\mathbf{S}(\mathbf{S} + \mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{F})^{-1}\boldsymbol{\mathcal{Y}}^{-1}d \\ &= (\boldsymbol{\mathcal{Y}}\mathbf{F}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}s_{\mathrm{WF}},\end{aligned} \tag{5.15}$$

after plugging in the effective data vector given by Eq. (5.14).

This problem therefore reduces to one that can be solved directly using the dual messenger algorithm, requiring the same computational time as in the white noise case. The only additional steps required are a simple pre-whitening, followed by a post-smoothing operation with $\mathbf{F}^{-1}$ and $\mathbf{F}$, respectively. For this particular case, we do not demonstrate the application of DANTE, as the implementation is straightforward.

5.3.2 Modulated correlated noise

The second generalization of the noise model corresponds to modulating the amplitude of spatially correlated noise. This is a more realistic noise model, typical of CMB experiments such as Planck, resulting from the scanning strategy of the instrument. The noise covariance, in pixel space, now

takes the following form:

$$\mathbf{N} = \mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger\mathbf{D}, \tag{5.16}$$

where $\mathbf{C}$ is the isotropic noise covariance, encoding the inverse frequency ($1/f$) noise correlation on the large scales, typically associated with atmospheric noise, and therefore diagonal in harmonic space. The modulation, described by $\mathbf{D}$, is sparse in pixel space. The power spectrum C_ℓ of the non-modulated part of the noise can be expressed as:

$$C_\ell = \frac{\sigma_N^2}{\beta}\left[1 + \left(\frac{\ell_{\rm knee}}{\ell}\right)^{\alpha_{\rm knee}}\right], \tag{5.17}$$

with the characteristic scale $\ell_{\rm knee}$ and tilt $\alpha_{\rm knee}$ of the power-law component, with σ_N being the noise amplitude per pixel.

The modulation takes into account the variation of noise amplitude due to the amount of integration time spent in any single pixel. The isotropic noise covariance in the map is expected to be a good approximation for scan strategies that cross each pixel in many directions, thereby isotropizing the way the time-ordered data noise correlations project onto the sky. But even in cases where the distribution of scan directions per pixel is not entirely isotropic, such as for the Planck data, this noise model was found to be of sufficient quality to derive the optimal primordial non-Gaussianity estimators (Planck Collaboration et al., 2016c; Planck Collaboration et al., 2019).

As a result of the above dense noise covariance, Eq. (5.3) is no longer algebraically solvable due to the required inversion of a fully dense system. But since we are free to choose the covariance $\mathbf{T}$ of the messenger field $\boldsymbol{t}$, we set $\mathbf{T} = \mathbf{D}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)\mathbf{D}$, where $\phi = \min(\mathrm{diag}(\mathbf{C}))$. This yields the following system of equations:

$$\boldsymbol{u} = (\bar{\mathbf{S}} + \mathbf{U})\left[\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\mathbf{D}^{-1}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\mathbf{D}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}(\bar{\mathbf{S}} + \mathbf{U}) + \mathbb{1}\right]^{-1}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\mathbf{D}^{-1}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\mathbf{D}^{-1}\boldsymbol{t} \tag{5.18}$$

$$\boldsymbol{t} = \mathbf{D}\left[(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger - \boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1} + (\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\right]^{-1}\left[(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\mathbf{D}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u} + (\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger - \boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\mathbf{D}^{-1}\boldsymbol{d}\right], \tag{5.19}$$

where the second equation can be simplified to the following form via straightforward linear algebraic manipulations:

$$\tilde{\boldsymbol{t}} \equiv \mathbf{D}^{-1}\boldsymbol{t} = \boldsymbol{\mathcal{Y}}(\mathbf{C} - \phi\mathbb{1})\boldsymbol{\mathcal{Y}}^\dagger(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\mathbf{D}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u} + (\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\mathbf{D}^{-1}\boldsymbol{d}, \tag{5.20}$$

which can now be solved trivially to obtain the messenger field. The first Eq. (5.18), however, cannot be solved directly, but it can be conveniently expanded using an extra messenger field $\boldsymbol{v}$, with covariance $\mathbf{V} = \omega(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)\mathbb{1}$, where $\omega \equiv \min(\mathrm{diag}(\mathbf{D}^2))$, yielding the following two trivially solvable equations:

$$\boldsymbol{v} = \omega\boldsymbol{\mathcal{M}}\mathbf{D}^{-1}\boldsymbol{\mathcal{M}}^{-1}\tilde{\boldsymbol{t}} + \left[\mathbb{1} - \omega\boldsymbol{\mathcal{M}}\mathbf{D}^{-1}\boldsymbol{\mathcal{M}}^{-1}\mathbf{D}^{-1}\right]\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u} \tag{5.21}$$

$$\boldsymbol{u} = \left[\phi\omega(\bar{\mathbf{S}} + \mathbf{U})^{-1} + \mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}\right]^{-1}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{v}, \tag{5.22}$$

where the coupling matrix $\boldsymbol{\mathcal{M}}$ is defined as $\boldsymbol{\mathcal{M}} \equiv \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^\dagger$. We therefore must solve the above system of three Eqs. (5.20)-(5.22) when accounting for modulated correlated noise. Eq. (5.22) can be written explicitly in terms of the relevant basis transformations as follows:

$$\boldsymbol{u} = \mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) \left[\mathfrak{R}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}\mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) + \phi\omega\mathbb{1}\right]^{-1} \mathfrak{R}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}v, \tag{5.23}$$

where, as before, $\mathbf{S} = \mathfrak{R}^\dagger\mathcal{S}\mathfrak{R}$. An in-depth derivation of these equations is laid out in Appendix B.4. We encode the mask by doing the decomposition, $\mathbf{D} = \boldsymbol{\Sigma}\widetilde{\mathbf{D}}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{P}^\dagger\boldsymbol{\Delta}\mathbf{P}\boldsymbol{\Sigma}$, as described in Section 5.2.2, and setting $\boldsymbol{\Sigma}^{-1}|_{\text{mask}} = \mathbf{0}$. This prescription corresponds exactly to dropping the contribution of observations that are considered masked out, which may be deduced from Eq. (5.1). We apply the cooling scheme to $\boldsymbol{\xi} = \mathbf{U} + \phi\omega\mathbb{1}$, as described in Section 5.5.

5.3.3 Nested Jacobi relaxation

As mentioned in Section 5.2, the above Eqs. (5.20), (5.21) and (5.23) may be simplified significantly using the approximation $\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{Y}}\mathbb{1} \approx \beta\mathbb{1}$, thereby reducing the required number of SHTs. This approximation is not exact due to the coupling of the SHTs on the pixelized sky. In this work, we employ Jacobi relaxation to correct for the operations $\boldsymbol{\mathcal{M}}^{-1} \equiv (\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^\dagger)^{-1}$ and $(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger)^{-1}$.

We follow a similar rationale and employ the same notation as in Section 5.2.1, with the Jacobi iteration given by Eq. (5.7), where $\boldsymbol{b}$ is any arbitrary vector in harmonic space. For the case of $\boldsymbol{\mathcal{A}} = (\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger)^{-1}$, the respective operations are as follows:

$$\widetilde{\boldsymbol{\mathcal{A}}} = (\beta^{-1}\boldsymbol{\mathcal{Y}})\mathbf{C}^{-1}(\beta^{-1}\boldsymbol{\mathcal{Y}}^\dagger), \; \boldsymbol{\mathcal{A}}^{-1} = \boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger. \tag{5.24}$$

The correction for $\boldsymbol{\mathcal{M}}^{-1}$ is analogous to the above, with $\mathbf{C}$ set to identity matrix, and must be embedded within the less trivial Jacobi relaxation for Eq. (5.23). The resulting nested relaxation scheme requires the following operations:

$$\boldsymbol{\mathcal{A}} = \mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) \left[\mathfrak{R}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}\mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) + \phi\omega\mathbb{1}\right]^{-1} \mathfrak{R} \tag{5.25}$$

$$\widetilde{\boldsymbol{\mathcal{A}}} = \mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) \left[\mathfrak{R}\mathbf{B}^\dagger\mathbf{B}\mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) + \phi\omega\mathbb{1}\right]^{-1} \mathfrak{R}, \tag{5.26}$$

and the corresponding inverse of operator $\boldsymbol{\mathcal{A}}$ given by:

$$\boldsymbol{\mathcal{A}}^{-1} = \mathfrak{R}^\dagger \left[\mathfrak{R}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}\mathfrak{R}^\dagger(\bar{\mathcal{S}} + \mathbf{U}) + \phi\omega\mathbb{1}\right] (\bar{\mathcal{S}} + \mathbf{U})^{-1}\mathfrak{R}, \tag{5.27}$$

with the basis operations included, and $\boldsymbol{b} = \mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}v$.

5.4 Pure $\mathcal{E}$ / $\mathcal{B}$ decomposition via Wiener filtering

In a finite patch of sky, the polarization field cannot be uniquely decomposed into pure $\mathcal{E}$ and pure $\mathcal{B}$ modes. Nevertheless, the polarization map can be uniquely decomposed into three distinct components, commonly referred to as the "pure $\mathcal{E}$", "pure $\mathcal{B}$" and "ambiguous" modes (Lewis, Challinor, and Turok, 2002; Bunn et al., 2003). This new set of ambiguous modes receives contributions from

both $\mathcal{E}$ and $\mathcal{B}$ modes. In such a framework, the signal vector space is divided into three orthogonal subspaces. The pure $\mathcal{B}$ modes exist on the vector subspace orthogonal to that of all $\mathcal{E}$ modes, and similarly for the pure $\mathcal{E}$ modes. The ambiguous component, however, lies in the subspace orthogonal to both pure $\mathcal{E}$ and $\mathcal{B}$ subspaces. This decomposition ensures that a reconstructed pure $\mathcal{B}$ map is not contaminated by $\mathcal{E}$ modes.

$\mathcal{E}$ / $\mathcal{B}$ separation methods based on this pure-ambiguous decomposition originally involved the construction of an eigenbasis for the various orthonormal subspaces, but this is a tedious and numerically slow procedure. The $\mathcal{E}$ / $\mathcal{B}$ decomposition is trivial for exact methods such as Gibbs sampling (Larson et al., 2007), which infer the posterior statistics of a full-sky signal conditional on the data. Gibbs sampling requires a complete sky sample, i.e. optimally filtered data augmented to cover the whole sky via constrained generalizations (see also KLW18). This is the basis of the motivation behind the Wiener filtering approach proposed by Bunn and Wandelt (2017).

We briefly review the rationale and the formalism behind this new method, and describe how it can be incorporated in DANTE. A more comprehensive description is provided in Bunn and Wandelt (2017).

5.4.1 Background and notation

Considering only polarization measurements, the data set can be described as a $2N_{\rm pix}$ dimensional vector of the Stokes parameters $\mathcal{Q}$ and $\mathcal{U}$, i.e. $\boldsymbol{d} = (\boldsymbol{d}_\mathcal{Q}, \boldsymbol{d}_\mathcal{U})$, where $N_{\rm pix}$ corresponds to the dimension of the pixelized map of a given Stokes parameter. We account for the contribution from the temperature anisotropy, Stokes $\mathcal{I}$, in a future section. We assume a data model as given by Eq. (2.1), and Gaussian white noise, although the results presented below would still be relevant for more complex noise covariance.

The signal can be expressed as a spherical harmonic expansion,

$$\boldsymbol{s} = \boldsymbol{s}^\mathcal{E} + \boldsymbol{s}^\mathcal{B} = \sum_{\ell,m} \left[a^\mathcal{E}_{\ell m} Y^\mathcal{E}_{z,\ell m}(\hat{r}_j) + a^\mathcal{B}_{\ell m} Y^\mathcal{B}_{z,\ell m}(\hat{r}_j) \right], \tag{5.28}$$

where $\hat{r}_j$ labels the pixel corresponding to measurement j, while the index $z \in \{\mathcal{Q},\mathcal{U}\}$ denotes the associated Stokes parameter. The functions Y can be expressed in terms of spin-2 spherical harmonics:

$$Y^\mathcal{E}_{\mathcal{Q},\ell m} = Y^\mathcal{B}_{\mathcal{U},\ell m} = -\frac{1}{2}({}_2Y_{\ell m} + {}_{-2}Y_{\ell m}) \tag{5.29}$$

$$Y^\mathcal{B}_{\mathcal{Q},\ell m} = -Y^\mathcal{E}_{\mathcal{U},\ell m} = -\frac{1}{2}({}_2Y_{\ell m} - {}_{-2}Y_{\ell m}). \tag{5.30}$$

The signal can therefore be written as

$$\boldsymbol{s} = \mathbf{Y}_\mathcal{E}\boldsymbol{e} + \mathbf{Y}_\mathcal{B}\boldsymbol{b}, \tag{5.31}$$

with the coefficients $a^\mathcal{E}_{\ell m}$ and $a^\mathcal{B}_{\ell m}$ encoded in the vectors $\boldsymbol{e}$ and $\boldsymbol{b}$, respectively. The matrices $\mathbf{Y}_\mathcal{Z}$, for $\mathcal{Z} \in \{\mathcal{E},\mathcal{B}\}$, consist of the spherical harmonics evaluated at the given pixel locations.

Under the assumption of data sourced by a statistically isotropic and random Gaussian process, the signal is uniquely described by the following covariance:

$$\mathbf{S} \equiv \langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle = \langle \boldsymbol{s}^{\mathcal{E}}\boldsymbol{s}^{\mathcal{E}\dagger} \rangle + \langle \boldsymbol{s}^{\mathcal{B}}\boldsymbol{s}^{\mathcal{B}\dagger} \rangle \equiv \mathbf{S}_{\mathcal{E}} + \mathbf{S}_{\mathcal{B}}. \tag{5.32}$$

For a full-sky data set, where $\boldsymbol{d}$ covers the whole sky, the matrices $\mathbf{Y}_{\mathcal{E}}$ and $\mathbf{Y}_{\mathcal{B}}$ span orthogonal spaces, and hence the $\mathcal{E}$-$\mathcal{B}$ coupling issue does not arise and the decomposition is straightforward. Incomplete sky coverage, however, results in the ambiguous modes that lie in both subspaces at the same time. The pure $\mathcal{B}$ space can therefore be described as the orthogonal complement of the space spanned by $\mathbf{Y}_{\mathcal{E}}$, with an analogous definition for the pure $\mathcal{E}$ space. The signal vector space can consequently be divided into three orthogonal subspaces, with the third ambiguous space being orthogonal to both of the pure subspaces. By projecting the data vector $\boldsymbol{d}$ on the pure $\mathcal{B}$ subspace, the $\mathcal{E}$-mode signal is mapped onto the null space of $\mathbf{Y}_{\mathcal{B}}$, ensuring no contamination of $\mathcal{E}$ modes in the resulting pure $\mathcal{B}$ map.

5.4.2 Purification framework

If the signal covariance $\mathbf{S}$ employed in the Wiener filter Eq. (2.3) contains covariances of both $\mathcal{E}$ and $\mathcal{B}$ signals, then the resulting Wiener-filtered map would have contributions from both the scalar and pseudo-scalar components, i.e. $\boldsymbol{s}_{\mathrm{WF}} = \boldsymbol{s}^{\mathcal{E}}_{\mathrm{WF}} + \boldsymbol{s}^{\mathcal{B}}_{\mathrm{WF}}$. Bunn and Wandelt (2017) proposed the following approach to isolate them from each other: Conceptually, the rationale is to treat one component as a source of noise. We can obtain the Wiener-filtered $\mathcal{B}$ map, for instance, via the following replacements: $\mathbf{S} \to \mathbf{S}_{\mathcal{B}}$ and $\mathbf{N} \to \mathbf{S}_{\mathcal{E}} + \mathbf{N}$, such that the Wiener filter Eq. (2.3) becomes:

$$\begin{aligned} \boldsymbol{s}^{\mathcal{B}}_{\mathrm{WF}} &= \left[\mathbf{S}_{\mathcal{B}}^{-1} + (\mathbf{S}_{\mathcal{E}} + \mathbf{N})^{-1}\right]^{-1} (\mathbf{S}_{\mathcal{E}} + \mathbf{N})^{-1}\boldsymbol{d} \\ &= \mathbf{S}_{\mathcal{B}} \left[\mathbf{S}_{\mathcal{B}} + (\mathbf{S}_{\mathcal{E}} + \mathbf{N})\right]^{-1} \boldsymbol{d}, \end{aligned} \tag{5.33}$$

with an analogous expression for $\boldsymbol{s}^{\mathcal{E}}_{\mathrm{WF}}$. Recall that $\boldsymbol{s}_{\mathrm{WF}} = \boldsymbol{s}^{\mathcal{E}}_{\mathrm{WF}} + \boldsymbol{s}^{\mathcal{B}}_{\mathrm{WF}}$.

However, the above Wiener-filtered maps are contaminated by the ambiguous modes. Due to our prior signal covariance having a much higher $\mathcal{E}$-mode power, the theory assigns the ambiguous modes with high signal-to-noise mostly to the $\mathcal{E}$ map. In order to ensure no cross-contamination, whereby a pure $\mathcal{B}$ map should have contributions strictly from $\mathcal{B}$ modes, Bunn and Wandelt (2017) suggested raising the signal covariance associated to the $\mathcal{E}$ component to infinity, and provided a proof that this gives the same result as doing a costly eigenmode decomposition and projecting out the ambiguous modes. We define

$$\mathbf{S}(\lambda) = \mathbf{S}_{\mathcal{B}} + \lambda \mathbf{S}_{\mathcal{E}} \tag{5.34}$$

as the signal covariance with the $\mathcal{E}$-mode power amplified by a factor of λ. Substituting $\mathbf{S}_{\mathcal{B}} + \mathbf{S}_{\mathcal{E}} \to \mathbf{S}(\lambda)$ in Eq. (5.33) yields

$$\begin{aligned} \boldsymbol{s}^{\mathcal{B}}_{\mathrm{WF}}(\lambda) &= \mathbf{S}_{\mathcal{B}} \left[\mathbf{S}(\lambda) + \mathbf{N}\right]^{-1} \boldsymbol{d} \\ &= \mathbf{S}_{\mathcal{B}}\mathbf{S}(\lambda)^{-1} \left[\mathbf{S}(\lambda)^{-1} + \mathbf{N}^{-1}\right]^{-1} \mathbf{N}^{-1}\boldsymbol{d}, \end{aligned} \tag{5.35}$$

such that in the limit $\lambda \to \infty$, only the pure $\mathcal{B}$ modes survive in the null space of $\mathbf{S}_\mathcal{E}$. A strictly analogous procedure holds for the pure $\mathcal{E}$ component.

5.4.3 Numerical implementation

To facilitate the numerical evaluation of the expressions above, it is more convenient to work with a full-sky data set, so that we can use fast transforms to move back and forth between the pixel and spherical harmonic spaces. Masked pixels are assigned infinite noise covariance, i.e. we set the inverse noise covariance to zero.

Assuming isotropic and Gaussian CMB anisotropies, the signal covariance $\mathbf{S}$ is diagonal in spherical harmonic basis,

$$\mathbf{S}_\mathcal{E} = \mathrm{diag}(C_2^{EE}, \ldots, C_{\ell_\mathrm{max}}^{EE}, 0, \ldots, 0) \tag{5.36}$$

$$\mathbf{S}_\mathcal{B} = \mathrm{diag}(0, \ldots, 0, C_2^{BB}, \ldots, C_{\ell_\mathrm{max}}^{BB}), \tag{5.37}$$

with the ordering convention of having the $\mathcal{E}$-mode component first. Hence, we have

$$\mathbf{S}(\lambda)^{-1} = \mathrm{diag}\Big[(\lambda C_2^{EE})^{-1}, \ldots, (\lambda C_{\ell_\mathrm{max}}^{EE})^{-1}, (C_2^{BB})^{-1}, \ldots, (C_{\ell_\mathrm{max}}^{BB})^{-1}\Big]. \tag{5.38}$$

As a result, $\mathbf{S}(\lambda)^{-1}|_{\lambda\to\infty} \to \mathbf{S}_\mathcal{B}^+$, where the pseudo-inverse $\mathbf{S}_\mathcal{B}^+$ is the inverse of $\mathbf{S}_\mathcal{B}$ within the subspace spanned by $\mathbf{Y}_\mathcal{B}$ and is zero in the orthogonal subspace of $\mathcal{E}$ modes. We consequently obtain the operator that projects onto the pure $\mathcal{E}$ subspace as $\mathbf{S}_\mathcal{B}\mathbf{S}_\mathcal{B}^+ = \mathbf{P}_\mathcal{B}$. Finally, we obtain the Wiener-filtered pure $\mathcal{B}$ map as

$$\boldsymbol{s}_\mathrm{WF}^{p\mathcal{B}} \equiv \lim_{\lambda\to\infty} \boldsymbol{s}_\mathrm{WF}^\mathcal{B}(\lambda) = \mathbf{S}_\mathcal{B}\mathbf{S}_\mathcal{B}^+(\mathbf{S}_\mathcal{B}^+ + \mathbf{N}^{-1})^{-1}\mathbf{N}^{-1}\boldsymbol{d}. \tag{5.39}$$

The above formalism can be generalized to include the correlation between polarization and temperature anisotropies, where the data, $\boldsymbol{d} = (\boldsymbol{d}_\mathcal{I}, \boldsymbol{d}_\mathcal{Q}, \boldsymbol{d}_\mathcal{U})$, are pixelized maps of the Stokes parameters, $\mathcal{I}$, $\mathcal{Q}$ and $\mathcal{U}$, where $\mathcal{I}$ here corresponds to the temperature anisotropy. The signal covariance matrix now has a block-diagonal structure in spherical harmonic space, with a 3×3 sub-matrix, for all multipole moments ℓ, as follows:

$$\mathbf{S}_\ell = \begin{pmatrix} C_\ell^{TT} & C_\ell^{TE} & 0 \\ C_\ell^{TE} & C_\ell^{EE} & 0 \\ 0 & 0 & C_\ell^{BB} \end{pmatrix}, \tag{5.40}$$

with the vanishing cross-spectra, C_ℓ^{TB} and C_ℓ^{EB}, set to zero.

To find the pure $\mathcal{B}$ map, we proceed as before, i.e. $\mathbf{S}_\mathcal{E} \to \lambda\mathbf{S}_\mathcal{E}$ with $\lambda \to \infty$. In this limit, the temperature and polarization components decouple, yielding the following signal covariance:

$$\mathbf{S}_\mathcal{B} \equiv \left[\lim_{\lambda\to\infty} \mathbf{S}(\lambda)^{-1}\right]^+ = \begin{pmatrix} C_\ell^{TT} - \frac{(C_\ell^{TE})^2}{C_\ell^{EE}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C_\ell^{BB} \end{pmatrix}, \tag{5.41}$$

such that the above Eq. (5.39) for the Wiener-filtered pure $\mathcal{B}$ map still holds.

A similar reasoning results in the following equation for the Wiener-filtered pure $\mathcal{E}$ map:

$$\boldsymbol{s}_{\rm WF}^{p\mathcal{E}} = \mathbf{S}_{\mathcal{E}}\mathbf{S}_{\mathcal{TE}}^{+}(\mathbf{S}_{\mathcal{TE}}^{+} + \mathbf{N}^{-1})^{-1}\mathbf{N}^{-1}\boldsymbol{d}, \tag{5.42}$$

where $\mathbf{S}_{\mathcal{TE}}$ is the signal covariance corresponding to:

$$\mathbf{S}_{\mathcal{TE}} \equiv \left[\lim_{\lambda\to\infty}\mathbf{S}(\lambda)^{-1}\right]^{+} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & C_\ell^{EE} - \frac{(C_\ell^{TE})^2}{C_\ell^{TT}} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.43}$$

$\mathbf{S}_{\mathcal{TE}}^{+}$ is then the pseudo-inverse associated to the subspace containing the temperature and $\mathcal{E}$ modes, i.e. all the $\mathcal{B}$ modes lie in the null space of both $\mathbf{S}_{\mathcal{TE}}$ and $\mathbf{S}_{\mathcal{TE}}^{+}$.

In this work, we encode this prescription in DANTE for optimal reconstruction of pure $\mathcal{E}$ and $\mathcal{B}$ maps via Wiener filtering. The numerical implementation for the pure $\mathcal{B}$ case entails the usual Wiener filtering procedure, i.e. solving Eq. (2.3), but assuming infinite covariance for the $\mathcal{E}$ component, followed by the application of the relevant projection operator $\mathbf{P}_{\mathcal{B}}$ to obtain the pure $\mathcal{B}$ map. An analogous procedure yields the pure $\mathcal{E}$ map.

The above formalism still holds in the presence of more complex noise models, such as the anisotropic correlated noise considered here. Since the signal covariance becomes fully diagonal when reconstructing the pure $\mathcal{E}$ or $\mathcal{B}$ map, we solve Eq. (5.22) itself since no basis transformations are required, as follows:

$$\boldsymbol{u} = (\bar{\mathbf{S}} + \mathbf{U})\left[\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B}(\bar{\mathbf{S}} + \mathbf{U}) + \phi\omega\mathbb{1}\right]^{-1}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{v}. \tag{5.44}$$

There is, nevertheless, a caveat in the implementation of the above equation. The signal covariance $\mathbf{S}$, as given by Eqs. (5.41) and (5.43), is actually the pseudo-inverse of the well-defined $\mathbf{S}(\lambda)^{-1}$ in the limit $\lambda \to \infty$. The trivial implementation is to set the relevant components of $\mathbf{S}_{\mathcal{TE}}$ and $\mathbf{S}_{\mathcal{B}}$ to a numerically large value. Alternatively, Eq. (5.44) can be expressed in the following numerically convenient form:

$$\boldsymbol{u} = \left[\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B} + \phi\omega\bar{\mathbf{S}}^{-1}(\bar{\mathbf{S}}^{-1} + \mathbf{U}^{-1})^{-1}\mathbf{U}^{-1}\right]^{-1}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{v}, \tag{5.45}$$

which, for the final step of the cooling scheme, i.e. $\mathbf{U} = \mathbf{0}$, reduces to:

$$\boldsymbol{u} = \left[\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Y}}\mathbf{B} + \phi\omega\mathbf{S}^{-1}\right]^{-1}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{v}. \tag{5.46}$$

We verified that both implementations resulted in identical solutions, within the limit of numerical errors.

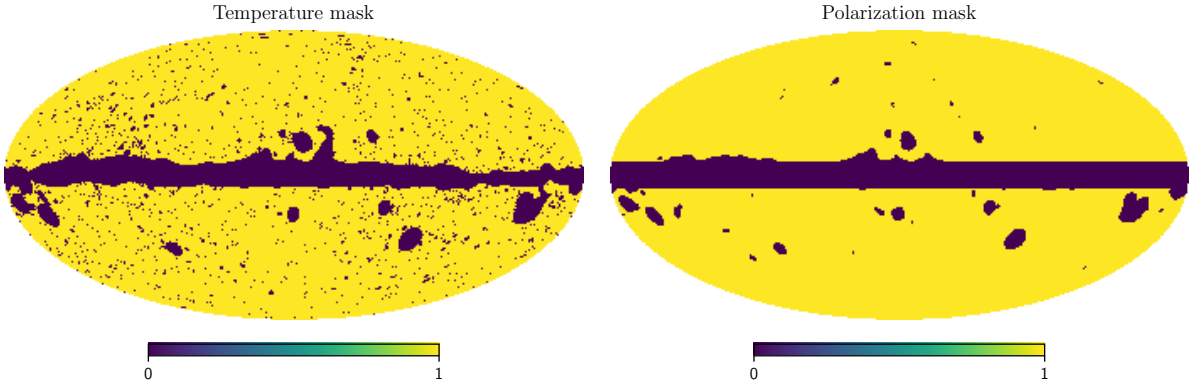


FIGURE 5.1: The temperature and polarization masks employed in the data analysis, corresponding to sky fractions of $f_{\rm sky}^{\mathcal{T}} = 0.76$ and $f_{\rm sky}^{\mathcal{P}} = 0.84$, respectively. Our artificially generated data set emulates the features of polarized Planck CMB maps.

5.5 Numerical experiments

In this section, we demonstrate the application of DANTE to an artificially generated but realistic CMB polarization data set. We present the procedure for the mock generation, followed by a description of the different steps in the data analysis pipeline.

5.5.1 Mock generation of polarized CMB maps

To simulate joint temperature and polarization maps on the sphere, we make use of HEALPY to generate realizations of $a_\ell^{\mathcal{T}}$, $a_\ell^{\mathcal{E}}$ and $a_\ell^{\mathcal{B}}$ signals with the correct covariance properties (cf. Eq. (5.40)), taking into account the correlation between CMB temperature anisotropy and polarization. We employed CAMB[3] (Lewis, Challinor, and Lasenby, 2000) to generate the input angular power spectra, C_ℓ^{TT}, C_ℓ^{EE}, C_ℓ^{BB} and C_ℓ^{TE}, from which the corresponding CMB signals are drawn. We assume a standard ΛCDM cosmology with the set of cosmological parameters ($\Omega_{\rm m} = 0.32$, $\Omega_\Lambda = 0.69$, $\Omega_{\rm b} = 0.05$, $h = 0.67$, $\sigma_8 = 0.83$, $n_{\rm s} = 0.97$) from Planck (Planck Collaboration et al., 2016b). We can then construct the input $\mathcal{Q}$ and $\mathcal{U}$ maps by transforming realizations of $\mathcal{E}$ and $\mathcal{B}$ signals (cf. Eq. (B.3) in Appendix B.2) on the sphere, with HEALPIX resolution of $N_{\rm side} = 128$ and $\ell_{\rm max} = 128$, such that the total number of pixels is $N_{\rm pix} = 12 \times N_{\rm side}^2 \approx 2 \times 10^5$. The input Stokes parameters' maps are subsequently contaminated with modulated correlated noise, as described by Eq. (5.47) below, with a white noise amplitude of $\sigma_N = 40.0\,\mu$K per pixel, typical of high-sensitivity CMB experiments tailored for the detection of $\mathcal{B}$ modes, and the corresponding $1/f$ noise parameters of $\ell_{\rm knee} = 10$ and $\alpha_{\rm knee} = 1.5$ (cf. Eq. (5.17)). We employ the SMICA Planck temperature and polarization masks (Planck Collaboration et al., 2016a), corresponding to sky fractions of $f_{\rm sky}^{\mathcal{T}} = 0.76$ and $f_{\rm sky}^{\mathcal{P}} = 0.84$, respectively, as depicted in Fig. 5.1. While our formalism and code account for the effect of a beam, we set the beam operator to identity for our present test cases.

[3] http://camb.info

5.5.2 Estimation of noise covariance

We now present a posterior optimization method to estimate the noise covariance using Monte Carlo (MC) simulations. For the case of modulated correlated noise, the data can be modelled as follows:

$$\boldsymbol{d} = \boldsymbol{s} + \mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{\mathcal{Y}}^{\dagger}\boldsymbol{n}, \tag{5.47}$$

following the notation of Eq. (2.1), where $\mathbf{D}$ is diagonal in pixel space. We also use liberally the notation $\mathbf{C}^{1/2}$ to indicate the positive square root matrix of $\mathbf{C}$. As described in Section 5.3.2, the noise covariance is now given by

$$\mathbf{N} = \mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}, \tag{5.48}$$

with $\mathbf{C}$ being the isotropic, homogeneous, noise covariance, which incorporates the inverse frequency $(1/f)$ noise correlation on the large scales. The overall aim is to estimate $\mathbf{D}$ and $\mathbf{C}$ using MC simulations by casting the covariance estimation problem as a two-level optimization scheme. The noise realizations can be modelled as the following linear combination:

$$\boldsymbol{n} = \mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{z} + \boldsymbol{k}, \tag{5.49}$$

where $\boldsymbol{z}$ and $\boldsymbol{k}$ are Gaussian random fields with covariances, $\langle \boldsymbol{z}\boldsymbol{z}^{\dagger}\rangle = \mathbb{1}$ and $\langle \boldsymbol{k}\boldsymbol{k}^{\dagger}\rangle = \kappa^2\mathbb{1}$, respectively. The corresponding χ^2, as the negative of the logarithm of the posterior distribution, with the sum over the contribution of each MC realization, can be written as:

$$\chi^2 = \sum_{i=1}^{\mathrm{N_{MC}}} \left[\frac{1}{\kappa^2} \left(\boldsymbol{n}_i - \mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{z}_i \right)^{\dagger} \left(\boldsymbol{n}_i - \mathbf{D}\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{z}_i \right) + \boldsymbol{z}_i^{\dagger}\boldsymbol{z}_i \right]. \tag{5.50}$$

To obtain the maximum *a posteriori* estimate of $\mathbf{D}$ and $\mathbf{C}$, we must optimize the above χ^2 with respect to $\boldsymbol{z}_i$ and $\mathbf{D}$, in the limit $\kappa \to 0$. The χ^2 optimization with respect to $\boldsymbol{z}_i$ yields, for a given MC simulation,

$$\tilde{\boldsymbol{z}}_i = \left(\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}^2\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2} + \kappa^2\mathbf{C}^{-1/2} \right)^{-1} \boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}\boldsymbol{n}_i, \tag{5.51}$$

which, in the limit $\kappa \to 0$, simplifies to

$$\begin{aligned} \lim_{\kappa\to 0} \tilde{\boldsymbol{z}}_i &= \left(\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}^2\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2} \right)^{-1} \boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}\boldsymbol{n}_i \\ &= \mathbf{C}^{-1/2}\boldsymbol{\mathcal{Y}}^{-1}\mathbf{D}^{-2}\tilde{\Pi}^{\dagger}\mathbf{D}\boldsymbol{n}_i, \end{aligned} \tag{5.52}$$

where $\tilde{\Pi}^{\dagger} = (\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{-1})^{\dagger}$ is a projector onto the pixel subspace. The inversion per matrix is acceptable for the term in parenthesis because the operation $\boldsymbol{\mathcal{Y}}^{\dagger}$ already projects on the subspace of maps bandwidth limited to ℓ_{max}. Within that space the $\boldsymbol{\mathcal{Y}}$ operator becomes invertible, though at some

cost. Optimizing the χ^2 with respect to $\mathbf{D}$ leads to

$$\begin{aligned}\widetilde{\mathbf{D}} &= \left[\sum_{i=1}^{\mathrm{N_{MC}}} \left(\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{z}_i\right)^\dagger \left(\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{z}_i\right)\right]^{-1} \sum_{i=1}^{\mathrm{N_{MC}}} \left(\boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{z}_i\right)^\dagger \boldsymbol{n}_i \\ &\equiv \left(\sum_{i=1}^{\mathrm{N_{MC}}} \boldsymbol{m}_i^\dagger \boldsymbol{m}_i\right)^{-1} \sum_{i=1}^{\mathrm{N_{MC}}} \boldsymbol{m}_i^\dagger \boldsymbol{n}_i \\ &\equiv \boldsymbol{\mathcal{E}}^{-1} \sum_{i=1}^{\mathrm{N_{MC}}} \boldsymbol{m}_i^\dagger \boldsymbol{n}_i, \end{aligned} \tag{5.53}$$

where we defined $\boldsymbol{\mathcal{E}} \equiv \sum_i \boldsymbol{m}_i^\dagger \boldsymbol{m}_i$, with $\boldsymbol{m}_i$ estimated as follows:

$$\boldsymbol{m}_i = \boldsymbol{\mathcal{Y}}\mathbf{C}^{1/2}\boldsymbol{z}_i = \tilde{\Pi}\mathbf{I}\mathbf{D}^{-2}\tilde{\Pi}\Pi^\dagger \mathbf{D}\boldsymbol{n}_i, \tag{5.54}$$

where $\tilde{\Pi}\mathrm{I} = \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{-1}$ is the projector onto the spherical harmonic space.

The algorithm for the noise covariance estimator proceeds according to the following iterative scheme: Compute $\tilde{\boldsymbol{m}}_i$ using Eq. (5.54), and subsequently $\boldsymbol{\mathcal{E}}^{-1}$ to obtain $\widetilde{\mathbf{D}}$ using Eq. (5.53), followed by a power spectrum update to obtain $\widetilde{\mathbf{C}}$ via $\widetilde{\mathbf{C}} = \sum_i \langle \hat{\boldsymbol{m}}_i \hat{\boldsymbol{m}}_i^\dagger \rangle / \mathrm{N_{MC}}$. In the above, we have defined the harmonic representation of the map with $\hat{\boldsymbol{m}}_i = \boldsymbol{\mathcal{Y}}^{-1}\boldsymbol{m}_i$. We solve Eq. (5.53) by implementing fixed point iterations, but this fixed point is not an attractor. We consequently employ the Babylonian method (e.g. Fowler and Robson, 1998; Friberg, 2007) to stabilize the fixed point and obtain an updated $\widetilde{\mathbf{D}}$, as follows:

$$\widetilde{\mathbf{D}} = \frac{1}{2}\left(\mathbf{D} + \boldsymbol{\mathcal{E}}^{-1} \sum_{i=1}^{\mathrm{N_{MC}}} \boldsymbol{m}_i^\dagger \boldsymbol{n}_i\right). \tag{5.55}$$

We may verify that the fixed point of the above equation is exactly the desired $\mathbf{D}$ matrix. Note that due to the degeneracy between the amplitudes of $\mathbf{C}$ and $\mathbf{D}$, we need to anchor the amplitude of the updated $\widetilde{\mathbf{C}}$ via the required re-scaling.

We therefore solve the above Eq. (5.53) iteratively, using $\mathbf{D}^2 = \sum_i (\boldsymbol{n}_i^\dagger \boldsymbol{n}_i)/\mathrm{N_{MC}}$ as an initial guess. We generate 10^4 MC noise simulations using as template the modulated noise covariance provided by the Planck data analysis pipeline,[4] as an estimate for the $\boldsymbol{n}_i$ above. The map estimates, after only five iterations, for the diagonal and off-diagonal components of the noise covariance matrix, along with their corresponding reference and residual maps, are displayed in Figs. 5.2 and 5.3, respectively. Visually, the distinct components of the covariance matrix are adequately recovered, with residuals at the level of $\sim 0.3\%$ and $\sim 6\%$ for the diagonal and off-diagonal components, respectively. As a quantitative diagnostic, we verify the relative deviation in the angular power spectra reconstructed from the maps, as a function of scale, with respect to their reference components: $\sqrt{C_\ell(\hat{\mathbf{D}} - \mathbf{D}_{\rm ref})/C_\ell(\mathbf{D}_{\rm ref})}$, illustrated in Fig. 5.4. This demonstrates the accuracy of reconstruction of our noise covariance estimator across the range of scales considered, with only five iterations.

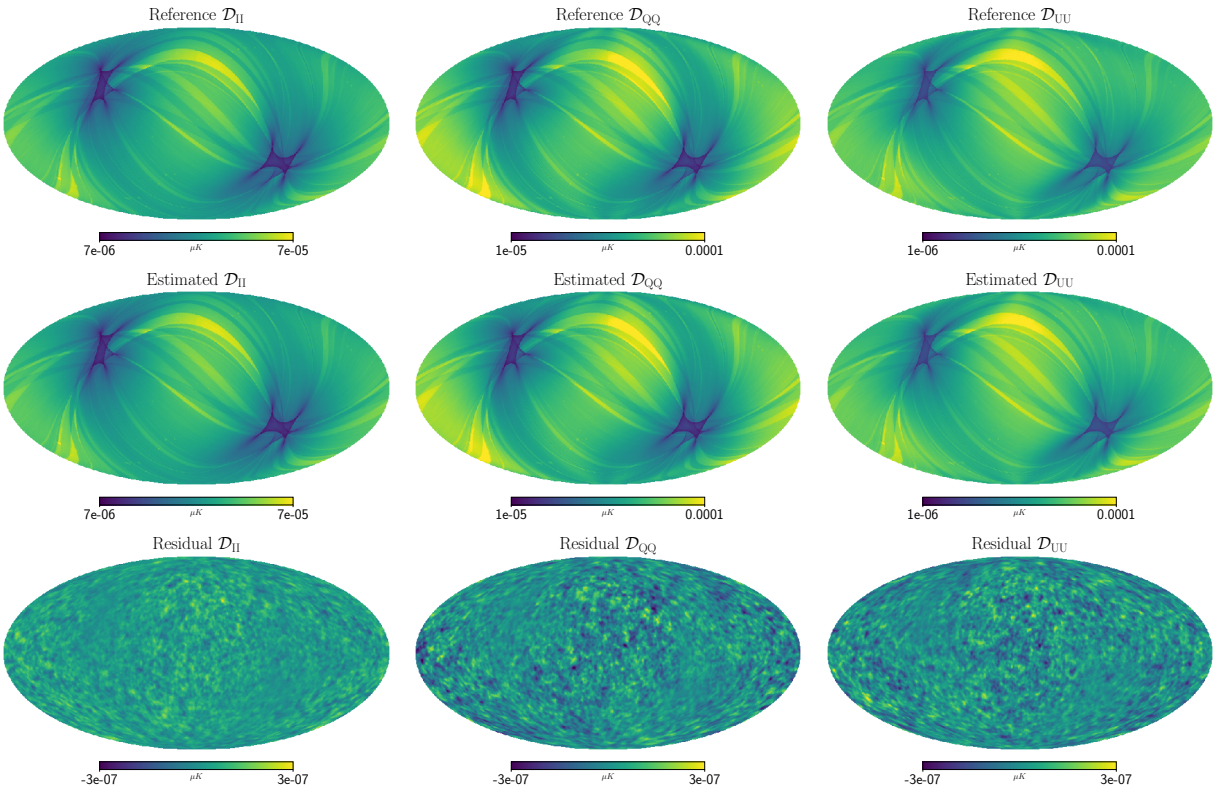


FIGURE 5.2: *Top row:* The reference modulated noise covariance used as inputs for the generation of the noise simulations. *Middle row:* The corresponding components recovered by the noise covariance estimator display the expected modulation patterns, indicating qualitatively the efficacy of our estimator. *Bottom row:* The residuals, generated by computing the difference between the reference and estimated noise covariance components, demonstrate the high-fidelity reconstructions. Note that for the relatively low residuals at the level of $\sim 0.3\%$ to be visible, we employ a different colour bar scale for the residual maps.

5.5.3 Analysis of CMB polarization on the sphere

In this section, we showcase the application of DANTE in Wiener filtering polarized CMB maps contaminated with anisotropic correlated noise, and also illustrate its efficacy in generating pure $\mathcal{E}$ and $\mathcal{B}$ maps, guaranteed to be free from any cross-contamination. This corresponds to three distinct runs using the same realization of mock data, generated as described above in Section 5.5.1, labelled as "WF", "pure $\mathcal{E}$" and "pure $\mathcal{B}$", respectively. The "pure $\mathcal{B}$" run yields a "pure" temperature map as by-product, which corresponds to the map of temperature anisotropies without any contribution from the $\mathcal{E}$ modes. We anchor the choice of hyperparameter values, described below, for all three cases.

We make an initial truncation in the power spectrum at $\ell = 50$, corresponding to a given value of $\boldsymbol{\xi}$ and the algorithm loops through the iterations until the fractional difference between successive iterations has reached a sufficiently low value, at which point $\boldsymbol{\xi}$ is reduced by a constant factor according to a given cooling scheme: $\boldsymbol{\xi} \to \boldsymbol{\xi}\eta$, where $\eta = 2/3$, until $\boldsymbol{\xi} \to \phi\omega\mathbb{1}$. We implement a "weak" criterion for convergence, $\|\boldsymbol{s}_{i+1} - \boldsymbol{s}_i\| \,/\, \|\boldsymbol{s}_i\| < \epsilon$, where $\epsilon = 10^{-5}$, as a cheap proxy for the strong criterion that is verified *a posteriori* in Fig. 5.7.

[4] Available from http://pla.esac.esa.int/pla/aio/product-action?MAP.MAP_ID=HFI_SkyMap_100_2048_R2.02_full.fits

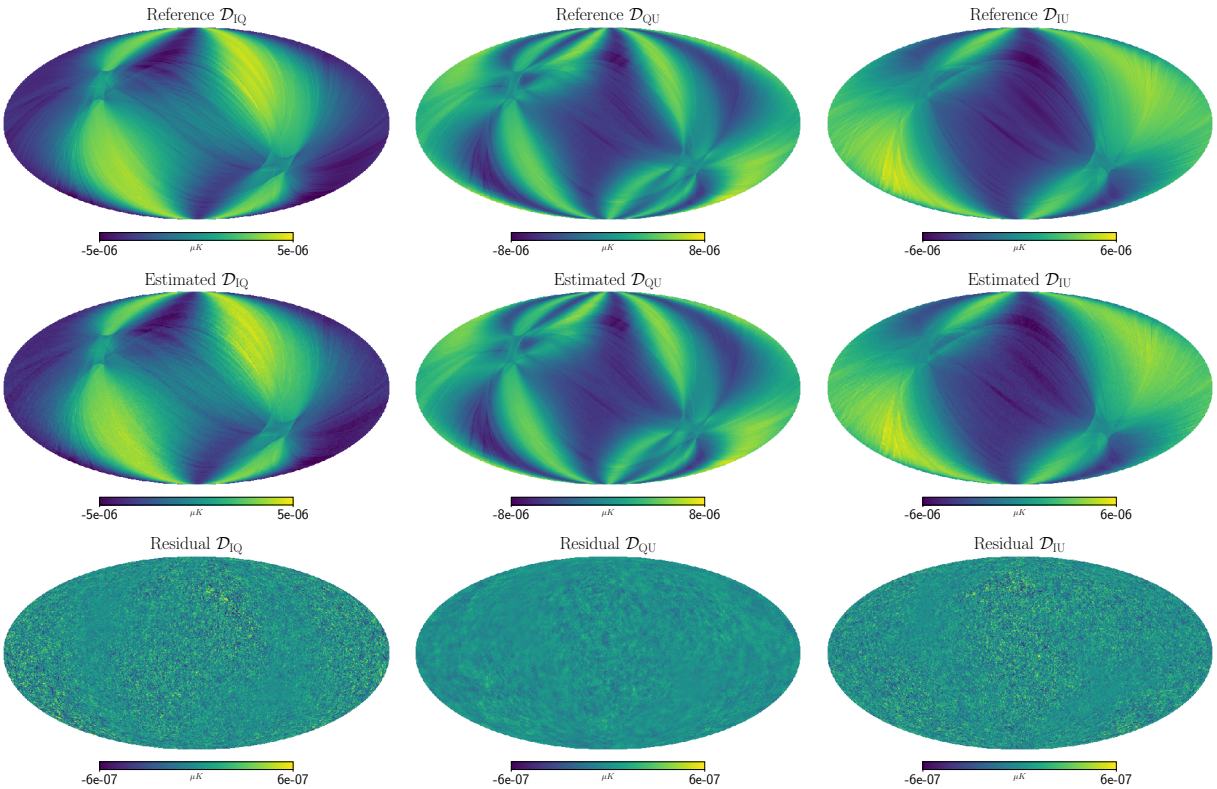


FIGURE 5.3: Same as Fig. 5.2, except for the off-diagonal components of the modulated noise covariance. As for their diagonal counterparts, the cross-correlations components are adequately recovered, with relatively insignificant residuals ($\sim 6\%$).

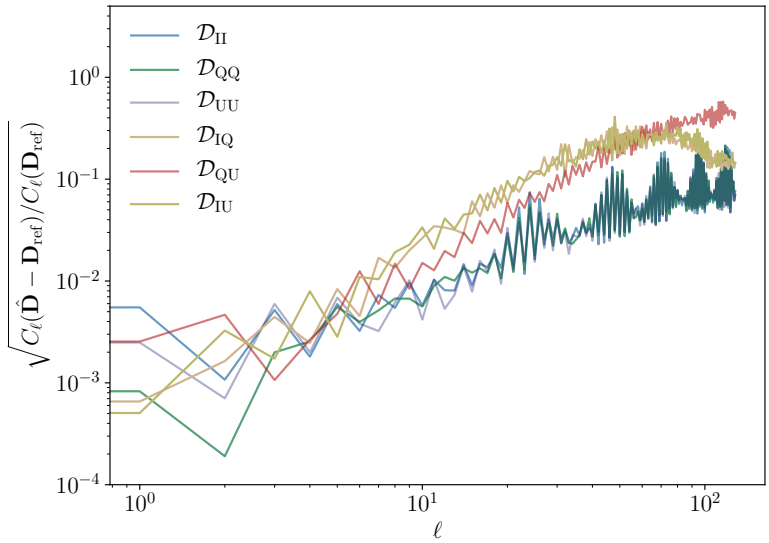


FIGURE 5.4: Relative deviation, as a function of scale, of the estimated components of the covariance matrix, with respect to their reference values. This diagnostic quantitatively demonstrates the performance of our noise covariance estimator with only five iterations.

The reconstructed angular power spectra for the temperature and polarization components are provided in Fig. 5.5, with the WF solution showing suppressed power on the small scales resulting from the noise and masked regions of the sky, which is a characteristic feature of Wiener filtering. For the temperature anisotropies, depicted in the left panel, the pure $\mathcal{B}$ run yields the temperature

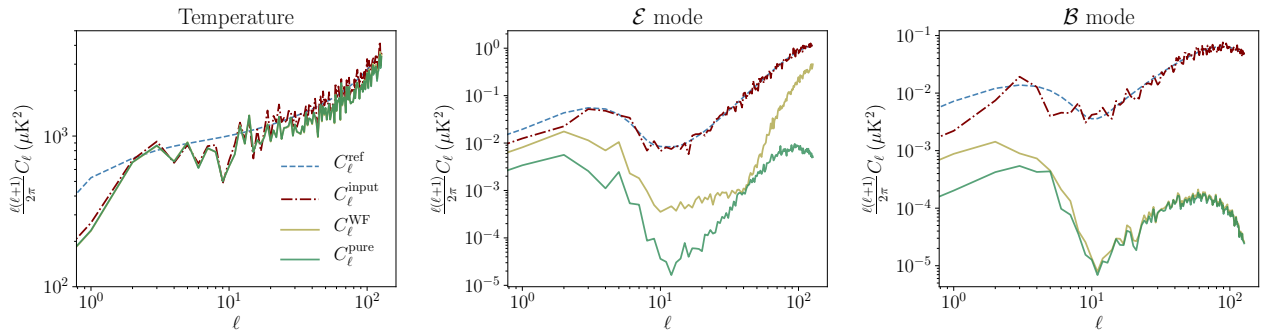


FIGURE 5.5: Reconstructed temperature, $\mathcal{E}$- and $\mathcal{B}$-mode angular power spectra from the WF and pure $\mathcal{E}/\mathcal{B}$ runs. The simulated realizations, depicted using dash-dotted lines, were drawn from the reference power spectra, denoted by dashed lines, and were subsequently contaminated with anisotropic correlated noise and masked. *Left panel:* The Wiener-filtered C_ℓ^{TT} is slightly suppressed on the small scales, as expected, in the low signal-to-noise regime. With the discarded $\mathcal{E}$-mode contribution being relatively low, the pure C_ℓ^{TT} matches the Wiener-filtered version. *Middle panel:* The pure C_ℓ^{EE} is substantially different from its Wiener-filtered counterpart, as the contribution of temperature anisotropies, by virtue of their larger power, is especially significant. *Right panel:* The contrast between the Wiener-filtered and pure C_ℓ^{BB} is more significant on the largest scales, as can be seen from their real-space maps displayed in the bottom row of Fig. 5.6.

power spectra that has been purified with respect to the $\mathcal{E}$ modes, and as such, corresponds to the prediction solely from the temperature data, with no contribution from the polarization component. This pure temperature power spectrum does not display any significant difference compared to the Wiener-filtered one, as expected, due to the relatively small $\mathcal{E}$-mode contribution. The corresponding reconstructed power spectra for the $\mathcal{E}$ modes are depicted in the middle panel of Fig. 5.5. The pure $\mathcal{E}$-mode power spectrum displays a smooth functional behaviour that matches the shape of the input power spectrum, although substantially suppressed because of the noisy and masked regions and since the discarded temperature contribution is significant. The right panel displays the corresponding reconstructions for the $\mathcal{B}$-mode power spectrum. The pure reconstruction, as in the WF case, suppresses the modes in the low signal-to-noise regime, while also discarding the ambiguous modes, and shows a notable difference on the largest scales.

The real-space maps of the Wiener-filtered, pure $\mathcal{E}$ and $\mathcal{B}$ modes, together with their corresponding simulated maps, are illustrated in the middle and bottom rows of Fig. 5.6, respectively. The corresponding temperature maps are also displayed in the top row, for completeness. Note that the simulated maps are the generated signal realizations which were subsequently contaminated with anisotropic correlated noise and masked according to Fig. 5.1. The Wiener-filtered map, as the maximum *a posteriori* reconstruction, represents the CMB signal content of the data, with the reconstruction of the large-scale modes in the masked areas, based on the information content of the observed sky regions, being a natural consequence of Wiener filtering. Both the Wiener-filtered and pure temperature maps are similar in appearance, as expected from their reconstructed power spectra (cf. left panel of Fig. 5.5). This is not the case, however, for the reconstructed $\mathcal{E}$ maps, purified with respect to the temperature anisotropies and ambiguous modes. The pure $\mathcal{E}$ map, as a result, has a lower signal amplitude. Both the Wiener-filtered and pure $\mathcal{B}$ maps show the clear reconstruction of the large-scale modes with extremely low amplitude. The striking contrast between the two maps is on the largest scales near the mask, where the pure $\mathcal{B}$ map has lower power. This is

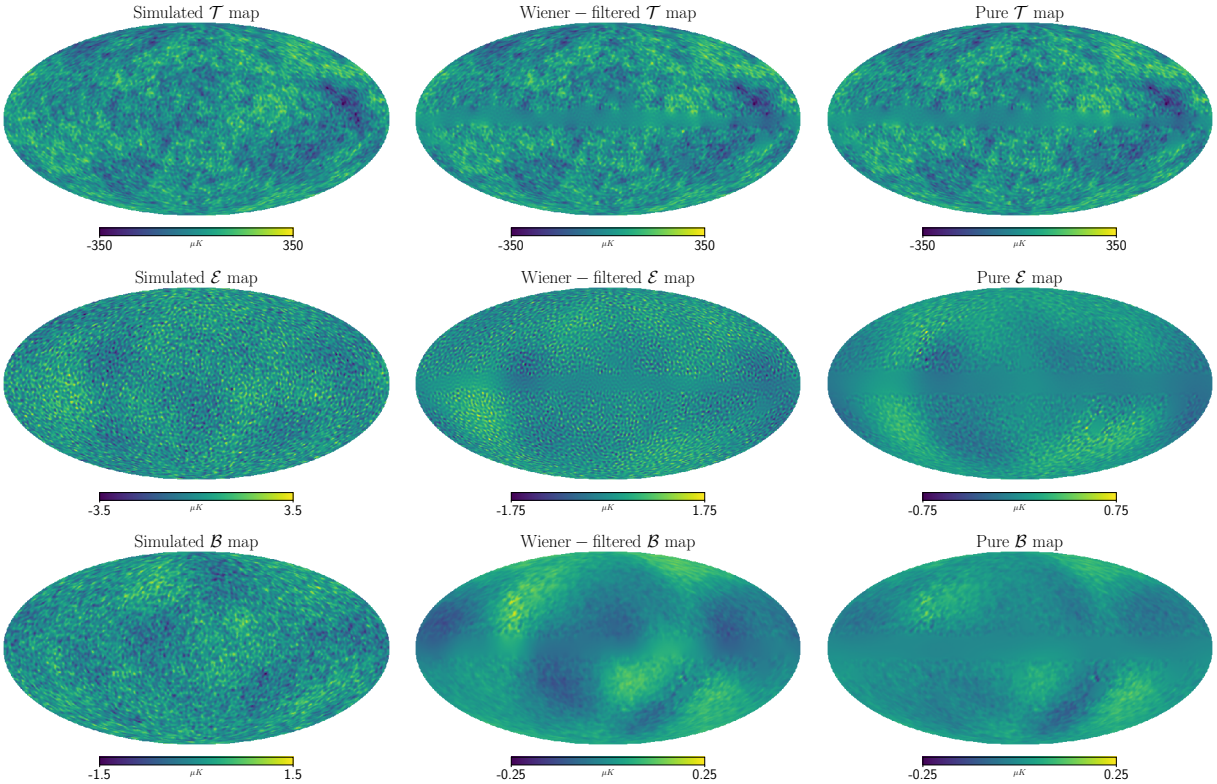


FIGURE 5.6: Simulated and reconstructed real-space maps of temperature anisotropies, $\mathcal{E}$ and $\mathcal{B}$ modes, from top to bottom, from the WF and pure $\mathcal{E}/\mathcal{B}$ runs. The Wiener-filtered maps for all three components exhibit the characteristic feature, whereby the signal is extrapolated into the masked regions. The level of anisotropic correlated noise smooths out the small-scale features for the $\mathcal{E}$ and $\mathcal{B}$ modes, with the suppression of the small-scale power being more significant for the latter due to its low amplitude. The pure $\mathcal{E}$ and $\mathcal{B}$ maps, after the removal of ambiguous modes, display a reduced signal content. This difference is more striking for the $\mathcal{E}$ map since the contribution from the temperature and $\mathcal{E}$-mode correlations is also discarded. The pure $\mathcal{B}$ map has lower power close to the masked regions, relative to the Wiener-filtered one, as expected, since ambiguous modes are known to have support primarily near the mask.

consistent with previous work (e.g. Bunn and Wandelt, 2017), where it was found that ambiguous modes have support primarily near the masked regions.

We illustrate the convergence behaviour of the three solutions via the corresponding variations of their residual error given by $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| \,/\, \|\boldsymbol{y}\|$, for a linear system of equations given by $\mathcal{A}\boldsymbol{x} = \boldsymbol{y}$, in Fig. 5.7. This residual error adequately characterizes the accuracy of the final solution, with the relevant equations as follows:

$$\boldsymbol{y} = \mathcal{S}^{1/2}\mathfrak{R}\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}^{-1}(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}\mathbf{D}^{-1}\boldsymbol{d} \tag{5.56}$$

and

$$\mathcal{A} = \mathbb{1} + \mathcal{S}^{1/2}\mathfrak{R}\boldsymbol{\mathcal{Y}}^{\dagger}\mathbf{D}^{-1}(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}\mathbf{D}^{-1}\boldsymbol{\mathcal{Y}}\mathfrak{R}^{\dagger}\mathcal{S}^{1/2}, \tag{5.57}$$

with $\boldsymbol{x} = \mathcal{S}^{-1/2}\mathfrak{R}\boldsymbol{s}$, following the notation from Section 5.3.2. The coupling matrix $(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^{\dagger})^{-1}$ requires Jacobi iterations for accurate evaluation of the above residual error. For the pure $\mathcal{E}/\mathcal{B}$ runs, this is non-trivial as the signal covariance $\mathbf{S}$ should have infinite values, as mentioned at end of Section 5.4.3. To simplify the residual error evaluations in these cases, we simply set the relevant

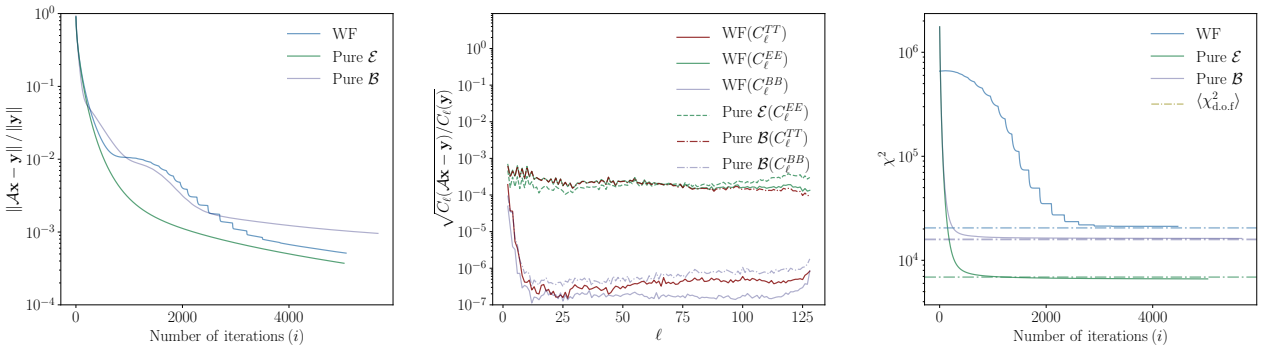


FIGURE 5.7: Convergence diagnostics for the reconstructions from the three runs. *Left panel:* Variation of residual error, given by $\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\| \,/\, \|\boldsymbol{y}\|$, as a function of iterations. This residual error is reduced monotonically as the iterations proceed, demonstrating the unconditional stability of the dual messenger algorithm in performing the three reconstructions. *Middle panel:* Variation of residual error as a function of angular scale for the final solutions. This error is sufficiently low for the range of scales considered, indicating the quality of the respective solutions. *Right panel:* Variation of χ^2 with number of iterations for the WF and pure $\mathcal{E}/\mathcal{B}$ solutions. Their respective χ^2 drop to a final value which is consistent with $\langle\chi^2_{\rm d.o.f}\rangle$ corresponding to the expectation value of the χ^2, given by the number of degrees of freedom (d.o.f), for the final solutions.

components of $\mathbf{S}$ to a numerically large value.

As we demonstrated in our previous work, a characteristic feature of the dual messenger algorithm is the monotonic decrease in the residual error as the iterations proceed, as illustrated in the left panel of Fig. 5.7, thereby demonstrating the unconditional stability of our method. This residual error, as a function of angular scale, for the final solutions from the three runs, are depicted in the middle panel. The Jacobi relaxation schemes employed is required to reduce this error to extremely low values across the range of scales considered.

The χ^2 is computed as follows:

$$\chi^2 = (\boldsymbol{d} - \boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{s})^\dagger \mathbf{D}^{-1} (\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger)^{-1} \mathbf{D}^{-1} (\boldsymbol{d} - \boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{s}) + \boldsymbol{s}^\dagger \mathbf{S}^{-1} \boldsymbol{s}, \tag{5.58}$$

where, as for the residual error evaluations above, we employ Jacobi relaxation for the composite operation $(\boldsymbol{\mathcal{Y}}\mathbf{C}\boldsymbol{\mathcal{Y}}^\dagger)^{-1}$. The corresponding χ^2 variation for the three different solutions is displayed in the right panel of Fig. 5.7. In all three cases, the respective χ^2 of the dual messenger solutions drop to a final value which matches $\langle\chi^2_{\rm d.o.f}\rangle$, the expectation value of the χ^2, given by the number of degrees of freedom (d.o.f), for the final solution. In the absence of masks, $\langle\chi^2_{\rm d.o.f}\rangle$ is given by the total number of harmonic modes of the temperature, $\mathcal{E}$ and $\mathcal{B}$ components. The computation of $\langle\chi^2_{\rm d.o.f}\rangle$ is, however, non-trivial when masks are involved. We estimated $\langle\chi^2_{\rm d.o.f}\rangle$ via Monte Carlo simulations. The convergence diagnostics discussed above, therefore, quantitatively demonstrate the efficacy of DANTE in performing the three distinct tasks.

Concerning the execution times for the WF, pure $\mathcal{E}$ and pure $\mathcal{B}$ runs, for the specific test case investigated, the algorithm runs to completion on four cores of a standard workstation, Intel Core i5-4690 CPU (3.50 GHz), in around three hours. Note that a conjugate gradient method can, in principle, deal with such anisotropic noise models, provided that an adequate preconditioner can be found and this is the major stumbling block. Devising an appropriate preconditioner for such

a complex problem is an extremely challenging task. For instance, the multi-grid preconditioner developed by Smith, Zahn, and Doré (2007) at WMAP resolution and sensitivity is already highly non-trivial. The preconditioner-free approach of the dual messenger algorithm, therefore, is the key advantage.

5.6 Conclusions and outlook

We presented a numerically robust and fast code, DANTE, for pure $\mathcal{E}/\mathcal{B}$ decomposition of CMB polarization maps. It accounts for complex and realistic noise models such as anisotropic correlated noise, encountered in typical CMB experiments such as Planck. DANTE is an augmented version of our dual messenger algorithm, adapted for the reconstruction of pure full-sky $\mathcal{E}$ and $\mathcal{B}$ maps on the sphere. The algorithm encodes a new method for the pure-ambiguous decomposition, based on a Wiener filtering approach, recently proposed by Bunn and Wandelt (2017), that guarantees no cross-contamination between the two maps. We also developed a noise covariance estimator to reconstruct the components of anisotropic noise covariance from Monte Carlo simulations, as required by the dual messenger algorithm.

We have demonstrated the capabilities of DANTE in dealing with large data sets and the associated high-dimensional covariance matrices. Moreover, as a preconditioner-free method, it is not hindered by ill-conditioned systems of equations inherent in CMB polarization problems, unlike standard PCG solvers, as demonstrated in KLW18. DANTE also has an in-built option for drawing constrained Gaussian realizations of the CMB sky, for applications requiring homogeneous coverage of the field of observations. We have not illustrated this particular feature in this work as this was shown previously in KLW18. DANTE will be rendered public in the near future.

The pure $\mathcal{E}/\mathcal{B}$ decomposition framework implemented in this work, as a maximum *a posteriori* probability approach, has several advantages over traditional methods. It exploits the sparsity of the $\mathcal{E}/\mathcal{B}$ decomposition in the spherical harmonic basis, rendering the implementation extremely efficient. It is therefore much faster and straightforward than methods relying on the construction of orthonormal bases or wavelet methods that require a certain degree of fine-tuning. Moreover, $\mathcal{E}/\mathcal{B}$ purification in the context of pseudo-C_ℓ estimators is only feasible when the mask is differentiable up to at least its second derivatives, which is usually achieved via an appropriate apodization (Alonso et al., 2019). An interesting aspect of our approach is that it is not hindered by such limitations.

We have showcased the performance of DANTE on a realistic mock data set, emulating the features of polarized Planck CMB maps. The next step in this series of investigations is to further augment DANTE with an adaptive upgrade. Despite the improvements made to render the analysis of high-resolution CMB polarization data sets numerically feasible, the statistically optimal approach for the separation of $\mathcal{E}$ and $\mathcal{B}$ modes requires exact global analyses such as Gibbs sampling. This would, however, require several applications of the Wiener filter to obtain one signal realization conditional on the polarization data (e.g. Larson et al., 2007). The algorithm would therefore benefit from a further level of sophistication. A particularly interesting upgrade is to exploit the hierarchical framework of the dual messenger algorithm by adapting the working resolution

progressively during execution, thereby substantially reducing the computation time. We also intend to explore the possibility of employing the dual messenger as a preconditioner in a standard PCG approach, in an attempt to drastically improve the convergence rate. Our algorithm could also be used to provide better examples of Wiener-filtered maps for a filtering based on machine learning, as in Münchmeyer and Smith (2019) with the J_1 loss function. These examples would be much more expensive than the purely simulation-based approach of the J_2 loss function, but would also provide a solid validation step.

Ultimately, the underlying objective is to employ this efficient tool in exact global Bayesian analyses of high-resolution and high-sensitivity CMB observations from the latest release of Planck to yield scientific products of significant value and interest. The resulting reconstructed maps may potentially be employed for various applications such as power spectrum reconstruction, estimation of lensing potential and searches for non-Gaussianity and statistical anisotropy. Another key aspect is that the features of the real-space pure $\mathcal{B}$ maps allow the characterization of lensing-induced $\mathcal{B}$ modes which go beyond the power spectrum.

DANTE can be easily applied to other CMB data sets in straightforward fashion without major modifications of the source code, making it a potentially powerful and robust tool for other current and future high-resolution CMB missions such as South Pole Telescope, Advanced ACTPol, Simons Observatory and CMB-S4. The flexibility of the code can nevertheless be exploited in other cosmological contexts, due to the ubiquitous use of the Wiener filter, and even in more general scenarios involving spin field reconstruction.

Part III

Large-scale structures of the Universe

Chapter 6

Dynamics of cosmic structure formation and evolution

6.1 Growth of density perturbations

The study of cosmic structure formation, in essence, requires the modelling of the dynamics of a gravitating system in an expanding Universe. We provide a mathematical description of the physical evolution of cosmic structures in terms of theoretical perturbative analyses, followed by a brief discussion of numerical simulations to compute the highly non-linear dynamics in regimes where analytical solutions are no longer feasible.

6.1.1 Physical description of non-linear dynamics

To model the growth of cosmic structures, standard approaches to study the dynamics of density and velocity fields are based on the assumption that cold dark matter is well described by a fluid. The study of the fluid motion by focusing on specific locations in space to describe the time evolution of the fluid flow constitutes the basis of *Eulerian perturbation theory* (e.g. Bernardeau et al., 2002).

An alternative approach is the Lagrangian description, where the description of the field relies on the determination of the trajectories of particles, with this physical description of structure growth yielding the *Lagrangian perturbation theory* (LPT) (e.g. Bouchet et al., 1995), which is employed in Chapters 8, 9 and 12 as an approximate physical model of cosmic structure formation. In the following sections, we outline the physical description of large-scale structures to the degree required for the following chapters. An excellent in-depth review of Eulerian and Lagrangian perturbation theory, including non-linear approximations to gravitational instability, is provided in Leclercq (2015).

Linear structure formation

In the linear regime, on large scales and during the early stages of gravitational evolution, under the assumption that density fluctuations are relatively small, cosmic structure formation is governed by a homogeneous growth function $D^+(a)$ acting on a density contrast according to $\delta(\boldsymbol{x}, a) = D^+(a)\delta(\boldsymbol{x}, a = 1)$. The growth factor $D^+(a)$ is obtained as the numerical solution to the linear growth equation, for a general cosmology (e.g. Turner and White, 1997; Wang and Steinhardt,

1998; Linder and Jenkins, 2003):

$$\frac{\mathrm{d}^2 D^+(a)}{\mathrm{d}a^2} + \frac{1}{a}\left(3 + \frac{\mathrm{d}\ln H}{\mathrm{d}\ln a}\right)\frac{\mathrm{d}D^+(a)}{\mathrm{d}a} - \frac{3}{2}\frac{\Omega_\mathrm{m}(a)D^+(a)}{a^2} = 0, \tag{6.1}$$

where the cosmological parameter $\Omega_\mathrm{m}(a)$ depends on the scale factor (cf. Eq. (1.7)).

Lagrangian perturbation theory

The Lagrangian framework is based on the *displacement field* $\boldsymbol{\Psi}(\boldsymbol{q})$, rather than the position of particles as in the Eulerian approach, yielding the Lagrangian coordinates. The latter label an infinite number of infinitesimally light particles, which are homogeneously distributed in the limit of a completely homogeneous Universe for redshift $z_i \to \infty$. A perturbation in the particle positions, triggered by a given physical phenomenon, results in a perturbation in density. The key point is that the corresponding displacements are non-zero at very high redshifts. τ is the conformal time, defined by $\mathrm{d}t = a(\tau)\mathrm{d}\tau$, $\mathcal{H} \equiv a'/a = aH$ indicates the conformal expansion rate.

The displacement field maps the initial comoving position $\boldsymbol{q}$ of a particle to its final (Eulerian) position $\boldsymbol{x}(\boldsymbol{q}, a_i)$ at a later time (e.g. Buchert, 1989; Bouchet et al., 1995; Bernardeau et al., 2002):

$$\boldsymbol{x}(\boldsymbol{q}, a_i) = \boldsymbol{q} + \boldsymbol{\Psi}(\boldsymbol{q}, a_i), \tag{6.2}$$

such that the physical position is given by $\boldsymbol{r} = a\boldsymbol{x}$ and the conformal velocity described by $\boldsymbol{u} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\tau$. Inserting the above parameterization in the equation of dynamics of perturbations with a FLRW metric yields

$$\frac{\partial \boldsymbol{u}}{\partial \tau} + \mathcal{H}(\tau)\boldsymbol{u} = -\boldsymbol{\nabla}_x \Phi, \tag{6.3}$$

where Φ is the cosmological gravitational potential and $\boldsymbol{\nabla}_x$ is the gradient operator in (Eulerian) comoving coordinates $\boldsymbol{x}$,

$$\Delta\Phi = \frac{3}{2}\Omega_\mathrm{m}(\tau)\mathcal{H}^2(\tau)\delta_\mathrm{m}(\boldsymbol{x}) \tag{6.4}$$

is the cosmological Poisson equation, and

$$\delta_\mathrm{m}(\boldsymbol{x}) \equiv \frac{\bar{\rho}}{\rho} - 1 = \left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{q}}\right|^{-1} - 1 \tag{6.5}$$

is the density contrast expressed in terms of the Jacobian of the transformation between Eulerian and Lagrangian coordinates, obtained via the conservation of mass (continuity equation) in the relationship between the density contrast and the trajectories, i.e. $\rho(\boldsymbol{x}, \tau)\mathrm{d}^3\boldsymbol{x} = \rho(\boldsymbol{q})\mathrm{d}^3\boldsymbol{q}$.

Zel'dovich Approximation

A first order expansion in the displacement field $\boldsymbol{\Psi}$ leads to the equation of motion in Lagrangian coordinates:

$$\boldsymbol{\nabla} \cdot \left(\frac{\partial^2 \boldsymbol{\Psi}}{\partial \tau^2} + \mathcal{H}(\tau)\frac{\partial \boldsymbol{\Psi}}{\partial \tau}\right) = \frac{3}{2}\Omega_\mathrm{m}(\tau)\mathcal{H}^2(\tau)\boldsymbol{\nabla} \cdot \boldsymbol{\Psi}. \tag{6.6}$$

Defining $\psi \equiv \boldsymbol{\nabla} \cdot \boldsymbol{\Psi}$, and assuming that $\boldsymbol{\Psi}$ is vorticity-free, the above equation can be simplified by separating the temporal and spatial dependence to

$$\psi'' + \mathcal{H}(\tau)\psi' - \frac{3}{2}\Omega_{\rm m}(\tau)\mathcal{H}^2(\tau)\psi = 0, \tag{6.7}$$

where the prime denotes differentiation with respect to τ.

The solution to the above second order differential equation is the sum of two terms; strictly growing and decaying functions, D_1^+ and D_1^-, respectively. The former encodes the essential physics that leads to the clustering of structures, from small perturbations to large concentrations of matter. The linear solution can be expressed as

$$\psi^{(1)}(\boldsymbol{q}, \tau) \equiv \nabla_{\boldsymbol{q}} \cdot \Psi^{(1)}(\boldsymbol{q}, \tau) = -D_1^+(\tau)\delta(\boldsymbol{q}), \tag{6.8}$$

where $\delta(\boldsymbol{q})$ is the growing mode of the initial density contrast in Lagrangian coordinates. Considering only this specific part of the solution, the coordinate transformation, to first order, behaves as:

$$\boldsymbol{x}(\boldsymbol{q}, a) = \boldsymbol{q} + \frac{D_1^+(a)}{D_1^+(a_0)}\boldsymbol{\Psi}_0(\boldsymbol{q}), \tag{6.9}$$

where $\boldsymbol{\Psi}_0(\boldsymbol{q})$ is the Lagrangian displacement at some cosmic scale a_0. The above equation is the well-known *Zel'dovich approximation* (ZA) (Zel'Dovich, 1970), which describes the particle dynamics as straight inertial motion following the initial velocity vector and at first glance, implies ballistic motion. This is not the case, however, as the linear growth factor D_1^+ incorporates feedback effect from cosmic structure formation and cosmic expansion.

Second-order Lagrangian perturbation theory

Extensions to this approximation have been developed extensively in the literature, with the immediate refinement brought by an expansion to higher orders (e.g. Bouchet et al., 1995). Due to the ZA being local, it does not provide a sufficient physical description at non-linear stages when particles form gravitationally bound structures, rather than follow straight trajectories. The addition of second order terms, therefore, is prerequisite to account for the fact that gravitational instability is non-local via the inclusion of corrections due to gravitational tidal effects (Bernardeau et al., 2002), and this provides a remarkable improvement in describing the global properties of density and velocity fields (Melott, Buchert, and Weiss, 1995). Second-order Lagrangian perturbation theory (LPT) yields the following solution:

$$\boldsymbol{x}(\boldsymbol{q}, a) = \boldsymbol{q} + \Psi(\boldsymbol{q}, a) = \boldsymbol{q} + \Psi^{(1)}(\boldsymbol{q}, a) + \Psi^{(2)}(\boldsymbol{q}, a), \tag{6.10}$$

where the divergence of the first-order solution resulting in the ZA, i.e. Eq. (6.8), with the divergence of the second-order solution describing the tidal effects,

$$\psi^{(2)}(\boldsymbol{q}, \tau) = \nabla_{\boldsymbol{q}} \cdot \Psi^{(2)}(\boldsymbol{q}, \tau) = \frac{1}{2}\frac{D_2^+(\tau)}{[D_1^+(\tau)]^2}\sum_{i \neq j}\left[\Psi_{i,i}^{(1)}\Psi_{j,j}^{(1)} - \Psi_{i,j}^{(1)}\Psi_{j,i}^{(1)}\right], \tag{6.11}$$

where $\Psi^{(1)}_{i,j} \equiv \partial \Psi^{(1)}_i / \partial \boldsymbol{q}_j$ and $D_2(\tau)$ is the second-order growth factor given as

$$D_2^+(a) \approx -\frac{3}{7}\left[D_1^+(a)\right]^2 \Omega_{\mathrm{m}}^{-\frac{1}{143}}, \tag{6.12}$$

which is valid for a flat ΛCDM Universe (Bouchet et al., 1995; Scoccimarro, 1998; Bernardeau et al., 2002).

6.1.2 Numerical simulations of cosmic structure growth

While LPT provides an adequate description of the growth of cosmic structures on the largest scales, it is no longer valid on the smaller scales since it relies on the convergence of series expansion, as described in the previous section. As such, this expansion cannot accurately describe the dynamics of gravitational evolution of dark matter on scales $r < 10h^{-1}$ Mpc (e.g. Melott, Buchert, and Weiss, 1995; Tassev and Zaldarriaga, 2012), with the non-linear dynamics yielding a strong coupling between different modes $\delta(\boldsymbol{k})$ in Fourier space, such that a simple analytic approach is not feasible. In particular, the heterogeneous structure formation violates the above relation of $\delta(\boldsymbol{x}, a) = D^+(a)\delta(\boldsymbol{x})$.

To overcome such limitations and account for non-linear physical evolution, we can perform full N-body simulations using a numerical particle mesh model (e.g. Klypin and Shandarin, 1983; Efstathiou et al., 1985; Hockney and Eastwood, 1988; Klypin and Holtzman, 1997). The purpose of the particle mesh code is to solve the gravitational N-body problem by following the dynamical trajectories of a cloud of dark matter particles in a given cosmological volume, while accounting for their mutual gravitational interactions. We outline the general principles involved in this approach. The particle mesh code solves the following equations of motion, from some predefined initial conditions, for positions $\boldsymbol{x}$ and momenta $\boldsymbol{p}$ of dark matter particles, to the present state of large-scale structures (Jasche and Lavaux, 2019):

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}a} = \frac{\boldsymbol{p}}{\dot{a}a^2} \tag{6.13}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}a} = -\frac{\boldsymbol{\nabla}_x \Phi}{aH(a)}, \tag{6.14}$$

with $\dot{a}$ being its the first time derivative of the cosmic scale factor, and $\boldsymbol{p} = a^2\dot{\boldsymbol{x}}$. The gravitational potential is encoded in the Poisson equation:

$$\nabla_x^2 \Phi = \frac{3}{2} H_0^2 \Omega_{\mathrm{m}} \frac{\delta_{\mathrm{m}}(\boldsymbol{x})}{a} = \frac{H_0}{a} \nabla_x^2 \tilde{\Phi}, \tag{6.15}$$

where $\tilde{\Phi}$ is the reduced gravitational potential, such that the above Poisson equation relating the density of particles to the potential $\tilde{\Phi}$ can be expressed as:

$$\nabla_x^2 \tilde{\Phi} = \frac{3}{2} H_0 \Omega_{\mathrm{m}} \delta_{\mathrm{m}}(\boldsymbol{x}). \tag{6.16}$$

The densities may subsequently be estimated from the particle positions via the cloud-in-cell (CIC) scheme (e.g. Hockney and Eastwood, 1988).

GADGET2 (Springel, 2005) is a popular tool for performing such numerical simulations to study the non-linear stages of cosmic structure formation, with recent extensions also including baryonic dynamics, magnetic fields and cosmic rays. In Chapter 12, we construct an efficient emulator of complex dynamics which predicts the 3D halo distribution using only the 2LPT density field, obviating the need for full N-body simulations and halo finding algorithms. As described in the next section, dark matter halos are the key building components of the cosmic large-scale structures.

6.2 Morphology of the cosmic web

In this section, we describe the cosmic matter distribution, followed by a description of the large-scale structures of the Universe, in terms of the relevant statistical tools.

6.2.1 Distribution of matter in the Universe

Structure formation in the Universe, as described by the equations governing the dynamics involved in the previous section, do not explain the origin of the initial seed perturbations. The commonly accepted paradigm is that all the observed cosmic structures were formed via gravitational instability from primordial microscopic quantum fluctuations generated in the earliest moments of the Universe, during the so-called *Planck* epoch. While the origin and statistical properties of the initial density perturbations are not completely understood, inflationary models currently provide the most promising explanations (e.g. Guth, 1981; Albrecht and Steinhardt, 1982; Guth and Pi, 1982), as outlined in Chapter 1.2.

A common prediction of the various inflationary scenarios proposed is that the primordial density fluctuations can be characterized by a Gaussian distribution, as a consequence of the central limit theorem (Kendall and Stuart, 1968) for the superposition of an extremely large number of uncorrelated quantum fluctuations. This, in turn, implies that the generation of the initial density perturbations is a stochastic process and is beyond a description based purely from first principles. At the fundamental level, a probability distribution function of these fluctuations is prerequisite to characterize the cosmic large-scale structures. To infer the statistical properties of the density field, conventional analyses rely on observations of the CMB anisotropies (e.g. Hinshaw et al., 2007; Planck Collaboration et al., 2014a) or galaxy distribution from redshift surveys (e.g. York et al., 2000; Alam et al., 2015). The standard practice in cosmology is to employ summary statistics as an adequate representation of the cosmological density fields, as outlined in the next section.

The gravitational amplification of the initial seed perturbations leads to the hierarchical formation of the observed structures in the Universe, such as galaxies and clusters of galaxies, in a complex network known as the *cosmic web*, consisting of filaments, walls and voids (Klypin and Shandarin, 1983; Padmanabhan, 1993; Bond, Kofman, and Pogosyan, 1996). The backbone of the cosmic web is the filamentary distribution of dark matter, resulting in a gravitational scaffold (Massey et al., 2007), which subsequently yields potential wells into which baryons fall, leading to the formation of luminous galaxies (Mukhanov, 2005).

6.2.2 Statistical description of cosmological large-scale structures

According to our standard cosmological model, the primordial density fluctuations can be characterized as a Gaussian random field, as mentioned in the previous section, with observations of the CMB anisotropies (e.g. Hinshaw et al., 2007; Planck Collaboration et al., 2014a) providing substantial evidence for this paradigm. An in-depth mathematical description of Gaussian random fields is provided in Leclercq, Pisani, and Wandelt (2014).

The statistical properties of a cosmic Gaussian field are completely described by the two-point correlation function:

$$\xi(\boldsymbol{r}) \equiv \langle \delta(\boldsymbol{r}')\delta(\boldsymbol{r}'+\boldsymbol{r})\rangle, \tag{6.17}$$

defined in configuration space as the joint ensemble average, taken over all possible realizations of the Universe, of the field at two different locations. It depends only on the norm of $\boldsymbol{r}$, i.e. $\xi(r) = \xi(|\boldsymbol{r}|)$, for a homogeneous and isotropic Universe, leaving only the amplitude $|\boldsymbol{r}|$ in the above definition.

The scalar field $\delta(\boldsymbol{r})$ corresponds to the density contrast in the context of large-scale structures, and can be expressed in terms of its Fourier components,[1]

$$\delta(\boldsymbol{r}) = \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\, \hat{\delta}(\boldsymbol{k}) \exp\left(i\boldsymbol{k}\cdot\boldsymbol{r}\right), \tag{6.18}$$

or equivalently, via its Fourier transform,

$$\hat{\delta}(\boldsymbol{k}) = \int \mathrm{d}^3\boldsymbol{r}\, \delta(\boldsymbol{r}) \exp\left(-i\boldsymbol{k}\cdot\boldsymbol{r}\right), \tag{6.19}$$

where the 3D wavevector of the plane wave is denoted by $\boldsymbol{k}$, with its amplitude (wavenumber) being k, related to the wavelength λ via $k = 2\pi/\lambda$. Note that the quantities $\hat{\delta}(\boldsymbol{k})$ are complex random variables. For a real scalar field, we have $\hat{\delta}(-\boldsymbol{k}) = \hat{\delta}^*(\boldsymbol{k})$, such that half of the Fourier space is sufficient to contain all the information.

Computing the two-point correlation function for $\hat{\delta}(\boldsymbol{k})$ in Fourier space yields the following relation:

$$\begin{aligned}
\langle \hat{\delta}^*(\boldsymbol{k})\hat{\delta}(\boldsymbol{k}')\rangle &= \int \mathrm{d}^3\boldsymbol{r}' \int \mathrm{d}^3\boldsymbol{r}\, \langle \delta^*(\boldsymbol{r}')\delta(\boldsymbol{r}'+\boldsymbol{r})\rangle \exp\left[i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}' - i\boldsymbol{k}'\cdot\boldsymbol{r}\right] \\
&= \int \mathrm{d}^3\boldsymbol{r}' \int \mathrm{d}^3\boldsymbol{r}\, \xi(\boldsymbol{r}) \exp\left[i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}' - i\boldsymbol{k}'\cdot\boldsymbol{r}\right] \\
&= (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}') \int \mathrm{d}^3\boldsymbol{r}\, \xi(\boldsymbol{r}) \exp\left(i\boldsymbol{k}\cdot\boldsymbol{r}\right) \\
&= (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}')P(k),
\end{aligned} \tag{6.20}$$

where δ_{D} is the Dirac delta function, and the above result leads to the definition of the *power spectrum*:

$$P(k) = \int \mathrm{d}^3\boldsymbol{r}\, \xi(\boldsymbol{r}) \exp\left(i\boldsymbol{k}\cdot\boldsymbol{r}\right), \tag{6.21}$$

[1] The hat denotes Fourier space representation.

as the Fourier transform of the two-point correlation function $\xi(r)$. The power spectrum essentially describes the dispersion of the plane wave amplitudes at different scales in Fourier space. Nevertheless, it does not provide any information about the phases of the field.

Non-linear structure formation, as described in Section 6.1.2, yields non-Gaussian features via mode coupling as well as phase correlations, such that the statistical properties of the evolved matter fields cannot be uniquely described in terms of the above two-point correlation function or power spectrum. To characterize this non-Gaussian field, higher-order statistics are required. The above definition (6.17) can be generalized for higher-order correlation functions, such that the n^{th} order correlation function in Fourier space can be expressed as

$$\langle \hat{\delta}(\boldsymbol{k}_1)\hat{\delta}(\boldsymbol{k}_2)\dots\hat{\delta}(\boldsymbol{k}_n)\rangle \equiv (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}_1+\boldsymbol{k}_2+\dots+\boldsymbol{k}_n)P_n(\boldsymbol{k}_1,\boldsymbol{k}_2,\dots,\boldsymbol{k}_n). \tag{6.22}$$

The Dirac delta distribution δ_{D} enforces statistical homogeneity of the field by ensuring that $\boldsymbol{k}$-vector configurations form closed polygons, $\sum_i \boldsymbol{k}_i = \boldsymbol{0}$, such that the above polyspectrum depends solely on $(n-1)$ wavevectors in the n-point function. Moreover, for a statistically isotropic field, there is no dependence on the direction of the $\boldsymbol{k}$-vectors. Going beyond the power spectrum, the immediate statistic of interest is the bispectrum $B(\boldsymbol{k})$, corresponding to $n = 3$, defined as

$$\langle \hat{\delta}(\boldsymbol{k}_1)\hat{\delta}(\boldsymbol{k}_2)\hat{\delta}(\boldsymbol{k}_3)\rangle \equiv (2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)B(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3), \tag{6.23}$$

which allows us to characterize the spatial distribution of the field. In Chapter 12, we make use of the power spectrum and bispectrum as quantitative diagnostics of the performance of our neural network in predicting 3D halo distributions.

Chapter 7

Bayesian inference in cosmology

7.1 Bayesian probability theory

In this chapter, we present the rationale underlying the use of Bayesian inference techniques for cosmological data analysis. We describe the conceptual underpinnings of Bayesian probability theory, in contrast to the conventional frequentist approach. In particular, we discuss the parameter inference problem within the Bayesian framework and the role of the prior in Bayesian data analysis, as relevant to the ALTAIR and BORG algorithms in Chapters 8 and 9.

7.1.1 Frequentist vs. Bayesian statistics

To push the boundaries of our understanding of our physical Universe, we conduct cosmological observations to identify or investigate unknown cosmic phenomena or verify the validity of existing theories in the light of the observations. Therefore, we need precise and quantitative tools for data analysis to make progress by optimally extracting information from observations, while adequately quantifying the associated uncertainties, in order to establish a physical model of our Universe.

There are two distinct viewpoints pertaining to statistical data analysis, namely the conventional (*frequentist*) and Bayesian approaches (e.g. Kendall and Stuart, 1968). Although both schools of thought have their foundations in probability theory, their interpretation of probability when using probability distributions, and their respective approaches for the determination of characteristic values, testing of hypotheses and model comparison differ crucially (e.g. Trotta, 2008; Heavens, 2009). In essence, a Bayesian method yields more general and profound statements about measurements or observations, as argued in Jasche et al. (2010b). For instance, a frequentist approach seeks the answer to: "Given the true value s of a signal, what is the probability distribution of the measured values d?" In contrast, a Bayesian method answers a distinct question: "Given the observations d, what is the probability distribution of the true signal s?" As such, by inferring the true value of the signal from observations, the Bayesian framework provides the answer to the fundamental question inherent to any measurement problem, unlike conventional statistics. The Bayesian method, in a nutshell, provides an algorithm for knowledge acquisition. The probability distribution of s quantifies the information content of *our* knowledge of s. Another distinction is that the frequentist considers the model as carved in stone, but this is not correct in general. The validity of the model is limited by the knowledge that we have accrued. The Bayesian evidence, as outlined in Section 7.2, provides a straightforward path to assess models.

7.1.2 Definition of probability

First and foremost, the epistemological definition and interpretation of probability is inherently different from both viewpoints. For instance, we may consider the probability $\mathcal{P}(A)$ of an event A occurring. The frequentist definition of the probability $\mathcal{P}(A)$ is the relative frequency of occurrence of event A in repeated experiments, in the limit of infinite equiprobable repetitions (Heavens, 2009). The various shortcomings of this frequentist definition are clearly outlined in Trotta (2008). In contrast, from the Bayesian viewpoint, the probability $\mathcal{P}(A)$ represents the degree of belief associated with the occurrence of event A, under consideration of all available information (Heavens, 2009). This intuitive concept of probability represents a state of knowledge in the presence of incomplete information (Jaynes, 2003).

7.1.3 Bayes identity

The basis of Bayesian statistics lies in a set of fundamental consistency requirements and philosophical principles for plausible reasoning, known as Cox's axioms (Cox, 1946), and can be derived from one of the axioms of probability theory, involving conditional probabilities (e.g. Trotta, 2008; Trotta, 2017):

$$\mathcal{P}(y|x)\mathcal{P}(x) = \mathcal{P}(x|y)\mathcal{P}(y) = \mathcal{P}(x,y), \tag{7.1}$$

where the vertical bar represents conditionality, such that $\mathcal{P}(x|y)$ implies the probability of x given y and $\mathcal{P}(x,y)$ denotes the joint probability distribution of x and y. The above identity leads to the mathematical formulation for inverse problems, i.e. if we know how x arises from y, then we can use x to constrain y, providing a pathway from forward modelling to the inverse problem (Leclercq, Pisani, and Wandelt, 2014).

7.2 Bayesian data analysis

The Bayes identity, from Eq. (7.1), is usually written as follows:

$$\mathcal{P}(\theta|d) = \frac{\mathcal{P}(d|\theta)\mathcal{P}(\theta)}{\mathcal{P}(d)}, \tag{7.2}$$

where θ represents a given set of model parameters corresponding to a particular theory, and d is the data set, with both θ and d written as vectors. This constitutes the mathematical foundation of Bayesian analysis (e.g. Trotta, 2008; Heavens, 2009; Verde, 2010):

- $\mathcal{P}(d|\theta)$ is the *likelihood* and it corresponds to the probability of the data before it is known, given the theory;
- $\mathcal{P}(\theta)$ is the *prior* probability distribution which is the probability of the theory in the absence of data;
- $\mathcal{P}(\theta|d)$ is the *posterior* probability distribution, indicating the probability of the theory after the data is known;

- $\mathcal{P}(d)$ is known as the *evidence* and it is the probability of the data before it is known, without making any assumptions about the theory.

In a Bayesian analysis, any uncertain quantities have an associated probability distribution function, such that the key outcome is the *posterior density* which provides a quantitative measure of the relative degree of rational belief in a range of parameter values, given the combined information from the prior and the data. This is a stark contrast to a frequentist analysis, which yields an estimator for the parameters of interest (Leclercq, Pisani, and Wandelt, 2014). The choice of the prior distribution is discussed in Section 7.2.2.

7.2.1 The parameter inference problem

Bayesian data analysis typically takes the form of parameter inference and model comparison. The former is of interest to Chapters 8 and 9, and is hence outlined here. In essence, this allows the interpretation of observations in terms of a model, which is a theoretical framework incorporating some assumptions, characterized by some free parameters (Heavens, 2009).

Within the Bayesian framework, the parameter inference problem can be expressed as (Verde, 2010): "What is the probability distribution of the model parameters given the available data?" This can formulated, in a general context, as follows: Given a vector of parameters θ associated to a physical model $\mathcal{M}$, we should specify priors $\mathcal{P}(\theta|\mathcal{M})$ for each parameter, followed by the construction of an appropriate likelihood function $\mathcal{P}(d|\theta, \mathcal{M})$. The likelihood encodes a generative and probabilistic model of the data, i.e. it reflects the data acquisition procedure. For instance, a measurement with Gaussian noise is adequately modelled by a normal distribution (Leclercq, Pisani, and Wandelt, 2014).

We subsequently obtain the posterior distribution for the model parameters of interest by applying the Bayes identity (cf. Eq. (7.2)), yielding

$$\mathcal{P}(\theta|d, \mathcal{M}) \propto \mathcal{P}(d|\theta, \mathcal{M})\mathcal{P}(\theta|\mathcal{M}), \tag{7.3}$$

after specifying the prior and encoding the information from the data in the likelihood function (Trotta, 2008). The Bayesian evidence $\mathcal{P}(d|\mathcal{M})$ is simply a normalizing constant for parameter inference and is therefore omitted above (Heavens, 2009). It is, however, crucial for model comparison. In practice, θ can correspond to a set of parameters, with some physically interesting quantities ζ and some nuisance parameters ψ. The above posterior is therefore the *joint* posterior distribution for $\theta = \{\zeta, \psi\}$. To obtain the *marginal* posterior of interest, we must *marginalize* over the nuisance parameters, as follows (Leclercq, Pisani, and Wandelt, 2014):

$$\mathcal{P}(\zeta|d, \mathcal{M}) \propto \int \mathrm{d}\psi \, \mathcal{P}(d|\zeta, \psi, \mathcal{M})\mathcal{P}(\zeta, \psi|\mathcal{M}). \tag{7.4}$$

The key outcome of the above parameter inference problem is the posterior probability density function, $\mathcal{P}(\theta|d, \mathcal{M})$. To extract information from the latter, we usually make use of Markov Chain Monte Carlo methods for efficient exploration of the posterior in high-dimensional parameter spaces, as explained in Section 7.3.

7.2.2 Choice of priors

The choice of a suitable prior is a key component of any Bayesian analysis. Due to the absence of a unique prescription for selecting a prior, it has attracted considerable criticism in the past, as summarized by Gelman (2008). The choice of priors is nevertheless not a limitation of the Bayesian approach, nor does it undermine the objectivity of an analysis. Since there can be no inference without assumptions, the prior, within Bayesian probability theory, is a systematic way of quantifying one's assumptions and *a priori* state of knowledge about the specific problem before actually looking at the data. In general, if the likelihood is more informative than the prior, the posterior converges to a common function, irrespective of choice of prior, guaranteeing objectivity. Moreover, specifying priors ensures transparency, as unreasonable or unjustified assumptions are exposed to scientific criticism. In comparison, assumptions adopted in a frequentist analysis may not be so evident. A more in-depth discussion about priors is provided in Leclercq, Pisani, and Wandelt (2014) and Trotta (2017).

7.3 Markov Chain Monte Carlo methods

This section contains an introduction to Markov Chain Monte Carlo (hereafter MCMC) methods, as employed in Chapters 8 and 9. We outline the elementary notions underlying MCMC for the exploration of high-dimensional posterior distributions and briefly review some standard MCMC methods which have proven to be extremely efficient.

7.3.1 The Monte Carlo rationale

The posterior distribution, as the primary result of the Bayesian analysis, encodes all the relevant information about the physical system under investigation. Any statistical inference, however, requires appropriate statistical summaries, such as the mean, median or mode, drawn from the posterior (Leclercq, Pisani, and Wandelt, 2014). Unfortunately, the representation of such a distribution, especially in high dimensions, is numerically challenging in practice (Heavens, 2009). Due to computational limitations, direct numerical integration of Eq. (7.4) is not feasible.

A possible solution is to map the posterior density numerically via a *sampled* representation in the high-dimensional space using Monte Carlo simulations (Trotta, 2008). The underlying rationale is to approximate the posterior by a set of samples drawn from the true posterior distribution:

$$\mathcal{P}(\theta|d) \approx \mathcal{P}_N(\theta|d) = \frac{1}{N}\sum_{i=1}^{N}\delta_D(\theta-\theta_i), \tag{7.5}$$

where a sum of N Dirac delta distributions provides an estimate to the desired posterior distribution. By design, the posterior probability, at any given point, is proportional to the local density of samples in the parameter space (Leclercq, Pisani, and Wandelt, 2014). Using Eq. (7.5), we can compute any integrals to obtain any desired statistical summary, such as the mean and variance (Heavens, 2009). For instance, the expectation value of a given function of the parameters $f(\theta)$ can

be approximated by the following tractable sum:

$$\langle f(\theta) \rangle = \int \mathrm{d}\theta \, f(\theta) \mathcal{P}(\theta) \approx \frac{1}{N} \sum_{i=1}^{N} f(\theta_i). \tag{7.6}$$

An advantage of the above approach is that marginalization consequently becomes trivial as we only have to histogram the number of samples falling within different bins of a given subset of parameters, while disregarding the remaining parameters, obviating the need for any high-dimensional integration (Trotta, 2017). Moreover, uncertainty quantification becomes conceptually straightforward as the variation between the samples provides a quantitative measure of the uncertainty. This is especially convenient for a Bayesian approach, as there is no inherent distinction between *statistical* and *systematic* uncertainties. The former is due to the stochastic nature of the experiment, while the latter emanates from the deterministic effects that are not perfectly known (Leclercq, Pisani, and Wandelt, 2014).

7.3.2 Markov chains for posterior exploration

In Chapters 8 and 9, we investigate physical systems with $\mathcal{O}(10^7)$ free parameters, such that direct evaluation of the posterior is never a possibility and we have to rely on the numerical approximation of sampling. The success of Monte Carlo methods, nevertheless, relies on the ability to efficiently draw samples from the target distribution $\mathcal{P}(\theta)$. One method of sampling is the sophisticated Markov Chain Monte Carlo (MCMC) technique. This allows us to map out numerically the posterior by constructing a sequence or "chain" of points in parameter space, whose density is proportional to the target posterior density (Trotta, 2008).

By definition, a Markov chain is a sequence of random elements $\{\theta_1, \theta_2, \ldots, \theta_n, \ldots\}$ of some "state space" χ, where the conditional distribution of θ_{n+1}, given all the previous n elements, depends solely on θ_n, i.e. the immediate preceding element (Trotta, 2017). It can be shown that Markov chains converge (asymptotically) to a stationary state, i.e. there is no dependence on n, where the successive elements in the chain are randomly drawn samples corresponding to the target distribution $\mathcal{P}(\theta|d)$ (Trotta, 2008; Heavens, 2009), such that

$$\mathcal{P}^{(n)}(\theta) \to \mathcal{P}(\theta) \text{ as } n \to \infty, \text{ for any } \theta_0. \tag{7.7}$$

The above crucial property of *stationary transition probability* constitutes the basis of the MCMC technique (Leclercq, Pisani, and Wandelt, 2014). The state of the chain, after some initial steps in the parameter space, depending on the initial position θ_0, before reaching the high-density regions of the target distribution is known as the "burn-in" phase. Once the chain has converged, yielding a sufficiently large number of samples, we can compute summary statistics of the target posterior distribution as explained in the previous section (cf. Eq. (7.6)). There are several well-known techniques for constructing Markov chains, with the choice of algorithm depending on the problem being investigated. In the following section, we review two such algorithms, as relevant to the future chapters of this thesis.

7.3.3 Metropolis-Hastings algorithm

A simple and widely used MCMC method is the *Metropolis-Hastings* algorithm (Metropolis et al., 1953; Hastings, 1970). To construct the Markov chain, the basic procedure is to initially choose an arbitrary starting point θ_0 in the parameter space as the first sample, then propose a new point θ' from the current point θ via a specified *proposal distribution* $\mathcal{Q}(\theta'|\theta)$, for a jump from θ to θ'. The new point is accepted with a probability which depends on the ratio of the final and initial target densities (Heavens, 2009), computed at each step using

$$\mathcal{P}(\text{acceptance}) = \min\left[1, \frac{\mathcal{P}(\theta')\mathcal{Q}(\theta'|\theta)}{\mathcal{P}(\theta)\mathcal{Q}(\theta|\theta')}\right]. \tag{7.8}$$

If accepted, the proposed point becomes the new state in the chain, otherwise the chain stays at θ (Leclercq, Pisani, and Wandelt, 2014), with the above procedure repeating in iterative fashion. Nevertheless, this method is sub-optimal for high-dimensional problems, with the proposal distribution dictating the step size being a key component that requires some fine tuning for efficient exploration of the parameter space (Trotta, 2008).

7.3.4 Gibbs sampling framework

The *Gibbs* sampling framework, as proposed by Geman and Geman (1987) and Gelfand and Smith (1990), is particularly suited to problems where the joint posterior distribution is not known explicitly, but it is straightforward and computationally cheap to sample from the conditional distribution of one of the free parameters at a given time (Trotta, 2017). This has been shown to work well for cosmological inference problems with $\mathcal{O}(10^7)$ parameters, such as the statistical reconstruction of large-scale structures from galaxy surveys (e.g. Jasche et al., 2010b).

Gibbs sampling entails updating each of the parameters in turn by drawing the proposal distribution of the given variable, conditional on the remaining ones (Smith and Roberts, 1993). For a 2D parameter space, where $\theta = \{x, y\}$ and d is the relevant data, we obtain the n^{th} sample by drawing:

$$x^{(n)} \curvearrowleft \mathcal{P}\left(x|y = y^{(n-1)}, d\right) \tag{7.9}$$

$$y^{(n)} \curvearrowleft \mathcal{P}\left(y|x = x^{(n)}, d\right), \tag{7.10}$$

where the symbol $\curvearrowleft$ denotes sampling from the distribution on the right. Here, the target distribution is the joint posterior distribution $\mathcal{P}(x, y|d)$. In the second step above, the draw for y is conditional on the value of x that has been updated to the value corresponding to the latest draw of x, i.e. $x^{(n)}$ (Trotta, 2017).

The mathematical formalism justifying the above being equivalent to sampling the joint posterior, under certain conditions, is provided in Casella and George (1992). Gibbs sampling can therefore be considered as a special case of Metropolis-Hastings sampling, with one-dimensional proposal distributions and an acceptance rate of unity (Trotta, 2017), which is the major advantage of this technique. This can also be generalized to blocks of variables, known as *block Gibbs* sampling, where all the variables are updated conditional on the others.

In Chapters 8 and 9, we employ an efficient method for sampling extremely high-dimensional parameter spaces, known as the *Hamiltonian Monte Carlo* (HMC) technique, within a block Gibbs sampling framework, as described in Section 8.4.3. Section 8.5 outlines the conceptual framework of the HMC sampling method.

Chapter 8

Precision cosmology with cosmic expansion

The work presented in this chapter is based on Kodi Ramanah et al. (2019).

8.1 Introduction

The past few decades have witnessed the advent of an array of galaxy redshift surveys, with the state-of-the-art catalogues mapping millions of galaxies with precision positioning and accurate redshifts. The Sloan Digital Sky Survey (SDSS) (York et al., 2000; Abazajian et al., 2009; Ahn et al., 2014; Alam et al., 2015) and the Six Degree Field Galaxy Redshift Survey (6dFGRS) (Jones et al., 2009) are two notable examples. Future cutting-edge surveys from the Euclid (Laureijs et al., 2011; Racca et al., 2016; Amendola et al., 2016) and Large Synoptic Survey Telescope (LSST) (Ivezic et al., 2008) missions, currently under construction, further highlight the wealth of galaxy redshift data sets which would be available within a five to ten year time frame. Sophisticated and optimal data analysis techniques, in particular large-scale structure analysis methods, are in increasing demand to cope with the present and upcoming avalanches of cosmological and astrophysical data, and therefore optimize the scientific returns of the missions.

With the metamorphosis of cosmology into a precision (and data-driven) science, the three-dimensional (3D) large-scale structures have emerged as an essential probe of the dynamics of structure formation and evolution to further our understanding of the Universe. The two-point statistics of the 3D matter distribution have developed into key tools to investigate various cosmological models and test different inflationary scenarios. Various techniques to measure the power spectrum and several reconstruction methods attempting to recover the underlying density field from galaxy observations are described in literature (e.g. Bertschinger and Dekel, 1989; Bertschinger and Dekel, 1991; Hoffman, 1994; Lahav et al., 1994; Fisher et al., 1995; Sheth, 1995; Webster, Lahav, and Fisher, 1997; Bistolas and Hoffman, 1998; Schmoldt et al., 1999; Saunders and Ballinger, 2000; Zaroubi, Hoffman, and Dekel, 1999; Zaroubi, 2002; Erdoğdu et al., 2004; Erdoğdu et al., 2006), with the recent focus being on large-scale Bayesian inference methods (e.g. Kitaura and Enßlin, 2008; Kitaura et al., 2009; Jasche and Kitaura, 2010; Jasche et al., 2010b; Jasche and Wandelt, 2012; Jasche and Wandelt, 2013b; Jasche and Lavaux, 2015; Jasche and Lavaux, 2019). A formal and rigorous Bayesian framework provides the ideal setting to solve the ill-posed problem of inferring signals from noisy observations, while quantifying the corresponding statistical uncertainties.

The potential of such Bayesian algorithms to jointly infer cosmological constraints, nevertheless, has not yet been exploited. We present, for the first time, a non-linear Bayesian inference framework for cosmological parameter inference from galaxy redshift surveys via an implementation of the Alcock-Paczyński (AP, Alcock and Paczynski, 1979) test. We extend the hierarchical Bayesian inference machinery of BORG (Bayesian Origin Reconstruction from Galaxies) (Jasche and Wandelt, 2013a), originally developed for the non-linear reconstruction of large-scale structures, to constrain cosmological parameters. BORG encodes a physical model for gravitational structure formation, yielding a highly non-trivial Bayesian inverse problem. This consequently allows us to reformulate the standard problem of present 3D density field reconstruction as an inference problem for initial conditions at an earlier epoch from current galaxy observations. BORG builds upon the implementation of the Hamiltonian Monte Carlo (HMC) method (Neal, 1993), initially introduced in the HADES (HAmiltonian Density Estimation and Sampling) algorithm (Jasche and Kitaura, 2010), for efficiently sampling the high dimensional and non-linear parameter space of possible initial conditions at an earlier epoch.

In this work, the conceptual framework is to constrain the comoving-redshift coordinate transformation and therefore infer the appropriate cosmology which would result in isotropic correlations of the galaxy density field. The key aspect of this application of the AP test consequently lies in its robustness to a misspecified model and the approximations therein, yielding a near-optimal exploitation of the model predictions, without relying on its accuracy in modelling the scale dependence of the correlations of the density field. Here, we employ Lagrangian Perturbation Theory (LPT) as a physical description for the non-linear dynamics and perform a joint inference of initial conditions, and consequently the corresponding non-linearly evolved density fields and associated velocity fields, and cosmological parameters, from incomplete observations. This augmented framework with cosmological applications is designated as ALTAIR (ALcock-Paczyński consTrAIned Reconstruction).

This chapter is organized as follows. In Section 8.2, the underlying principles of the AP test are outlined, followed by a description of the forward modelling approach and data model implemented in Section 8.3. In Section 8.4, we describe the LPT-Poissonian posterior implemented in this work. We provide a brief overview of the Hamiltonian sampling approach in Section 8.5, and follow up by deriving the required equations of motion in Section 8.5.2, with the numerical implementation outlined in Section 8.6. We subsequently describe how we increase the efficiency of our cosmological parameter sampler via a rotation of the parameter space in Section 8.6.3. We then test the algorithm in Section 8.8 on an artificially generated galaxy survey, with the mock generation procedure described in the preceding Section 8.7, by investigating its performance via statistical efficiency and consistency tests. In Section 8.9, we summarize the main aspects of our work and discuss further possible extensions to our algorithm in order to fully exploit its potential in deriving cosmological constraints.

8.2 The Alcock-Paczyński test

The Alcock-Paczyński (AP) test (Alcock and Paczynski, 1979) is a cosmological test of the expansion of the Universe and its geometry. The main advantage of this test is that it is independent of

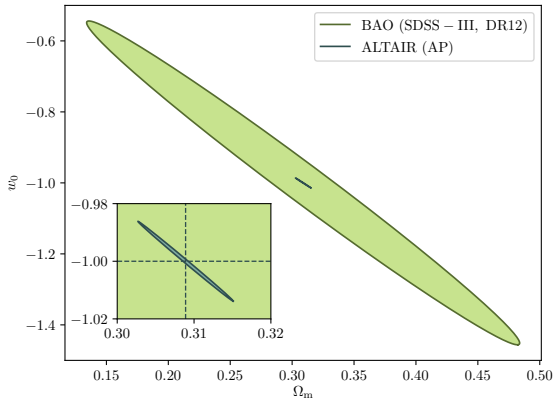


FIGURE 8.1: Comparison of cosmological constraints from BAO measurements and our implementation of AP test in ALTAIR. The grey and green lines denote the 1σ confidence region, centred on the fiducial cosmological parameters, obtained from our AP test and BAO constraints from SDSS-III (Data Release 12) (Alam et al., 2017), respectively. The BAO constraints have not been combined with *Planck* CMB measurements. This demonstrates the potentially unprecedented constraining power of our AP test compared to standard BAO analyses, as discussed in Section 8.2, with the inset focusing on the ALTAIR constraints where the fiducial cosmology is depicted in dashed lines. This error forecast is validated on a simulated analysis (cf. Fig. 8.8).

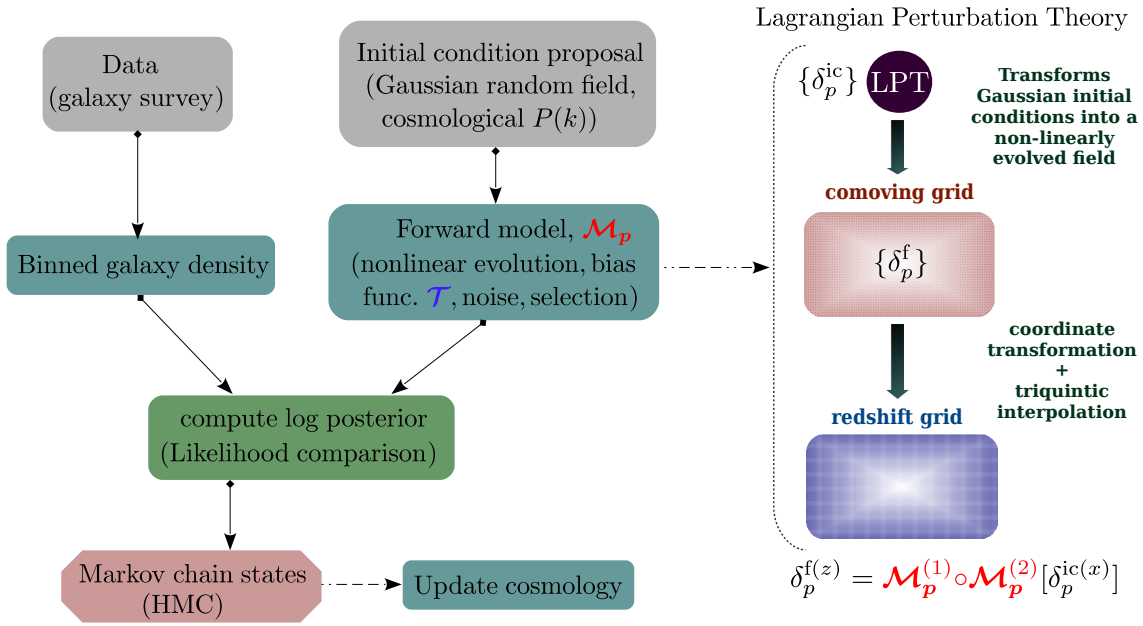


FIGURE 8.2: Schematic representation of the reconstruction pipeline. The forward model consists of a chain of various components for the non-linear evolution from initial conditions and the subsequent transformation from comoving to redshift space for the application of the AP test. This consequently transforms the initial density field into a set of predicted observables, i.e. a galaxy distribution in redshift space, for comparison with data via a likelihood or posterior analysis.

the evolution of galaxies but depends only on the geometry of the Universe. The assumption of incorrect cosmological parameters in data analysis produces distortions in the appearance of any spherical object or isotropic statistical distribution. The AP test provides a pathway to exploit this resulting spurious anisotropy to constrain the cosmological parameters. Here, we invoke the AP test to ensure that the underlying geometrical properties of isotropy of the Universe (Friedmann, 1922; Friedmann, 1924; Lemaître, 1927; Lemaître, 1931; Lemaître, 1933; Robertson, 1935; Robertson, 1936a; Robertson, 1936b; Walker, 1937; Saadeh et al., 2016) are maintained. As such, the key underlying assumption adopted in this work relies purely on the geometrical symmetries of the cosmological principle. As a result, such a test does not employ the growth of structures to constrain cosmology, unlike cluster abundance (e.g. Wang and Steinhardt, 1998).

The AP test, and various formulations thereof, have been studied extensively in the context of galaxy and quasar surveys (e.g. Phillipps, 1994; Ryden, 1995; Ballinger, Peacock, and Heavens, 1996; Matsubara and Suto, 1996; Popowski et al., 1998; de Laix and Starkman, 1998; López-Corredoira, 2014). Variants of the AP test have also been successfully applied to cosmic voids (e.g. Sutter et al., 2012; Lavaux and Wandelt, 2012; Sutter et al., 2014; Hamaus et al., 2014; Hamaus et al., 2015; Hamaus et al., 2016) and also to other cosmological observables like supernovæ (Blake et al., 2011), the Lyman-α forest (Hui, Stebbins, and Burles, 1999) and 21 centimetre emission maps (Nusser, 2005; Barkana, 2006).

With baryon acoustic oscillations (BAOs) being a robust standard ruler, the AP test has been utilized for the simultaneous measurement of the Hubble parameter and angular diameter distance of distant galaxies (e.g. Seo and Eisenstein, 2003; Blake and Glazebrook, 2003; Glazebrook and Blake, 2005; Padmanabhan and White, 2008; Shoji, Jeong, and Komatsu, 2009). In Fig. 8.1, we depict the 1σ confidence region of the cosmological constraints inferred via our implementation of the AP test. As a comparison, we also indicate the corresponding confidence region obtained via BAO measurements from the SDSS-III (Data Release 12) (Alam et al., 2017). These BAO constraints have not been combined with *Planck* measurements, which would significantly tighten the constraints. Nevertheless, this highlights the significant potential constraining power of our AP test compared to standard BAO analyses, while being at least as robust. While this improvement is extremely substantial for the mock SDSS-III survey considered here, we will investigate to what extent the above promise holds when applied to actual SDSS-III data in a follow-up work, as unknown systematics represent a potential caveat. We discuss Fig. 8.1 in more depth in Section 8.8.

The crucial aspect of our AP test is that it does not assume that the correlation function is correctly modelled. This robustness to a misspecified model is illustrated explicitly in Section 8.8, where we demonstrate that the shape of the prior power spectrum adopted in the inference framework does not impact on the inferred cosmological constraints (cf. Fig. 8.12). As a result of this robustness, our AP test has a definite edge over standard approaches. Moreover, it has been pointed out that other cosmological tests, such as the luminosity distance - redshift relation, can be considered as generalized formulations of the AP test (Mukherjee and Wandelt, 2018), further underlining the strength of the approach presented in this work.

8.3 The forward modelling approach

The large-scale structure (LSS) posterior implemented in this work, based on the BORG framework (Jasche and Wandelt, 2013a), is described in depth in Section 8.4. A key component of the inference framework is the forward model $\mathcal{M}_p$ which links the initial conditions $\delta_p^{\mathrm{ic},(r)}$ to the redshift space representation of the evolved density field $\delta_p^{\mathrm{f},(z)}$ as follows:

$$\begin{aligned}\delta_p^{\mathrm{f}(z)} &= \mathcal{M}_p\left(1+\delta_p^{\mathrm{ic}(r)}\right) = \mathcal{M}_p^{(1)} \circ \mathcal{M}_p^{(2)}\left(1+\delta_p^{\mathrm{ic}(r)}\right) = \mathcal{M}_p^{(1)}\left(\rho_p^{\mathrm{f}(r)}\right) \\ &= \mathcal{J}_p\left(\sum_{i,j} \mathcal{E}_{ij}^{-1} \rho_j^{\mathrm{f}(r)} x_p^{\alpha(i)} y_p^{\beta(i)} z_p^{\gamma(i)}\right) - 1, \end{aligned} \tag{8.1}$$

where $\rho_p^{\mathrm{f}(r)} \equiv 1+\delta_p^{\mathrm{f}(r)}$ is the final density field in comoving space. The forward model consists of two components, $\mathcal{M}_p = \mathcal{M}_p^{(1)} \circ \mathcal{M}_p^{(2)}$. The first component, $\mathcal{M}_p^{(2)} \equiv \mathcal{G}_p(a, \{\delta_p^{\mathrm{ic}}\})$, contains a physical description of the non-linear dynamics, and consequently propagates the initial conditions forward in time using LPT, yielding a non-linearly evolved final density field in comoving space, $\delta_p^{\mathrm{f}(r)}$.

To encode the AP test, we incorporate another component in the forward model that takes care of the coordinate transformation from comoving $(\boldsymbol{r})$ to redshift $(\boldsymbol{z})$ space, encoded in $\mathcal{M}_p^{(1)}$ (cf. Fig. 8.2). Schematically, we construct a second grid in redshift space, which involves a triquintic interpolation (fifth order interpolation scheme in three dimensions) on the comoving grid. This interpolation scheme is described in Section 8.5.3, with the notation employed in Eq. (8.1) clearly laid out. The corresponding Jacobian factor of this transformation, $|\mathcal{J}_r^z|$ (cf. Section 8.3.4), entails cosmological dependence and is consequently included in the AP test as well as through the direct coordinate dependence $\mathcal{E}_{ij}$.

The redshift space representation then allows for comparison with data via the likelihood or posterior. The essence of this AP test to constrain cosmological parameters can be summarized as follows: The Bayesian inference machinery explores the various cosmological expansion histories and selects the cosmology-dependent evolution pathways which result in isotropic correlations of the galaxy density field.

Fig. 8.2 illustrates the reconstruction scheme implemented in ALTAIR. First, galaxies are projected from the survey onto a 3D grid, such that we have a distribution of galaxies in redshift space and this constitutes our observable. We then generate a 3D density field according to Gaussian initial conditions (homogeneous prior) with a reference power spectrum, typically ΛCDM cosmology. The forward model subsequently transforms the initial density field into a set of predicted observables which are then compared to data via a likelihood or posterior analysis. And conversely, given the position of galaxies, we can infer this density field.

While we implement LPT to approximately describe gravitational non-linear structure formation in this work, other more adequate physical descriptions such as 2LPT or the non-perturbative particle mesh (see recent upgrade of BORG in Jasche and Lavaux, 2019) can be straightforwardly employed, within the flexible block sampling approach described in Section 8.4.3, by upgrading the

first component, $\mathcal{M}_p^{(2)} \equiv \mathcal{G}_p(a, \{\delta_p^{\rm ic}\})$, of our forward model (cf. Fig. 8.2). Nevertheless, our implementation of the AP test exploits essentially the isotropy of the correlation function, such that there is no explicit dependence on the accuracy of modelling the scale dependence of the correlations, rendering this method robust to a misspecified model and the approximations therein.

8.3.1 The galaxy data model

Galaxies can be considered as tracers of the matter fluctuations since they follow the gravitational potential of the underlying matter distribution, with the statistical uncertainty due to the discrete nature of the galaxy distribution usually modelled by a Poissonian distribution (e.g. Layzer, 1956a; Peebles, 1980). Poissonian likelihoods have emerged as the standard for non-linear LSS inference (e.g. Jasche et al., 2010a; Jasche and Kitaura, 2010; Kitaura, Jasche, and Metcalf, 2010). The Poissonian likelihood distribution implemented in ALTAIR, for multiple subcatalogues or galaxy observations labelled by the index g, can be expressed as follows:

$$\mathcal{L}\left(\{N_p^g\}|\lambda_p^g\right) = \prod_p \frac{(\lambda_p^g)^{N_p^g} \mathrm{e}^{-\lambda_p^g}}{N_p^g!}, \tag{8.2}$$

where N_p^g is the observed galaxy number counts in redshift space in the given voxel p. λ_p^g is the expected number of galaxies at this given position and is related to the final density field $\delta_p^{\rm f}$, in redshift space, via

$$\lambda_p^g\left(\{\delta_p^{\rm f}\}, \{\theta_i\}\right) = R_p^g(\theta_i)\mathcal{T}\left[1+\delta_p^{\rm f}\right], \tag{8.3}$$

where R_p^g is the overall linear response operator of the survey that incorporates the survey geometry and selection effects, and θ_i corresponds to a set of cosmological parameters. $\mathcal{T}$ is the galaxy biasing model which accounts for the fact that galaxies do not trace exactly the underlying matter distribution, and are therefore biased tracers with clustering properties that do not exactly mirror those of dark matter (Kaiser, 1984). This is currently one of the most challenging and unresolved issues hindering the analysis of galaxy distributions in non-linear regimes (e.g. see the review by Desjacques, Jeong, and Schmidt, 2016; Schmidt et al., 2018). In this work, we adopt the standard approach of a local, but non-linear bias function, in particular, the phenomenological model proposed by Neyrinck et al. (2014), such that the above Poisson intensity field can be expressed as

$$\lambda_p^g\left(\{\delta_p^{\rm f}\}, \{\theta_i\}, \{\bar{N}^g\}, \{b_i^g\}\right) = R_p^g \bar{N}^g \left[1+\delta_p^{\rm f}\right]^{\beta} \mathrm{e}^{-\rho^g\left[1+\delta_p^{\rm f}\right]^{-\epsilon^g}}, \tag{8.4}$$

where the bias function, described by four parameters, $\bar{N}^g$, the mean density of tracers, and $\{b_i^g\} = \{\beta, \rho^g, \epsilon^g\}$, is a truncated power-law bias model with the additional exponential function suppressing galaxy clustering in under dense regions. This bias model, with a power law and an exponential at low densities, were found to be in good agreement with standard excursion set and local-growth-factor models (for more details, see Neyrinck et al., 2014). The main limitation of this bias model is that it is purely local. Nevertheless, it is more adequate than a simplistic linear bias model and mitigates in practice the deficiencies of our physical model (LPT) at the considered resolution. The expected number of galaxies can subsequently be related to the initial conditions $\delta_p^{\rm ic}$ via the forward

model $\mathcal{M}_p$, as described above, due to the deterministic nature of structure formation, i.e. the Dirac delta function in Eq. (8.20).

The logarithm of the likelihood from Eq. (8.2) can therefore be expressed, in terms of the initial conditions, as

$$\begin{aligned}&\ln \mathcal{L}\left[\{N_p^g\}|\mathcal{M}_p\left(\{\delta_p^{\rm ic}\}\right),\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right]\\&=-\sum_p\left\{\lambda_p^g\left(\{\delta_p^{\rm f}\},\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right)-N_p^g\ln\left[\lambda_p^g\left(\{\delta_p^{\rm f}\},\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right)\right]+\ln\left(N_p^g!\right)\right\}.\end{aligned}\tag{8.5}$$

We therefore have a likelihood distribution that encodes the statistical process describing the generation of galaxy observations given a specific realization of 3D initial conditions. This data model is inherently non-linear as a result of the galaxy biasing model employed and also due to the signal dependence of Poissonian noise, which does not behave as an additive nuisance.

8.3.2 The augmented joint posterior distribution

The augmented joint posterior distribution corresponds to the following:

$$\begin{aligned}&\mathcal{P}\left(\{\delta_p^{\rm ic}\},\{\bar{N}^g\},\{b_i^g\},\{\theta_i\}|\{N_p^g\},\mathbf{S}\right)\\&\qquad\propto\mathcal{L}\left[\{N_p^g\}|\mathcal{M}_p(\{\delta_p^{\rm ic}\}),\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right]\Pi\left(\{\delta_p^{\rm ic}\}|\,\mathbf{S}\right)\Pi\left(\{\bar{N}^g\},\{b_i^g\}\right)\Pi\left(\{\theta_i\}\right),\end{aligned}\tag{8.6}$$

where the Π's correspond to the respective priors for each parameter. Hence, given our forward model $\mathcal{M}$, which incorporates sequential components of structure formation and coordinate transformation, as described above, the complex task of modelling accurate priors for the statistical behaviour of present-day matter fluctuations can be recast into a Bayesian inference problem for the initial conditions. From this joint posterior distribution, we can construct the various conditional posterior distributions for each parameter of interest. The modular statistical programming approach adopted in outlined in Section 8.4.3.

Since the non-linear LSS analysis has been reformulated as an initial conditions statistical inference problem, as described by the joint posterior distribution (8.6), this method depends solely on forward evaluations, and consequently has a definite edge over traditional approaches of initial conditions inference that require backward integration of the equations of motion or the inversion of the flow of time (e.g. Nusser and Dekel, 1992). The latter methods are prone to erroneous fluctuations in the initial density and velocity fields, resulting from spurious growth of decaying modes. Moreover, such schemes are hindered by survey incompleteness which requires the knowledge of the complex and, as yet, unknown multivariate probability distribution for the matter fluctuations, to render the backward integration of non-linear models physically meaningful via constrained realizations. In comparison, the forward modelling approach here conveniently accounts for survey masks and statistical uncertainties in the initial conditions, which amounts to modelling straightforward uncorrelated Gaussian processes, to generate data constrained realizations of the initial and evolved density fields.

8.3.3 The cosmological parameter posterior distribution

In this work, we sample the present-day values of matter density and dark energy equation of state parameters, $\{\theta_i\} = \{\Omega_\mathrm{m}, w_0\}$, via the following conditional posterior distribution,

$$\mathcal{P}\left(\{\theta_i\}|\{N_p^g\},\{\delta_p^\mathrm{ic}\},\{\bar{N}^g\},\{b_i^g\},\mathbf{S}\right) = \mathcal{P}\left(\{\theta_i\}|\{N_p^g\},\{\delta_p^\mathrm{ic}\},\{\bar{N}^g\},\{b_i^g\}\right), \tag{8.7}$$

assuming conditional independence of the cosmological power spectrum, i.e. Fourier transform of the covariance matrix $\mathbf{S}$, once the density field is known. This assumption holds as we are only probing the cosmological expansion in this work, with the power spectrum anchored with a fiducial cosmology. We quantitatively demonstrate the validity of this assumption in Section 8.8 by comparing the entropy of prior information against that of posterior information (cf. Fig. 8.11). We defer power spectrum sampling to a future work. Applying Bayes' identity, and using the joint posterior distribution from Eq. (8.6), we obtain

$$\begin{aligned}
\mathcal{P}\left(\{\theta_i\}|\{N_p^g\},\{\delta_p^\mathrm{ic}\},\{\bar{N}^g\},\{b_i^g\}\right) &= \frac{\mathcal{P}\left(\{\theta_i\},\{\delta_p^\mathrm{ic}\},\{\bar{N}^g\},\{b_i^g\}|\{N_p^g\}\right)}{\mathcal{P}\left(\{\delta_p^\mathrm{ic}\},\{\bar{N}^g\},\{b_i^g\}|\{N_p^g\}\right)} \\
&= \frac{\mathcal{L}\left[\{N_p^g\}|\mathcal{M}_p\left(\{\delta_p^\mathrm{ic}\}\right),\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right]}{\mathcal{P}\left(\{\delta_p^\mathrm{ic}\},\{\bar{N}^g\},\{b_i^g\}|\{N_p^g\}\right)} \\
&\quad \times \Pi\left(\{\delta_p^\mathrm{ic}\}|\,\mathbf{S}\right)\Pi\left(\{\bar{N}^g\},\{b_i^g\}\right)\Pi\left(\{\theta_i\}\right) \\
&\propto \mathcal{L}\left[\{N_p^g\}|\mathcal{M}_p\left(\{\delta_p^\mathrm{ic}\}\right),\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right] \times \Pi\left(\{\theta_i\}\right),
\end{aligned} \tag{8.8}$$

after omitting the terms without any cosmological parameter dependence.

Since this work attempts to demonstrate the capabilities of ALTAIR to constrain the cosmological parameters, as proof of concept, we set uniform prior distributions on θ_i. To sample from the above marginal posterior distribution, we make use of a slice sampling procedure (e.g. Neal, 2000; Neal, 2003). After obtaining a realization of θ_i, we need to update the comoving-redshift coordinate transformation in the forward model.

We adopt a dynamical dark energy model, in particular, the standard Chevallier-Polarski-Linder (CPL) parameterization (Chevallier and Polarski, 2001; Linder, 2003), where the evolution of the dark energy equation of state parameter w is a linear function of the scale factor a, as follows:

$$w = w_0 + (1-a)w_a. \tag{8.9}$$

In this work, we set $w_a = 0$ and infer the present-day value w_0. Moreover, we impose the assumption of flatness, i.e. $\Omega_\mathrm{k} = 0$, such that the dark energy density is $\Omega_\mathrm{de} = 1 - \Omega_\mathrm{m}$.

Due to the correlation between Ω_m and w_0, we perform a rotation of the (Ω_m, w_0) parameter space, using orthonormal basis transformations derived from the covariance matrix, to improve the efficiency of the slice sampler. This procedure is outlined in Section 8.6.3, with the significant gain in efficiency illustrated in Fig. 8.4. The corresponding sampler for the bias parameters is outlined in Section 8.4.4.

8.3.4 Jacobian of comoving-redshift transformation

In this section, we derive the Jacobian, $\mathcal{J}_r^z$, of the coordinate transformation between comoving $(\boldsymbol{r})$ and redshift $(\boldsymbol{z})$ space. Since we desire the redshift space representation of the density field,

$$\rho_z[\boldsymbol{z}(\boldsymbol{r})] = \rho_r(\boldsymbol{r}) \left|\frac{\partial z}{\partial r}\right|^{-1} = \rho_r(\boldsymbol{r}) \left|\mathcal{J}_z^r\right|^{-1} = \rho_r(\boldsymbol{r}) \left|\mathcal{J}_r^z\right|, \tag{8.10}$$

or in terms of the density contrast,

$$\delta_z[\boldsymbol{z}(\boldsymbol{r})] = [1 + \delta_r(\boldsymbol{r})] \left|\mathcal{J}_r^z\right| - 1, \tag{8.11}$$

where $\mathcal{J}_z^r$ is the corresponding Jacobian for the converse (redshift to comoving) transformation.

The set of three indices $\{i, j, k\}$ below are spatial coordinates $\{1, 2, 3\}$. With the functional dependence of $\boldsymbol{z} = f(\boldsymbol{r})$, we have

$$z_i = \frac{r_i}{r} f(\boldsymbol{r}) = \bar{z}\,(r)\,\frac{r_i}{r}, \tag{8.12}$$

where we defined $r = \sqrt{\sum_j r_j^2}$, and $\bar{z}$ is the cosmological redshift. Taking the derivative yields

$$\begin{aligned} \mathrm{d}z_i &= \mathrm{d}\bar{z}\,(r)\,\frac{r_i}{r} + \bar{z}\,(r)\,\mathrm{d}\left(\frac{r_i}{r}\right) \\ &= \sum_j \frac{\partial \bar{z}}{\partial r}\frac{\partial r}{\partial r_j}\mathrm{d}r_j \frac{r_i}{r} + \bar{z}\frac{\mathrm{d}r_i}{r} + \bar{z} r_i \mathrm{d}\left(\frac{1}{r}\right), \end{aligned} \tag{8.13}$$

after applying the chain rule. Differentiating with respect to the comoving coordinate, and using $\mathrm{d}\,(1/r) = -\sum_j (1/r^3) r_j \mathrm{d}r_j$,

$$\begin{aligned} \frac{\partial z_i}{\partial r_k} &= \frac{\partial \bar{z}}{\partial r}\frac{\partial r}{\partial r_k}\frac{r_i}{r} + \frac{\bar{z}}{r}\delta_{ik}^{\mathrm{k}} - \frac{z r_i r_k}{r^3} \\ &= \left(\frac{\partial \bar{z}}{\partial r} - \frac{\bar{z}}{r}\right)\frac{r_i r_k}{r^2} + \frac{\bar{z}}{r}\delta_{ik}^{\mathrm{k}}, \end{aligned} \tag{8.14}$$

where we made use of $r\mathrm{d}r = \sum_k r_k \mathrm{d}r_k$ which follows from the definition $r = \sqrt{\sum_k r_k^2}$. To compute $\partial\bar{z}/\partial r$, we start from the definition of comoving distance,

$$r = \frac{c}{H_0}\int_0^{\bar{z}} \frac{\mathrm{d}z}{\mathcal{E}(z)}, \tag{8.15}$$

with the conventional definition for $\mathcal{E}(z)$,

$$\mathcal{E}(z) = \sqrt{\Omega_{\mathrm{de}}(1+z)^{3(1+w)} + \Omega_{\mathrm{k}}(1+z)^2 + \Omega_{\mathrm{m}}(1+z)^3 + \Omega_{\mathrm{r}}(1+z)^4}, \tag{8.16}$$

where c is the speed of light and H_0 is the Hubble constant today, such that

$$\frac{\partial \bar{z}}{\partial r} = \left(\frac{H_0}{c}\right)\mathcal{E}(\bar{z}). \tag{8.17}$$

Finally, the Jacobian matrix can be expressed as follows:

$$\mathcal{J}^{r}_{z\{ik\}} = \frac{\partial z_i}{\partial r_k} = \left(\frac{H_0}{c}\mathcal{E}(\bar{z}) - \frac{\bar{z}}{r}\right)\frac{r_i r_k}{r^2} + \frac{\bar{z}}{r}\delta^{\mathrm{k}}_{ik}, \tag{8.18}$$

and we obtain the desired $\left|\mathcal{J}^{z}_{r\{ik\}}\right|$ by taking the reciprocal of $\left|\mathcal{J}^{r}_{z\{ik\}}\right|$.

8.4 The LPT-Poissonian posterior

In this section, we describe the large-scale structure (LSS) posterior distribution implemented in this work. We demonstrate how the complex problem of exploring the high dimensional joint posterior distribution can be reduced to a set of distinct steps via a multiple block sampling scheme.

8.4.1 The density posterior distribution

The primary objective is to fully characterize the 3D cosmic LSS in observations via a numerical representation of the corresponding LSS posterior using sophisticated Markov Chain Monte Carlo (MCMC) techniques, in particular to provide data constrained realizations of a set of plausible 3D density contrast amplitudes underlying a given set of galaxy observations. The posterior distribution for the evolved density field fluctuation δ^{f}_p can be formulated, in a general context, via Bayes' identity as:

$$\mathcal{P}\left(\{\delta^{\mathrm{f}}_p\}|\{N^g_p\}\right) = \frac{\mathcal{L}\left(\{N^g_p\}|\{\delta^{\mathrm{f}}_p\}\right)\Pi\left(\{\delta^{\mathrm{f}}_p\}\right)}{\Pi\left(\{N^g_p\}\right)}, \tag{8.19}$$

where N^g_p is the observed number of galaxies in voxel p, at position $\boldsymbol{x}_p$ in the sky, in redshift space, for the g^{th} galaxy sample, with the prior $\Pi(\{\delta^{\mathrm{f}}_p\})$ incorporating our *a priori* knowledge of the present-day matter fluctuations in the Universe, the likelihood $\mathcal{L}(\{N^g_p\}|\{\delta^{\mathrm{f}}_p\})$ and the normalizing factor given by the evidence $\Pi(\{N^g_p\})$.

A major stumbling block consequently arises, as discussed extensively in Jasche and Wandelt (2013a), since the inference framework requires a suitable prior $\Pi(\{\delta^{\mathrm{f}}_p\})$ which adequately describes the physical behaviour of the gravitationally evolved density field. Nevertheless, as elaborated in the following section, most attempts made in this direction so far have been based on heuristic approximations and the absence of a closed form description of the present day matter fluctuations encoded in a multivariate probability density distribution still persists.

However, Jasche and Wandelt (2013a) proposed an elegant approach to circumvent this key impediment based on the following assertions: There is substantial evidence that primordial seed fluctuations at redshifts $z \sim 1000$ can be modelled as a Gaussian random field to great accuracy (e.g. Linde, 2008; Komatsu et al., 2011; Planck Collaboration et al., 2016d), consistent with inflationary theories and CMB observations. Moreover, the evolution of the initial conditions relies solely on deterministic gravitational structure formation processes. Therefore, a conceptually reasonable alternative to modelling the complex statistical behaviour of the non-linear matter distribution is to formulate the inference problem at the level of the initial conditions adequately described by

Gaussian statistics. This constitutes the conceptual foundation of the BORG framework (Jasche and Wandelt, 2013a).

Given a forward model $\mathcal{M}_p$ that connects the initial conditions $\delta_p^{\mathrm{ic},(r)}$, in comoving $(\boldsymbol{r})$ space, to the redshift $(\boldsymbol{z})$ space representation of the final density field $\delta_p^{\mathrm{f},(z)}$, we can therefore express the conditional posterior for the evolved density field as

$$\mathcal{P}\left(\{\delta_p^{\mathrm{f}}\}|\{\delta_p^{\mathrm{ic}}\}\right) = \prod_p \delta^{\mathrm{D}}\left[\delta_p^{\mathrm{f}} - \mathcal{M}_p\left(\{\delta_p^{\mathrm{ic}}\}\right)\right], \tag{8.20}$$

where $\delta^{\mathrm{D}}(x)$ denoting the Dirac delta distribution encapsulates the assumption that the structure formation process is deterministic. Within this generic framework, the forward model may be generalized to a chain of arbitrary components linking the primordial density fluctuations to the present-day density contrast. Nevertheless, at its crux lies a cosmic structure formation model $\mathcal{G}_p(a, \{\delta_p^{\mathrm{ic}}\})$ for the non-linear evolution of initial conditions into a final density field at a given scale factor a, i.e. $\delta_p^{\mathrm{f},(r)} = \mathcal{G}_p(a, \{\delta_p^{\mathrm{ic},(r)}\})$. The forward model implemented in this work to encode the AP test is described in Section 8.3 above.

We can then obtain a prior distribution for δ_p^{f} via a two-step sampling procedure: First, a realization of δ_p^{ic} is generated from the prior distribution $\Pi(\{\delta_p^{\mathrm{ic}}\})$ and subsequently evolved with a given forward model $\mathcal{M}_p(\{\delta_p^{\mathrm{ic}}\})$. This essentially implies generating samples from the joint prior distribution of δ_p^{ic} and δ_p^{f}:

$$\begin{aligned}\mathcal{P}\left(\{\delta_p^{\mathrm{f}}\}, \{\delta_p^{\mathrm{ic}}\}\right) &= \Pi\left(\{\delta_p^{\mathrm{ic}}\}\right)\mathcal{P}\left(\{\delta_p^{\mathrm{f}}\}|\{\delta_p^{\mathrm{ic}}\}\right) \\ &= \Pi\left(\{\delta_p^{\mathrm{ic}}\}\right)\prod_p \delta^{\mathrm{D}}\left[\delta_p^{\mathrm{f}} - \mathcal{M}_p\left(\{\delta_p^{\mathrm{ic}}\}\right)\right],\end{aligned} \tag{8.21}$$

after plugging in the conditional posterior distribution from Eq. (8.20).

Assuming a normal distribution with zero mean and covariance $\mathbf{S}$ corresponding to an underlying cosmological power spectrum for the initial conditions δ_p^{ic}, the joint prior distribution can be expressed as

$$\begin{aligned}\mathcal{P}\left(\{\delta_p^{\mathrm{f}}\}, \{\delta_p^{\mathrm{ic}}\}|\,\mathbf{S}\right) &= \Pi\left(\{\delta_p^{\mathrm{ic}}\}|\,\mathbf{S}\right)\prod_p \delta^{\mathrm{D}}\left[\delta_p^{\mathrm{f}} - \mathcal{M}_p\left(\{\delta_p^{\mathrm{ic}}\}\right)\right] \\ &= \frac{\mathrm{e}^{-\frac{1}{2}\sum_{p,q}\delta_p^{\mathrm{ic}}\mathbf{S}^{-1}\delta_q^{\mathrm{ic}}}}{\det(2\pi\,\mathbf{S})}\prod_p \delta^{\mathrm{D}}\left[\delta_p^{\mathrm{f}} - \mathcal{M}_p\left(\{\delta_p^{\mathrm{ic}}\}\right)\right].\end{aligned} \tag{8.22}$$

Reformulating the statistical inference problem in terms of the initial conditions δ_p^{ic} results in the following joint posterior distribution of δ_p^{ic} and δ_p^{f}, conditional on the observed galaxy number counts N_p^g,

$$\mathcal{P}\left(\{\delta_p^{\mathrm{f}}\}, \{\delta_p^{\mathrm{ic}}\}|\{N_p^g\}, \mathbf{S}\right) = \frac{\mathcal{L}\left(\{N_p^g\}|\{\delta_p^{\mathrm{f}}\}, \{\delta_p^{\mathrm{ic}}\}\right)\mathcal{P}\left(\{\delta_p^{\mathrm{f}}\}, \{\delta_p^{\mathrm{ic}}\}|\,\mathbf{S}\right)}{\Pi\left(\{N_p^g\}|\,\mathbf{S}\right)}, \tag{8.23}$$

after making the dependence on the underlying power spectrum explicit. Assuming that the galaxy observations are conditionally independent of the initial conditions once the final density field

is known, i.e. $\mathcal{L}(\{N_p^g\}|\{\delta_p^{\rm f}\},\{\delta_p^{\rm ic}\}) = \mathcal{L}(\{N_p^g\}|\{\delta_p^{\rm f}\})$ and using the joint prior distribution from Eq. (8.22) leads to the LSS posterior distribution:

$$\mathcal{P}\left(\{\delta_p^{\rm f}\},\{\delta_p^{\rm ic}\}|\{N_p^g\},\mathbf{S}\right) = \mathcal{L}\left(\{N_p^g\}|\{\delta_p^{\rm f}\}\right)\frac{\Pi\left(\{\delta_p^{\rm ic}\}|\,\mathbf{S}\right)}{\Pi\left(\{N_p^g\}|\,\mathbf{S}\right)} \times \prod_p \delta^{\rm D}\left[\delta_p^{\rm f} - \mathcal{M}_p\left(\{\delta_p^{\rm ic}\}\right)\right]. \tag{8.24}$$

By marginalizing over the final density field $\delta_p^{\rm f}$, we finally obtain our posterior distribution as follows:

$$\mathcal{P}\left(\{\delta_p^{\rm ic}\}|\{N_p^g\},\mathbf{S}\right) = \mathcal{L}\left[\{N_p^g\}|\mathcal{M}_p\left(\{\delta_p^{\rm ic}\}\right)\right]\frac{\Pi\left(\{\delta_p^{\rm ic}\}|\,\mathbf{S}\right)}{\Pi\left(\{N_p^g\}|\,\mathbf{S}\right)}, \tag{8.25}$$

thereby connecting present galaxy observations N_p^g to their corresponding primordial density fields $\delta_p^{\rm ic}$ via a given forward model, $\mathcal{M}_p(\{\delta_p^{\rm ic}\})$. We must therefore sample from the highly non-linear and non-Gaussian posterior above to obtain realizations of the 3D initial density fields conditioned on the galaxy observations via the sophisticated HMC method described below in Section 8.5.

8.4.2 Choice of density prior

Standard Wiener filtering approaches employ isotropic Gaussian priors on the present-day density contrast which is justified for inference on the largest scales (e.g. Zaroubi, Hoffman, and Dekel, 1999; Zaroubi, 2002; Erdoğdu et al., 2004; Erdoğdu et al., 2006; Kitaura and Enßlin, 2008; Kitaura et al., 2009; Jasche et al., 2010b; Jasche and Lavaux, 2015), where the density field can be reasonably approximated as a Gaussian random field. Although the exact probability distribution for the density field in the mildly and strongly non-linear regimes is not known, the deviation from Gaussianity is well-established and we therefore require a non-Gaussian prior to effectively capture the details of the highly complex filamentary nature of the present day cosmic web.

Lognormal density priors were subsequently proposed (Coles and Jones, 1991) and this proved to be an adequate phenomenological choice (e.g. Hubble, 1934; Peebles, 1980; Gaztanaga and Yokoyama, 1993; Kayo, Taruya, and Suto, 2001), albeit with some limitations, for modelling the evolved matter field in the mildly non-linear regime (Kitaura et al., 2009; Jasche and Kitaura, 2010; Jasche et al., 2010a). The use of Edgeworth expansions to construct prior distributions involving third order moments has also been proposed in the literature (Kitaura, 2012).

In this work, we rely on Lagrangian Perturbation Theory (LPT) to model cosmic structure formation, which accounts for non-local effects of gravitational mass transport from initial to final positions. It has been extensively shown that LPT provides a sufficient description of the cosmic LSS on large scales, where it is capable of reproducing the exact one-, two- and three-point statistics of the evolved field, while still being a reasonable approximation to the higher order statistics (e.g. Moutarde et al., 1991; Buchert, Melott, and Weiss, 1994; Bouchet et al., 1995; Scoccimarro, 2000; Scoccimarro and Sheth, 2002).

The essence of the above approach is that it provides a pathway to recover the high-order statistics of the matter distribution using only the 2-point statistics of the initial conditions, obviating the need for additional parameters to describe higher order statistics for the matter inference problem.

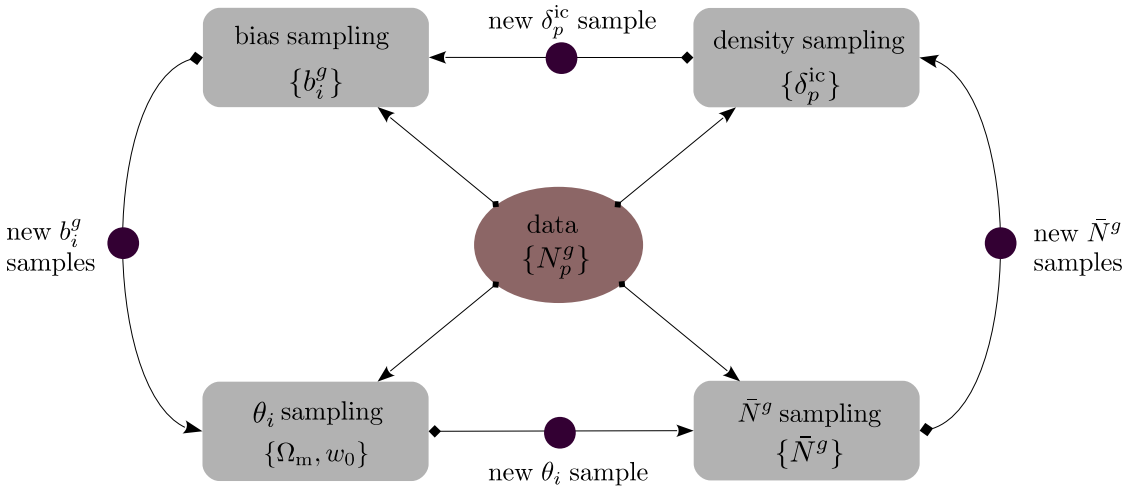


FIGURE 8.3: Schematic representation of the multi-step iterative block sampling procedure implemented in ALTAIR. Initially, a realization of the 3D density contrast is obtained conditional on galaxy observations, followed by the other parameters being sampled conditional on the respective previous samples. Subsequent iterations of this procedure yield samples from the full joint posterior distribution within a modular statistical programming framework.

Moreover, since our model encodes the dynamics, the reconstruction of the large-scale velocity field is automatically generated without incurring the expense of an increased parameter space.

8.4.3 Modular statistical programming

In this work, we exploit the modular statistical programming framework inherent in the BORG algorithm to encode the AP test as an additional component to the original block sampling machinery. This approach allows us to model any Bayesian hierarchical problem to take into account any observational systematics via data models of higher complexity. Here, we account for the unknown parameters $\{b_i^g\}$ of the galaxy bias model described in Section 8.3.1 and unknown normalizations $\{\bar{N}^g\}$ for distinct galaxy samples, as illustrated in Fig. 8.3. These last normalizations, in practice, scale as the noise amplitudes as we are using a per-voxel Poissonian likelihood.

Conceptually, within such a framework, the overall aim is to characterize fully the augmented joint posterior $\mathcal{P}(\{\delta_p^{\rm ic}\}, \{\bar{N}^g\}, \{b_i^g\}, \{\theta_i\}|\{N_p^g\}, \mathbf{S})$ of different tracer populations labelled by the index g. Since it is not computationally feasible and advisable to sample directly from the high dimensional joint posterior distribution, we make use of an important theorem on Metropolis-Hastings block sampling which allows us to break the high dimensional sampling problem into a series of lower dimensional ones (Hastings, 1970). We therefore dissect the exploration of the full joint parameter space into a sequence of conditional sampling procedures.

The block sampling approach consists of drawing samples from the following conditional probability distributions:

$$\begin{aligned}
&(1)\ \{\delta_p^{\rm ic}\}^{s+1} \curvearrowleft \mathcal{P}\left(\{\delta_p^{\rm ic}\}|\{\theta_i\}^s, \{N_p^g\}^s, \{\bar{N}^g\}^s, \{b_i^g\}^s, \mathbf{S}\right)\\
&(2)\ \{\bar{N}^g\}^{s+1} \curvearrowleft \mathcal{P}\left(\{\bar{N}^g\}|\{\theta_i\}^s, \{N_p^g\}^s, \{\delta_p^{\rm ic}\}^s, \{b_i^g\}^s, \mathbf{S}\right)\\
&(3)\ \{b_i^g\}^{s+1} \curvearrowleft \mathcal{P}\left(\{b_i^g\}|\{\theta_i\}^s, \{\bar{N}^g\}^s, \{N_p^g\}^s, \{\delta_p^{\rm ic}\}^s, \mathbf{S}\right)\\
&(4)\ \{\theta_i\}^{s+1} \curvearrowleft \mathcal{P}\left(\{\theta_i\}|\{\bar{N}^g\}^s, \{N_p^g\}^s, \{\delta_p^{\rm ic}\}^s, \{b_i^g\}^s, \mathbf{S}\right).
\end{aligned} \tag{8.26}$$

In the above expressions, s denotes the sampling step and the symbol $\curvearrowleft$ indicates sampling from the distribution on the right hand side. A series of iterations of the individual sampling steps above will converge to the target distribution, i.e. the full joint posterior distribution (Hastings, 1970). Hence, by simultaneously exploring the 3D initial density field $\{\delta_p^{\rm ic}\}$, the galaxy bias parameters $\{b_i^g\}$, the galaxy density contrast normalizations $\{\bar{N}^g\}$ and the cosmological parameters $\{\theta_i\}$ via an implementation of high dimensional MCMC methods in a multiple block sampling procedure, we can obtain samples from the desired joint posterior distribution.

8.4.4 The bias posterior distribution

The formalism for the data model is presented in a generic context in Section 8.3.1, such that it can be implemented for two or more different galaxy surveys, distinct subsamples of a given catalogue or even different cosmological probes of the LSS. The advantage of splitting a galaxy sample into various subsamples is that we can treat the respective systematic and statistical uncertainties of each subsample separately, thereby allowing us to account for the distinct clustering behaviour of galaxy populations in the LSS sample via their respective bias parameters.

These additional parameters can be trivially incorporated in the flexible block sampling approach adopted here, as described in Section 8.4.3. The mean numbers of galaxies, $\bar{N}^g$, for the various subsamples are essential for defining the density contrast of the galaxy distribution, with an erroneous value of $\bar{N}^g$ resulting in a non-zero value of the mean, yielding spurious large-scale power in the inferred density fields (Jasche and Wandelt, 2013b). Due to a lack of a priori knowledge, we must perform a joint inference of the set of four $\bar{N}^g$ and b_i^g bias parameters, together with initial and final density fields, to take into account possible cross-correlations and interdependencies. Unlike traditional approaches (e.g. Tegmark et al., 2004), here we infer the bias factors directly from the relation between the data and the density field.

The conditional posterior distribution for the bias parameters, given a realization of the density field and galaxy number counts for the respective subcatalogues, can be expressed as:

$$\mathcal{P}\left(\{\bar{N}^g\}, \{b_i^g\}|\{\theta_i\}, \{N_p^g\}, \{\delta_p^{\rm ic}\}\right) \propto \mathcal{L}\left[\{N_p^g\}|\mathcal{M}_p\left(\{\delta_p^{\rm ic}\}\right), \{\theta_i\}, \{\bar{N}^g\}, \{b_i^g\}\right] \times \Pi\left(\{\bar{N}^g\}, \{b_i^g\}\right), \tag{8.27}$$

following an analogous reasoning to that described in Section 8.3.3. Adopting a standard maximally agnostic philosophy, we set uniform prior distributions for the bias parameters:

$$\Pi\left(\{\bar{N}^g\}, \{b_i^g\}\right) = \Theta\left(\{\bar{N}^g\}\right)\Theta\left(\{b_i^g\}\right), \tag{8.28}$$

where the Heaviside function $\Theta(x)$ ensures that the parameters are positive, as required by the bias model. The non-linear shape of the galaxy biasing function, as given in Eq. (8.4), poses a particular challenge as no straightforward sampling procedure can be derived. We therefore sample from the above conditional bias posterior distribution via sequential slice sampling steps (e.g. Neal, 2000; Neal, 2003) to ensure unit acceptance rates.

8.5 Hamiltonian sampling

The exploration of the LPT-Poissonian posterior, described by Eq. (8.6), requires highly non-linear reconstruction methods and is therefore numerically intensive. From the Bayesian viewpoint considered in this work, a single estimate of the density field is not of particular interest, with the desired scientific output being a sampled representation of the multidimensional LSS posterior. Extracting any relevant statistical summary, such as mean, mode or variance, given this posterior representation, is then straightforward. Furthermore, the propagation of uncertainties to the final inferred quantities is smooth and coherent.

In the absence of a known procedure to directly sample from a LPT-Poissonian distribution, a Metropolis-Hastings sampling scheme with an accept-reject step must be implemented. But such a mechanism suffers from the well-known drawback of possible high rejection rate, where a low acceptance rate of proposed samples will render the method numerically inefficient. This is especially significant since the inference of 3D density fields typically involves extremely high number of inference parameters. For instance, in this work, there are over 2×10^6 free parameters, which correspond to primordial density fluctuation amplitudes δ_p^{ic} at respective grid nodes. Exploring this high dimensional parameter space via a random walk, as in conventional Markov Chain Monte Carlo (MCMC) methods, consequently results in a high rejection rate. This is an understatement, as in practice, the acceptance rate is virtually zero, with one chance over 10^6 of going in the right direction, without taking into account the step size.

To limit the number of samples and alleviate the numerical scaling of the method, we implement a Hamiltonian Monte Carlo (HMC) scheme. This method guarantees an acceptance rate of unity in the absence of numerical errors. The numerical efficiency of the Hamiltonian sampling lies in deterministically proposing new samples to the Metropolis-Hastings algorithm, based on techniques to follow dynamical particle motion in potentials. The HMC algorithm therefore maintains a reasonable sampling efficiency in high dimensional spaces by suppressing the dominant random walk behaviour.

The HMC method has been successfully implemented for the inference of density fields in the non-linear regime, and has been found to be very efficient (e.g. Jasche and Kitaura, 2010; Jasche et al., 2010a; Kitaura, Gallerani, and Ferrara, 2012; Jasche and Wandelt, 2013a; Jasche and Wandelt, 2013b). Moreover, this approach is less prone to spurious effects induced by numerical inaccuracies due to the final accept/reject step of the Metropolis-Hastings sampler ensuring correctness of the target density. Numerical inaccuracies would only be detrimental to the efficiency, without comprimising the correctness.

We briefly review the underlying framework and rationale of the HMC method. An excellent in-depth review of this approach is provided in Duane et al. (1987) and Neal (1993), with its application to LSS inference illustrated in Jasche and Kitaura (2010), Jasche et al. (2010a), Jasche and Wandelt (2013a) and Jasche and Wandelt (2013b) and Jasche and Lavaux (2019).

8.5.1 Theoretical background and review

The essence of the HMC method is as follows: If we want to generate random variates according to a desired probability distribution $\mathcal{P}(\{x_i\})$, with $\{x_i\}$ being a set of N elements x_i, then we may interpret the negative logarithm of the distribution as a potential $\psi(x)$,

$$\psi(x) = -\ln(\mathcal{P}(x)). \tag{8.29}$$

We can subsequently formulate a corresponding Hamiltonian that describes the dynamics in the multidimensional phase space by adding a kinetic term to the above potential. To this end, we introduce a "momentum" variable p_i and a "mass matrix" $\mathbf{M}$ as nuisance parameters, as follows:

$$H = \frac{1}{2}\sum_{i,j} p_i M^{-1}_{ij} p_j + \psi(x). \tag{8.30}$$

This Hamiltonian, as in classical mechanics, describes the dynamics in a high dimensional parameter space. We then obtain the new target probability distribution by exponentiating the above Hamiltonian:

$$e^{-H} = \mathcal{P}(\{x_i\}) \exp\left(\frac{1}{2}\sum_{i,j} p_i M^{-1}_{ij} p_j\right). \tag{8.31}$$

The form of the Hamiltonian ensures that the new joint distribution can be separated into a Gaussian distribution for the momenta $\{p_i\}$ and the target distribution $\mathcal{P}(\{x_i\})$. This implies that the two sets of variables $\{x_i\}$ and $\{p_i\}$ are independent, and hence, we can obtain samples from the target distribution $\mathcal{P}(\{x_i\})$ simply by marginalizing over the auxiliary momenta.

The next step is to generate a random variate from the joint distribution, which is proportional to $\exp(-H)$. We must first draw a set of momenta from the distribution defined by the kinetic energy term, which is an N-dimensional Gaussian with covariance $\mathbf{M}$. To obtain a new sample from the joint distribution, we allow the system to evolve deterministically, from a given initial point in the high dimensional parameter space $(\{x_i\}, \{p_i\})$, according to Hamilton's equations:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i} \tag{8.32}$$

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \frac{\partial H}{\partial x_i} = -\frac{\partial \psi(x)}{\partial x_i}. \tag{8.33}$$

Integrating the above equations of motion for the time variable t up to a fixed pseudo-time τ results in the new position $(\{x_i'\}, \{p_i'\})$ in phase space. This new point in phase space is accepted according to the standard Metropolis-Hastings acceptance rule:

$$\mathcal{P}_{\mathcal{A}} = \min\left\{1, \exp\left[-\left(H\left(\{x_i'\}, \{p_i'\}\right) - H\left(\{x_i\}, \{p_i\}\right)\right)\right]\right\}. \tag{8.34}$$

Since the equations of motion provide a solution to a Hamiltonian system, the energy or the Hamiltonian described by Eq. (8.30) must be conserved. A direct consequence of this conservation of the Hamiltonian is that the HMC approach yields an acceptance rate of unity. In practice, however, the acceptance rate may be lower due to numerical inaccuracies in the integration scheme. After accepting a new sample, the mechanism proceeds by discarding the momentum counterpart and restarts the sampling procedure by randomly drawing a new set of momenta. The individual momenta $\{p_i\}$ do not have to be stored in memory which is numerically convenient, and marginalization simply implies discarding them.

The HMC sampling scheme can therefore be summarized in two steps. The first step is essentially a Gibbs sampling step which results in a new set of Gaussian distributed momenta. The second step involves computing the deterministic dynamical trajectory on the posterior surface. We obtain samples of the desired target distribution by marginalizing over the nuisance parameters, i.e. by discarding the auxiliary momenta $\{p_i\}$.

8.5.2 Equations of motion for large-scale structure inference

As described above, the Hamiltonian Monte Carlo (HMC) approach allows the exploration of the non-linear LSS posterior via the study of the Hamiltonian dynamics in the high dimensional parameter space. To ensure a high acceptance rate, the HMC exploits the gradient of the logarithmic posterior distribution to optimally explore this parameter space, such that the algorithm also requires the derivatives of the posterior distribution, as outlined below.

From the LSS posterior illustrated in Eq. (8.6), we can derive the corresponding forces required to evaluate the Hamiltonian trajectories. The Hamiltonian potential $\psi(\delta^{\rm ic})$ can be written as

$$\begin{aligned}\psi(\delta^{\rm ic}) &= -\ln\left[\mathcal{P}\left(\{\delta_p^{\rm ic}\},\{\bar{N}^g\},\{b^g\},\{\theta_i\}|\{N_p^g\},\mathbf{S}\right)\right] \\ &= \psi_{\rm likelihood}(\delta^{\rm ic}) + \psi_{\rm prior}(\delta^{\rm ic}),\end{aligned} \tag{8.35}$$

where the potential $\psi_{\rm prior}(\delta^{\rm ic})$ corresponds to

$$\psi_{\rm prior}(\delta^{\rm ic}) = \frac{1}{2}\delta^{\rm ic,\dagger}\mathbf{S}^{-1}\delta^{\rm ic} = \frac{1}{2}\sum_{ij}\delta_i^{\rm ic,*}S_{ij}^{-1}\delta_j^{\rm ic}, \tag{8.36}$$

while $\psi_{\rm likelihood}(\delta^{\rm ic})$ is given by

$$\psi_{\rm likelihood}(\delta^{\rm ic}) = \sum_p\left\{\lambda_p^g\left(\{\delta_p^{\rm f}\},\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right) - N_p^g\ln\left[\lambda_p^g\left(\{\delta_p^{\rm f}\},\{\theta_i\},\{\bar{N}^g\},\{b_i^g\}\right)\right]\right\}. \tag{8.37}$$

Using the definition of the Hamiltonian potential $\psi(\delta^{\rm ic})$ from Eq. (8.35), we obtain the corresponding equations describing the Hamiltonian forces by differentiating with respect to $\delta^{\rm ic}$:

$$\frac{\partial\psi(\delta^{\rm ic})}{\partial\delta_i^{\rm ic}} = \frac{\partial\psi_{\rm prior}(\delta^{\rm ic})}{\partial\delta_i^{\rm ic}} + \frac{\partial\psi_{\rm likelihood}(\delta^{\rm ic})}{\partial\delta_i^{\rm ic}}. \tag{8.38}$$

The prior term is trivially obtained as

$$\frac{\partial \psi_{\text{prior}}(\delta^{\text{ic}})}{\partial \delta_i^{\text{ic}}} = \sum_j \delta_j^{\text{ic},*} S_{ij}^{-1}. \tag{8.39}$$

The corresponding likelihood term, however, cannot be derived in straightforward fashion. We compute this adjoint gradient as a sequential application of linear operations, as follows:

$$\frac{\partial \psi_{\text{likelihood}}(\delta^{\text{ic}})}{\partial \delta_s^{\text{ic}}} = \sum_q \frac{\partial \rho_q^{(r)}}{\partial \delta_s^{\text{ic}}} \sum_{\tilde{p}} \frac{\partial \rho_{\tilde{p}}^{(z)}}{\partial \rho_q^{(r)}} \sum_{p,g} \frac{\partial \lambda_p^g}{\partial \rho_{\tilde{p}}^{(z)}} \frac{\partial \psi_{\text{likelihood}}}{\partial \lambda_p^g}, \tag{8.40}$$

where we explicitly expressed the coordinate transformation from comoving (r) to redshift (z) space, and $\rho_p = 1 + \delta_p^{\text{f}}$ is the density field. Eq. (8.40) constitutes a sequence of matrix vector applications, as follows:

$$\widetilde{\mathcal{G}}_{qs} \equiv \frac{\partial \rho_q^{(r)}}{\partial \delta_s^{\text{ic}}} \tag{8.41}$$

$$\mathcal{Q}_{\tilde{p}q} \equiv \frac{\partial \rho_{\tilde{p}}^{(z)}}{\partial \rho_q^{(r)}} = \sum_i \mathcal{E}_{i\tilde{p}}^{-1} \sum_{\tilde{p}} \mathcal{J}_{\tilde{p}} x_{\tilde{p}}^{\alpha(i)} y_{\tilde{p}}^{\beta(i)} z_{\tilde{p}}^{\gamma(i)} \tag{8.42}$$

$$\mathcal{K}_{p\tilde{p}} \equiv \sum_g \frac{\partial \lambda_p^g}{\partial \rho_{\tilde{p}}^{(z)}} = \sum_g \delta_{p,\tilde{p}}^{\text{k}} \frac{\lambda_p^g}{\rho_{\tilde{p}}^{(r)}} \left[\rho^g \epsilon^g \rho_{\tilde{p}}^{-\epsilon_{g'}(r)} + \beta \right] \tag{8.43}$$

$$v_p \equiv \frac{\partial \psi_{\text{likelihood}}}{\partial \lambda_p^g} = 1 - \frac{N_p^g}{\lambda_p^g \left(\{\delta_p^{\text{f}}\}, \{\theta_i\}, \{\bar{N}^g\}, \{b_i^g\} \right)}, \tag{8.44}$$

where $\mathcal{J}$ is the Jacobian of the comoving-redshift transformation, $\widetilde{\mathcal{G}}$ indicates the derivative of LPT (cf. Appendix D in Jasche and Wandelt (2013a)) and $\delta_{p,\tilde{p}}^{\text{k}}$ is the Kronecker delta. The derivation of the adjoint gradient for the triquintic interpolation is illustrated in the following section.

Using the above results, the two equations of motion (8.32) and (8.33) for the Hamiltonian system can be expressed as follows:

$$\frac{\mathrm{d}\hat{\delta}_i^{\text{ic}}}{\mathrm{d}t} = \sum_j M_{ij}^{-1} p_j, \tag{8.45}$$

and

$$\frac{\mathrm{d}\hat{p}_i}{\mathrm{d}t} = -\sum_j \delta_j^{\text{ic},*} S_{ij}^{-1} - \sum_{q,\tilde{p},p} \widetilde{\mathcal{G}}_{qi}\, \mathcal{Q}_{\tilde{p}q}\, \mathcal{K}_{p\tilde{p}}\, v_p, \tag{8.46}$$

where the hats denote Fourier space representation. We can now obtain new samples from the LSS posterior by following the dynamical evolution of the Hamiltonian system, governed by Eqs. (8.45) and (8.46), in phase space.

8.5.3 Adjoint gradient for generic 3D interpolation

The 3D interpolation can be expressed generally, for any arbitrary n^{th} order of interpolation, as:

$$P_{\alpha\beta\gamma} = \sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n} a_{ijk} x_{\alpha}^{i} y_{\beta}^{j} z_{\gamma}^{k} = \sum_{i,j,k}^{n} a_{ijk} \mathcal{E}_{\alpha\beta\gamma}{}^{ijk}, \tag{8.47}$$

where $\mathbf{P}$ is the interpolated surface and $\boldsymbol{\mathcal{E}}$ is a $(n+1)^3 \times (n+1)^3$ matrix required for computing the vector of interpolation coefficients $\boldsymbol{a}$. In this work, we implement a triquintic interpolation scheme, which corresponds to $n = 5$, with the indices $\{i,j,k\}, \{\alpha,\beta,\gamma\} = \{0,1,2,3,4,5\}$ constituting the particular indexing scheme employed to denote the voxels involved. $\boldsymbol{\mathcal{E}}$ is therefore a 216×216 matrix for the triquintic scheme.

The above system of equations can be reformulated as a matrix for the linear system described by $\mathbf{P} = \boldsymbol{\mathcal{E}}\boldsymbol{a}$, such that the vector of interpolation coefficients can be computed in straightforward fashion through matrix inversion, $\boldsymbol{a} = \boldsymbol{\mathcal{E}}^{-1}\mathbf{P}$. The advantage of this approach is that $\boldsymbol{\mathcal{E}}^{-1}$ can computed only once and stored, and then used for interpolation at any location inside the cube (e.g. Lekien and Marsden, 2005).

The derivative of the generic 3D interpolation is required as a component of the adjoint gradient at the core of the HMC sampler described in Section 8.5.2. The 3D interpolated density field, in redshift space (cf. Eq. (8.10)), can be written explicitly as:

$$\tilde{\rho}_k = \mathcal{J}_k \left(\sum_{i,m} \mathcal{E}_{im}^{-1} \rho_m x_k^{\alpha(i)} y_k^{\beta(i)} z_k^{\gamma(i)} \right), \tag{8.48}$$

where $\boldsymbol{\mathcal{J}}$ is the Jacobian of the comoving-redshift transformation from Section 8.3.4, evaluated at the location of the mesh element k. Here, we employed a compact notation, where $\{x,y,z\}$ correspond to the fractional steps with respect to a reference node in the grid, i.e. interpolation weights, and $\{\alpha,\beta,\gamma\}$ label the powers of the elements of $\boldsymbol{\mathcal{E}}$ via:

$$i = \alpha(n+1)^2 + \beta(n+1) + \gamma, \tag{8.49}$$

with the index $m \in \{i,j,k\}$ from Eq. (8.47). Using the chain rule, the resulting gradient $\boldsymbol{w}$, after application of the incoming gradient $\boldsymbol{v}$ from the previous components in the forward model, is obtained as follows:

$$\begin{aligned} w_a &= \sum_k v_k \frac{\partial \tilde{\rho}_k}{\partial \rho_a} \\ &= \sum_k \mathcal{J}_k v_k \sum_{i,m} \mathcal{E}_{im}^{-1} \delta_{ma}^{\mathrm{k}} x_k^{\alpha(i)} y_k^{\beta(i)} z_k^{\gamma(i)} \\ &= \sum_i \mathcal{E}_{ia}^{-1} \sum_k \mathcal{J}_k v_k x_k^{\alpha(i)} y_k^{\beta(i)} z_k^{\gamma(i)} \\ &= \sum_{i,k} \mathcal{E}_{ia}^{-1} \xi_{ik}, \end{aligned} \tag{8.50}$$

where $\xi_{ik} \equiv \mathcal{J}_k v_k x_k^{\alpha(i)} y_k^{\beta(i)} z_k^{\gamma(i)}$ represents the contribution to the adjoint gradient from a specific voxel labelled by the indexing scheme adopted. This is convenient as $\boldsymbol{\mathcal{E}}^{-1}$ is already available from the execution of the 3D interpolation in the forward model.

8.6 Numerical implementation

Our numerical implementation of the augmented version of BORG, that incorporates cosmological parameter inference, while assuming a more realistic non-linear bias model, is designated as ALTAIR. It employs the FFTW3 library for fast Fourier transforms (Frigo and Johnson, 2005), whose feature of parallel transforms allows for straightforward parallelization of our code. For random number generation, we make use of the GNU scientific library (gsl) (Galassi et al., 2009), in particular, the Mersenne Twister MT19937, with 32-bit word length, from the `gsl_rng_mt19937` routine. The use of the Mersenne Twister algorithm as a pseudo-random number generator for Monte Carlo simulations has been proven to be efficient and dependable (Matsumoto and Nishimura, 1998).

8.6.1 The leapfrog scheme

In order to obtain samples from the LSS posterior, we must integrate the equations of motion (8.45) and (8.46) numerically over a pseudo-time interval τ by means of a leapfrog scheme. Ideally, we want to solve the equations exactly for an optimal acceptance rate. As such, the choice of this integrator is motivated by several reasons. Primarily, the leapfrog scheme is convenient as it is a symplectic integrator and is therefore exactly reversible, thereby maintaining detailed balance (Duane et al., 1987) of the Markov chain. The high accuracy of such an integration scheme is conducive to the conservation of the total Hamiltonian along a given trajectory, within the limit of numerical errors, resulting in high acceptance rates. It is also numerically stable and allows for simple propagation of errors. To avoid resonant trajectories, τ is drawn from a uniform distribution. The numerical implementation of the "kick-drift-kick" leapfrog scheme here follows closely the descriptions provided in Jasche and Kitaura (2010) and Jasche and Wandelt (2013a), which we refer the reader to for more elaborate and complementary explanations.

8.6.2 Hamiltonian mass

The "mass matrix" $\mathbf{M}$ contains a large number of parameters that must be tuned to optimize the performance of the HMC sampler. Conceptually, this Hamiltonian mass characterizes the inertia of individual parameters as they move through the parameter space. As a result, the choice of $\mathbf{M}$ is a compromise to limit slow exploration efficiency due to extremely large masses and avoid large numerical errors of the integration scheme due to extremely light masses, and is therefore a trade-off between efficiency and acceptance rate.

A general prescription for the Hamiltonian masses, when the density fluctuations δ_p^{f} can be characterized as a Gaussian random field, is to set them inversely proportional to the variance of that specific δ_p^{f} (Taylor, Ashdown, and Hobson, 2008). However, this prescription has been found to be effective even for the case of non-Gaussian distributions, such as the LPT-Poissonian posterior encountered in this work (Jasche and Wandelt, 2013a; Jasche, Leclercq, and Wandelt, 2015; Lavaux

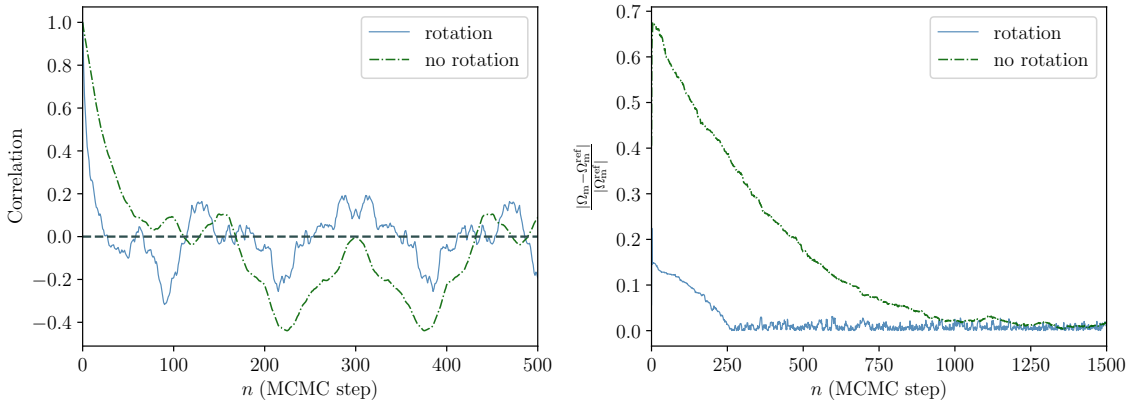


FIGURE 8.4: *Left panel:* The correlation length of the Ω_m Markov chain. The chain has a correlation length of the order of 25 samples. The dashed-dotted lines depict the longer correlation length without the rotation of the cosmological parameter space, by nearly a factor of five. *Right panel:* The burn-in phase for both scenarios, illustrated via the deviation from the reference cosmology. The dashed-dotted lines depict the MCMC chains obtained without rotating the parameter space, where the burn-in phase consequently requires around five times more iterations. The longer correlation length and burn-in phase highlight the loss in efficiency compared to the more sophisticated sampler.

and Jasche, 2016). Since the leapfrog scheme requires the inversion of $\mathbf{M}$, a diagonal representation of $\mathbf{M}$ is numerically convenient, given the extremely high dimensionality of the problem. We therefore choose $\mathbf{M}$ to be inversely proportional to the cosmological power spectrum, $\mathbf{M} = \mathbf{S}^{-1}$, i.e. diagonal in its Fourier basis. This choice, moreover, improves the efficiency of the HMC sampler and results in faster convergence since it accounts for the correlation structure of the underlying density field (Jasche and Wandelt, 2013a).

8.6.3 Rotation of cosmological parameter space

To increase the efficiency of the cosmological parameter sampler, we rotate the (Ω_m, w_0) parameter space, using orthogonal basis transformations, before slice sampling. The corresponding covariance matrix can be orthonormally decomposed as $\mathbf{C} = \mathbf{Q}^\dagger \mathbf{\Delta} \mathbf{Q}$, where $\mathbf{Q}^\dagger \mathbf{Q} = \mathbb{1}$.

We perform a rotation of the $\boldsymbol{\theta}$ vector of cosmological parameters, about the mean $\bar{\boldsymbol{\theta}}$, using the transformation:

$$\tilde{\boldsymbol{\theta}} = \mathbf{\Delta}^{-\frac{1}{2}} \mathbf{Q} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}), \tag{8.51}$$

where the tilde denotes the transformed variable. A key point is that we must rotate back to the original frame when computing the logarithm of the posterior via

$$\boldsymbol{\theta} = \mathbf{Q}^\dagger \mathbf{\Delta}^{\frac{1}{2}} \tilde{\boldsymbol{\theta}} + \bar{\boldsymbol{\theta}}. \tag{8.52}$$

This is implemented at each step of the Markov chain, such that the covariance used to compute the orthogonal basis operators must be learnt progressively as follows:

$$\mathbf{C}_{n+1} = \frac{n}{n+1} \mathbf{C}_n + \frac{(\boldsymbol{\theta}_n - \bar{\boldsymbol{\theta}})^\dagger (\boldsymbol{\theta}_n - \bar{\boldsymbol{\theta}})}{n}, \tag{8.53}$$

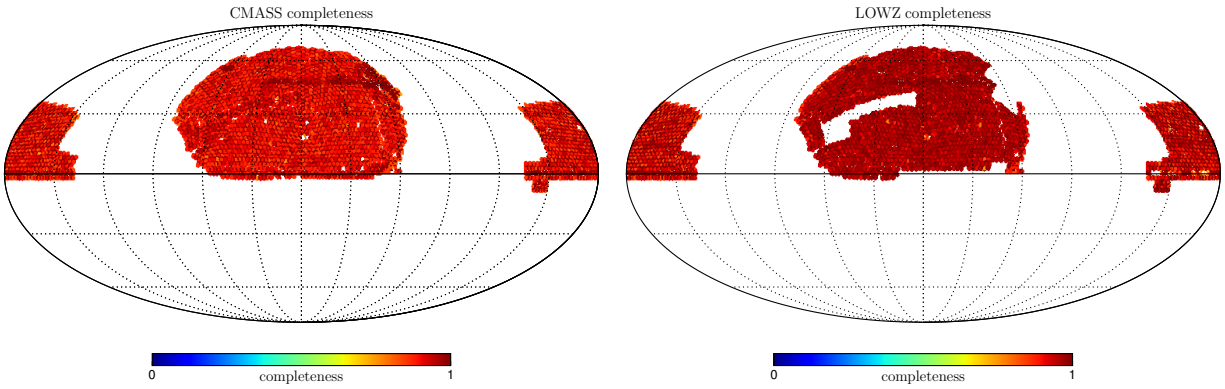


FIGURE 8.5: Observed sky completeness, used to generate and analyze the mock catalogue in this work, are illustrated in the left and right panels, corresponding to the CMASS and LOW-Z components of the SDSS-III survey, respectively.

at the given n^{th} step. The above transformation helps to decorrelate the variables involved when sampling, thereby reducing the burn-in period significantly. This procedure is especially effective with highly correlated parameters, as is the case for $(\Omega_{\rm m}, w_0)$ (cf. Fig. 8.8).

We investigate the statistical efficiency of our cosmological parameter sampler by computing the correlation length of the Markov chain. Subsequent samples in the chain are generally correlated and hence do not qualify as independent samples of the target posterior distribution. The correlation length characterizes the statistical efficiency of generating independent samples of a given parameter θ, as follows:

$$C(\theta)_n = \frac{1}{N-n} \sum_{i=0}^{N-n} \left(\frac{\theta^i - \langle\theta\rangle}{\sqrt{\mathrm{Var}(\theta)}} \frac{\theta^{i+n} - \langle\theta\rangle}{\sqrt{\mathrm{Var}(\theta)}} \right), \tag{8.54}$$

where n is the number of transition steps in the chain, $\langle\theta\rangle$ and $\mathrm{Var}(\theta)$ correspond to the mean and variance, respectively, i.e. $\langle\theta\rangle = 1/N \sum_i \theta^i$ and $\mathrm{Var}(\theta) = 1/N \sum_i (\theta^i - \langle\theta\rangle)^2$, while N is the total number of realizations in the chain. This correlation length therefore indicates the number of independent realizations that can be drawn from the MCMC chain with a given number of transition steps. For an unbiased comparison, we computed the correlation length using the same chain length for both scenarios, while ignoring their respective burn-in phases. From the left panel of Fig. 8.4, we deduce that the correlation length for the $\Omega_{\rm m}$ chain is of the order of 25 samples. Without the rotation of $(\Omega_{\rm m}, w_0)$ parameter space, the correlation length is longer by nearly a factor of five, further demonstrating the gain in efficiency obtained with the more sophisticated sampler.

8.7 Generation of a mock galaxy catalogue

We describe the generation of an artificial galaxy survey using as template the Sloan Digital Sky Survey (SDSS-III), consisting of four galaxy subcatalogues, to validate the methodology described in the previous sections. The procedure implemented here for the mock generation is essentially based on the descriptions provided in Jasche and Kitaura (2010) and Jasche and Wandelt (2013a).

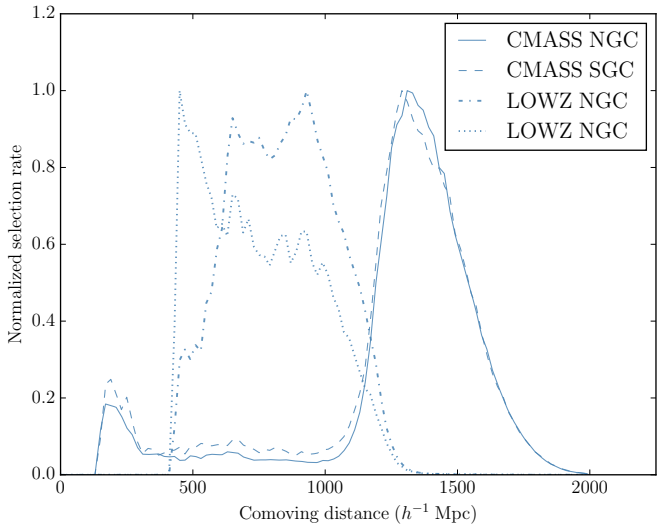


FIGURE 8.6: Radial selection functions for the CMASS sample, in solid and dashed lines for the North Galactic Cap (NGC) and South Galactic Cap (SGC), respectively. The corresponding radial selection functions for the LOW-Z sample are depicted in dash-dotted (NGC) and dotted lines (SGC). These selection functions are used to generate the mock data to emulate features of the actual SDSS-III BOSS data.

At first, we generate a realization for the initial density contrast $\delta_p^{\rm ic}$ from a normal distribution with zero mean and covariance corresponding to an underlying cosmological power spectrum for the matter distribution. This power spectrum, including baryonic wiggles, is computed following the prescription provided in Eisenstein and Hu (1998) and Eisenstein and Hu (1999), assuming a standard Λ cold dark matter (ΛCDM) cosmology with the set of cosmological parameters ($\Omega_{\rm m} = 0.3089$, $\Omega_\Lambda = 0.6911$, $\Omega_{\rm b} = 0.0486$, $h = 0.6774$, $\sigma_8 = 0.8159$, $n_{\rm s} = 0.9667$) from *Planck* (Planck Collaboration et al., 2016b). This yields a 3D Gaussian initial density field in an equidistant Cartesian grid with $N_{\rm side} = 128$, i.e. consisting of 128^3 voxels, where each voxel corresponds to a discretized volume element with a physical voxel size of 31.25 h^{-1} Mpc, and comoving box length of 4000 h^{-1} Mpc. As a result, this implies a total of $\sim 2.1 \times 10^6$ inference parameters, corresponding to the amplitude of the primordial density fluctuations at the respective grid nodes.

To ensure a sufficient resolution of inferred final density fields, we oversample the initial density field by a factor of eight, thereby evaluating the LPT model with 256^3 particles in every sampling step. The following step is to scale this 3D distribution of initial conditions to a cosmological scale factor of $a_{\rm init} = 0.001$ via multiplication with a cosmological growth factor $D^+(a_{\rm init})$. These initial conditions are subsequently evolved forward in time, in non-linear fashion, with LPT providing an approximate description of gravitational LSS formation. We then construct a final 3D non-linearly evolved density field $\delta_p^{\rm f}$ from the resulting particle distribution via the cloud-in-cell (CIC) method (e.g. Hockney and Eastwood, 1988).

To generate the mock galaxy survey, we essentially need to simulate the inhomogeneous Poissonian process described by Eq. (8.2), by drawing random samples from the distribution, on top of the final density field $\delta_p^{\rm f}$. In this work, we generate a mock data set with realistic features emulating the highly structured survey geometry and selection effects of the SDSS-III survey, with the observed sky completeness depicted in the left and right panels of Fig. 8.5, respectively, for the CMASS and

LOW-Z components of the SDSS-III survey. To account for the different selection effects in the northern and southern galactic planes, each component is further divided into two subcatalogues, corresponding to North/South Galactic Caps (NGC/SGC). The respective radial selection functions for these four subcatalogues are illustrated in Fig. 8.6. Here, the selection functions are numerical estimates obtained by binning the corresponding distribution of tracers $N(d)$ in the CMASS and LOW-Z components (e.g. Ross et al., 2017), where d is the comoving distance from the observer.

The projection of the completeness functions into the 3D volume produces the 3D observation mask C_p^g. The survey properties are described by the galaxy selection function and the 3D completeness function, with the product of these two functions yielding the survey response operator R_p^g. More specifically, it is the average of the product of the 2D survey geometry $C^g(\hat{x}) = C^g_{p(\hat{x})}$ and the selection function $f(x)$ at each volume element of the 3D grid:

$$R_p^g = \frac{1}{|\mathcal{V}_p|} \int_{\mathcal{V}_p} \mathrm{d}^3\boldsymbol{x}\, C^g(\hat{x}) f^g(x), \tag{8.55}$$

where the volume occupied by the p^{th} voxel is indicated by $\mathcal{V}_p$, with $|\mathcal{V}|$ the volume of the set $\mathcal{V}$.

Finally, we generate the four artificial galaxy subcatalogues, labelled by g, by Poisson sampling on the grid with $\bar{N}_p^g = \{110.42, 122.94, 71.43, 205.48\}$, resulting in a total number of 997828 galaxies. The values of $\bar{N}_p^g$ are chosen such that the mock catalogue reflects the characteristics of the actual SDSS-III data which contains around one million tracers.

8.8 Results and discussion

We perform a series of tests to evaluate the performance of our algorithm in a realistic context by applying ALTAIR on the simulated SDSS-III-like galaxy catalogue described in the previous section. In particular, the focus is on the burn-in and convergence behaviour of our method, which are key indicators of the overall numerical feasibility and statistical efficiency for real data applications.

8.8.1 Cosmological parameter inference

To validate the conceptual framework for cosmological parameter inference and the robustness of our implementation of the AP test, the reproducibility of the input cosmology and the correlations with the other inferred quantities such as galaxy bias are also of interest.

The Markov chains for the cosmological parameters, displayed in Fig. 8.7, were initialized with an over-dispersed state. This figure consequently illustrates an initial burn-in phase, lasting ∼250 MCMC steps, where the Markov chains follow a persistent drift towards the high probability region of the parameter space. The rotation of the (Ω_m, w_0) parameter space before slice sampling, as described in Section 8.6.3, reduces the burn-in period significantly, by roughly a factor of five, as shown in the right panel of Fig. 8.4, resulting in improved sampling efficiency.

The corresponding marginal and joint posterior distributions for the cosmological parameters are displayed in Fig. 8.8, demonstrating the capability of ALTAIR to infer tight constraints from

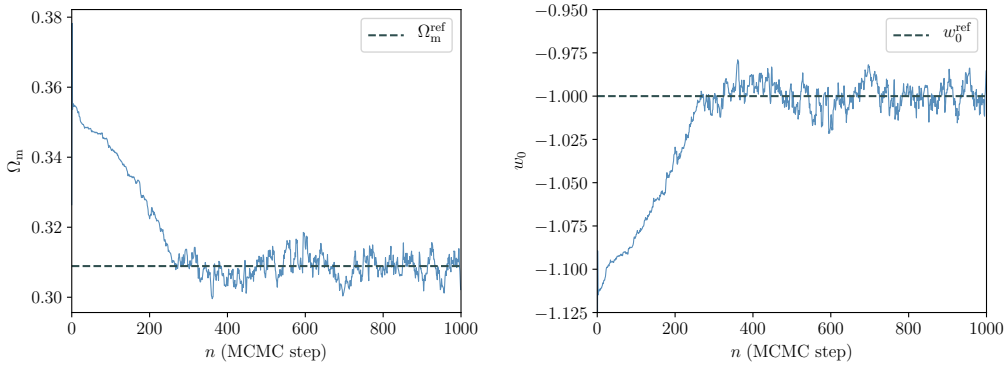


FIGURE 8.7: MCMC chains for the cosmological parameters, for the first 1000 samples, with the reference cosmology employed in the mock generation indicated by the horizontal dashed lines. An initial burn-in phase lasting ∼250 Markov transitions is illustrated by the coherent drift of the Markov chain towards the preferred region in parameter space.

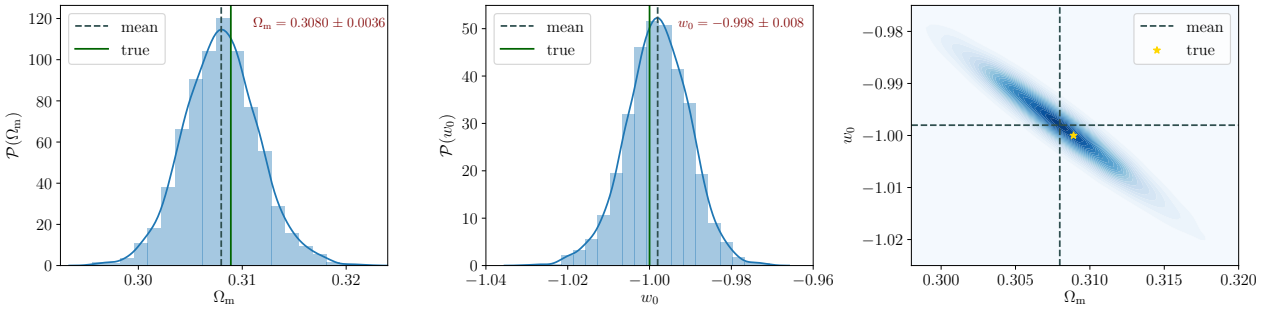


FIGURE 8.8: Marginal posteriors for Ω_m (*left panel*) and the dark energy equation of state, w_0 (*middle panel*), for ∼3000 MCMC realizations, ignoring the burn-in phase of ∼250 Markov steps. The corresponding mean and standard deviation for each parameter are indicated in the top right corner of each plot. The joint posterior (*right panel*) for Ω_m and w_0, depicting the high level of correlation between these two parameters. The highly informative distortion due to the cosmic expansion, as a result of probing a deep redshift range, yields extremely tight constraints on the above cosmological parameters. As a consistency test, this validates our implementation of the AP test to correctly recover the input cosmology.

galaxy redshift surveys. This robust AP test fully exploits the high information content from the cosmic expansion as a result of probing a deep redshift range, where the distortion is more pronounced, yielding the following cosmological constraints: $\Omega_m = 0.3080 \pm 0.0036$ and $w_0 = -0.998 \pm 0.008$. As a comparison, the SDSS-III (DR12, BAO + *Planck*) constraints are as follows: $\Omega_m = 0.310 \pm 0.005$ and $w_0 = -1.01 \pm 0.06$ (Alam et al., 2017), further highlighting the significant constraining power of our AP test. We acknowledge the significant difference in the size of uncertainties. A back of the envelope computation of the information gain is as follows: Considering a sphere of 100 h^{-1} Mpc for BAO against all voxels in 4000 h^{-1} Mpc, $N_{BAO} = 4000^3/(4\pi \times 100^3/3) = 15278$, compared to $N_{vox} = 128^3$, yields an improvement of $\sqrt{N_{vox}/N_{BAO}} = 11.7$, which provides an order of magnitude of our uncertainties on the cosmological parameters. This is an approximate attempt to quantify the information gain from including smaller scales (in our work, $\sim 0.17\ h^{-1}$ Mpc) than the BAO scale by essentially counting the number of modes. However, the above argument does not imply that employing finer resolutions will result in an infinite gain of information. There is a saturation of information at a certain resolution due to the slow variation in the density fields across

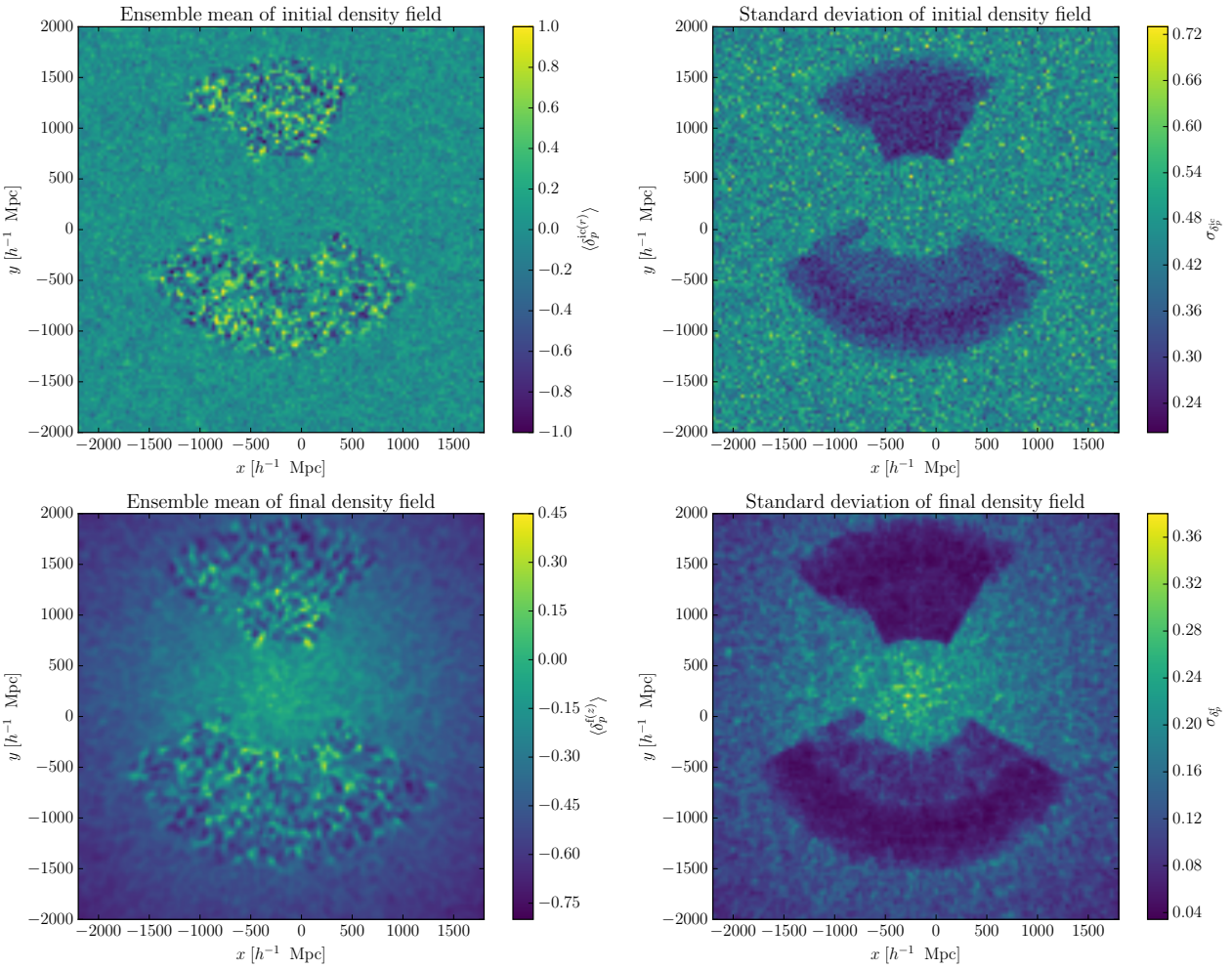


FIGURE 8.9: Mean and standard deviation maps for the initial (*top panel*) and final density fields (*bottom panel*), computed from the MCMC realizations, with the same slice through the 3D density fields being illustrated above. In unobserved or masked regions, the density fields are not constrained by data, and they average out to the cosmic mean density. However, in observed regions, the Gaussian nature of the initial conditions and the filamentary nature of the non-linearly evolved density field is manifest. The corresponding variance is therefore higher in regions devoid of data.

neighbouring voxels, such that further refinement beyond this limit will not yield any additional information.

To verify that the cosmological information stems purely from the geometric distortion due to the cosmic expansion, i.e. the AP test, we perform the following experiment: We deactivate the cosmic expansion component in our forward model and sample the cosmological parameters using only the LPT evolved density field. We consequently recover the corresponding prior distributions for the marginal posteriors of the cosmological parameters. As a result, this implies that the information derives purely from the geometry and not from the clustering of the non-linearly evolved density field, at least for the test case with the physical voxel size considered in this work.

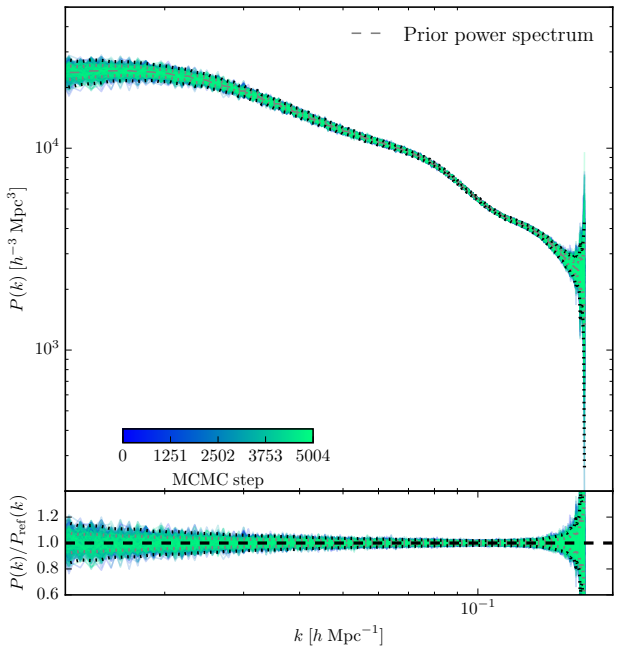


FIGURE 8.10: *Top panel:* Reconstructed power spectra from the inferred posterior initial density field realizations for all sampling steps of the Markov chain. The MCMC samples are colour-coded according to their sample number, as indicated by the colour bar. This is a self-consistency test to verify whether the sampled density fields are in accordance with the reference power spectrum, depicted in dashed lines, employed in the mock generation. *Bottom panel:* The deviation from the true underlying solution, with the thick dashed lines representing the prior power spectrum. The thin grey and and black dotted lines correspond to the Gaussian 1σ and 2σ limits, respectively.

8.8.2 Inference of 3D initial conditions and evolved matter fields

The mean initial and final density fields computed from $\sim$4000 realizations (ignoring the burn-in phase) are illustrated in the left panel of Fig. 8.9 for a particular slice of the 3D field. The maps reveal that on average we can recover highly detailed and well-defined structures in the observed regions. In particular, the filamentary nature of the non-linearly evolved density field can be clearly seen, while the Gaussian nature of the initial conditions can also be deduced visually. However, in poorly or not observed regions, the ensemble mean density field approaches the cosmic mean density, as expected, since regions lacking any observational information should on average reflect the cosmic mean. The uncertainty in these regions is accurately accounted for in the inference process, as demonstrated by the right panel, since each sampled density field is a constrained realization, i.e. these regions are augmented with statistically correct information.

8.8.3 Cosmological power spectra reconstruction

In the top panel of Fig. 8.10, we illustrate the power spectra reconstructed from all the realizations of 3D initial density field, obtained from the posterior via the HMC sampler. This is a self-consistency test to confirm that the sampled density fields are in agreement with the reference (prior) power spectrum adopted for the mock generation, as substantiated quantitatively in the bottom panel, where the ratio of the *a posteriori* power spectra to the prior distribution is shown. The measured power spectra therefore demonstrate that the individual realizations possess the correct power

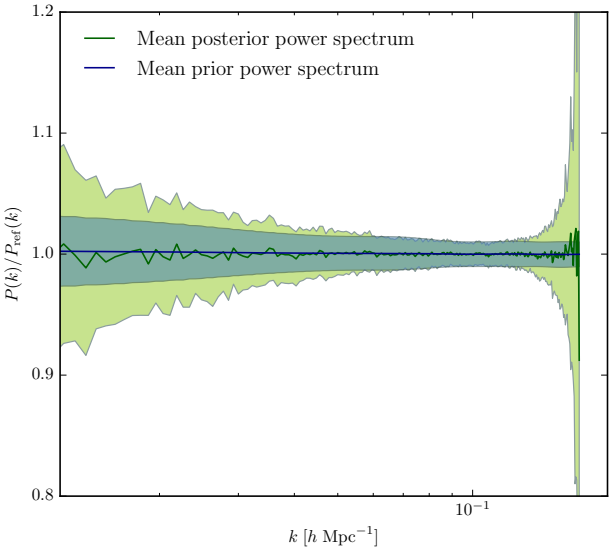


FIGURE 8.11: Summary statistics of the reconstructed power spectra from the inferred posterior initial density field realizations depicted in Fig. 8.10, with the ensemble mean indicated by a solid green line. The solid blue line corresponds to the ensemble mean of the power spectra realizations generated using the inferred cosmological parameters. The shaded regions indicate their respective 1σ confidence region, i.e. 68% probability volume. This plot shows that the prior information entropy is inferior to the posterior information entropy, due to the narrower distribution of the former. The prior power spectrum adopted, as a result, does not impact significantly on the cosmological parameter inference via the AP test.

throughout the entire domain of Fourier modes considered in this work. Moreover, they do not exhibit any spurious power artefacts typically resulting from erroneous treatment of survey characteristics, such as survey geometry and selection effects, and galaxy biases, implying that such effects have been properly accounted for.

In order to verify the impact of the prior power spectrum on the actual inference of cosmological parameters, we illustrate, in Fig. 8.11, the distributions of power spectra computed using the inferred cosmology and the prior analytic prescription (Eisenstein and Hu, 1998; Eisenstein and Hu, 1999) and the reconstructed power spectra from Fig. 8.10, via their respective summary statistics, normalized with respect to the fiducial power spectrum. The scatter in the latter *a posteriori* power spectra reconstructed from the sampled density field realizations is significantly higher than distribution of the former prior spectra. This implies that the entropy of prior information is much lower than that of posterior information, thereby demonstrating that the prior power spectrum does not influence the cosmological parameter inference via the AP test and justifying the assumption made in Section 8.3.3.

8.8.4 Robustness to model misspecification

We further demonstrate this robustness of our AP test by employing a modified prior power spectrum in the inference procedure. By adopting a different cosmology ($\Omega_{\rm m} = 0.40$, $w_0 = -0.85$), we modify the shape of the power spectrum, and subsequently apply ALTAIR on the same mock catalogue from Section 8.7. As shown in Fig. 8.12, we recover the fiducial cosmological parameters employed in the mock generation, although with slightly larger uncertainties than for the original

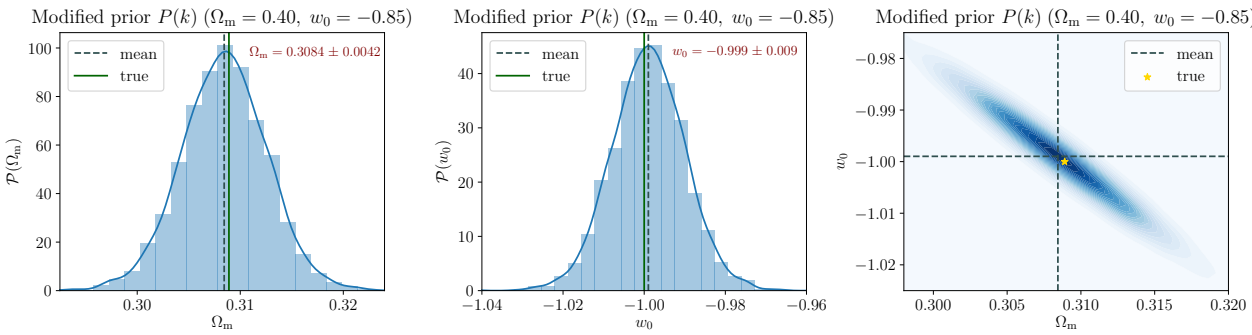


FIGURE 8.12: Same as Fig. 8.8, but employing a different prior power spectrum ($\Omega_m = 0.40$, $w_0 = -0.85$), for ∼3000 MCMC realizations. By recovering the fiducial cosmological parameters employed in the mock generation, this test case explicitly highlights the robustness of our approach to the shape of the prior power spectrum adopted. The corresponding uncertainties are slightly larger than for the original run by around 15%.

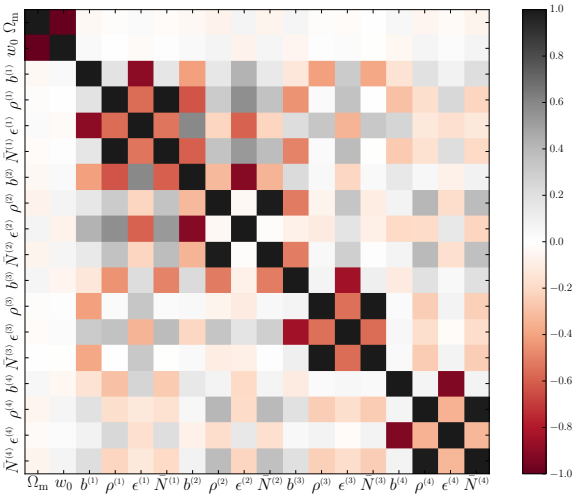


FIGURE 8.13: Correlation matrix of the galaxy bias and cosmological parameters, normalized using their respective variance. This illustrates the weak correlation between the inferred cosmological constraints and the galaxy bias. The lack of dependence on the currently unknown phenomenon of galaxy biasing is therefore a key highlight of our implementation of the AP test for cosmological parameter inference.

run by roughly 15%. This test case therefore explicitly highlights the robustness of our implementation of the AP test to a misspecified model since it does not optimize the information from the scale dependence of the correlations of the density field, but rather from the isotropy of the field.

8.8.5 Resilience of cosmological constraints to galaxy bias

From the correlation matrix of Ω_m, w_0 and the galaxy bias parameters for each of the four subcatalogues, illustrated in Fig. 8.13, we deduce the extremely weak correlation between the cosmological constraints and the bias. This is a key positive aspect of our method, as galaxy bias remains nevertheless a highly active and challenging field of research (see, for e.g., Desjacques, Jeong, and Schmidt, 2016), due to its complex non-linear behaviour on intermediate and small scales, which

may potentially limit the effectiveness of traditional methods of cosmological parameter inference (Pollina et al., 2018).

Moreover, this insensitivity to the galaxy bias implies that our method does not rely on absolute density fluctuation amplitudes, but on the actual location of matter. This entails that our AP test exploits the geometrical structure of the density field and not its absolute amplitude since the power spectrum does not influence the inferred cosmological constraints. To the best of our knowledge, this is a novel aspect of cosmological inference, with most of our current understanding of cosmology based on measurements of density contrast amplitudes. We present, therefore, one of the first methods which extracts a large fraction of information from statistics other than that of direct density contrast correlations, without relying on the power spectrum or bispectrum. Our method consequently yields complementary information to state-of-the-art methods.

8.9 Summary and conclusions

We presented the implementation of a robust AP test that performs a detailed fit of the cosmological expansion via a non-linear and hierarchical Bayesian LSS inference framework. This forward modelling approach employs LPT as a physical description for the non-linear dynamics and sequentially encodes the cosmic expansion effect for joint inference of the cosmological parameters and underlying 3D density fields, while also fitting the mean density of tracers and bias parameters. In essence, this inference machinery explores the various cosmological expansion histories and selects the cosmology-dependent evolution pathways which yield isotropic correlations of the galaxy density field, thereby constraining cosmology.

We demonstrated the application of our algorithm ALTAIR on an artificially generated galaxy catalogue, consisting of four subcatalogues, that emulates the highly structured survey geometry and selection effects of SDSS-III. We performed a series of statistical efficiency and consistency tests to validate the methodology adopted and showcased its potential to yield tight constraints on cosmological parameters from current and future galaxy redshift surveys. The main strength of our implementation of the AP test lies in its robustness to a misspecified model and its inherent approximations, thereby near-optimally exploiting the model predictions, without relying on its accuracy in modelling the scale dependence of the correlations of the field.

Moreover, another key aspect of our approach, resulting from the robustness to a misspecified model, is that the cosmological constraints show extremely weak dependence on galaxy bias. This yields two crucial advantages. First, this is especially interesting as the lack of a sufficient physical description of this bias remains a potential limiting factor for standard approaches of cosmological parameter inference from such redshift surveys. Furthermore, this lack of sensitivity to the bias also implies that our method does not depend on the absolute density fluctuation amplitudes. This is therefore among the first methods to extract a large amount of information from statistics other than that of direct density contrast correlations, without relying on the power spectrum or bispectrum, thereby providing complementary information to state-of-the-art techniques.

There is scope for further development of the ALTAIR framework, such as incorporating power spectrum inference, which is a highly non-trivial undertaking. We also intend to augment the current formalism to include the treatment of redshift space distortions, which is key for unbiased constraints on the cosmological parameters, and apply ALTAIR on state-of-the-art galaxy redshift catalogues for cosmological inference.

Chapter 9

Explicit Bayesian treatment of unknown foreground contaminations

The work presented in this chapter is based on Porqueres, Kodi Ramanah, Jasche, and Lavaux (2019).

9.1 Introduction

The next generation of galaxy surveys, such as Large Synoptic Survey Telescope (LSST) (Ivezic et al., 2008) or Euclid (Laureijs et al., 2011; Amendola et al., 2016; Racca et al., 2016), will not be limited by noise but by systematic effects. In particular, deep photometric observations will be subject to several foreground and target contamination effects, such as dust extinction, stars and seeing (e.g. Scranton et al., 2002; Ross et al., 2011; Ho et al., 2012; Huterer, Cunha, and Fang, 2013; Ho et al., 2015).

In the past, such effects have been addressed by generating templates for such contaminations and accounting for their overall template coefficients within a Bayesian framework. Leistedt and Peiris (2014), for instance, performed an extensive effort in compiling a total set of 220 foreground contaminations for the inference of the clustering signal of quasars in the Sloan Digital Sky Survey (SDSS-III) Baryon Oscillation Spectroscopic Survey (BOSS) (Bovy et al., 2012). Foreground contaminations are also dealt with in observations of the cosmic microwave background, where they are assumed to be an additive contribution to observed temperature fluctuations (e.g. Tegmark and Efstathiou, 1996; Tegmark et al., 1998; Hinshaw et al., 2007; Eriksen et al., 2008; Ho et al., 2015; Vansyngel et al., 2016; Sudevan et al., 2017; Elsner, Leistedt, and Peiris, 2017). In the context of large-scale structure analyses, Jasche and Lavaux (2017) presented a foreground sampling approach to account for multiplicative foreground effects which can affect the target and the number of observed objects across the sky.

All these methods rely on a sufficiently precise estimate of the map of expected foreground contaminants to be able to account for them in the statistical analysis. These approaches exploit the fact that the spatial and spectral dependence of the phenomena generating these foregrounds are well-known. But what if we are facing unknown foreground contaminations? Can we make progress in robustly recovering cosmological information from surveys subject to yet unknown contaminations? In this work, we describe an attempt to address these questions and develop an optimal and robust likelihood to deal with such effects. The capability to account for "unknown unknowns" is

FIGURE 9.1: Schematic to illustrate the colour indexing of the survey elements. Colours are assigned to voxels according to patches of a given angular scale. Voxels of the same colour belong to the same patch, and this colour indexing is subsequently employed in the computation of the robust likelihood.

also the primary motivation behind the blind method for the visibility mask reconstruction recently proposed by Monaco, Di Dio, and Sefusatti (2019).

This chapter is organized as follows. We outline the underlying principles of our novel likelihood in Section 9.2, followed by a description of the numerical implementation in Section 9.3. We subsequently assess the performance of our proposed likelihood via a comparison with a standard Poissonian likelihood in Section 9.5, with the specific problem considered illustrated in the preceding Section 9.4. The key aspects of our findings are finally summarized in Section 9.6.

9.2 Robust likelihood model

We describe the conceptual framework for the development of the robust likelihood which constitutes the crux of this work. The standard analysis of galaxy surveys assumes that the distribution of galaxies can be described as an inhomogeneous Poisson process (Layzer, 1956b; Peebles, 1980; Martínez and Saar, 2003) given by

$$\mathcal{P}(N|\lambda) = \prod_i \frac{e^{-\lambda_i}(\lambda_i)^{N_i}}{N_i!}, \tag{9.1}$$

where N_i is the observed number of galaxies at a given position in the sky i and λ_i is the expected number of galaxies at that position. The expected number of galaxies is related to the underlying dark matter density field ρ via

$$\lambda = S\bar{N}\rho^b \exp(-\rho_g \rho^{-\epsilon}), \tag{9.2}$$

where S encodes the selection function and geometry of the survey, $\bar{N}$ is the mean number of galaxies in the volume, and $\{b, \rho_g, \epsilon\}$ are the parameters of the non-linear bias model proposed by Neyrinck et al. (2014).

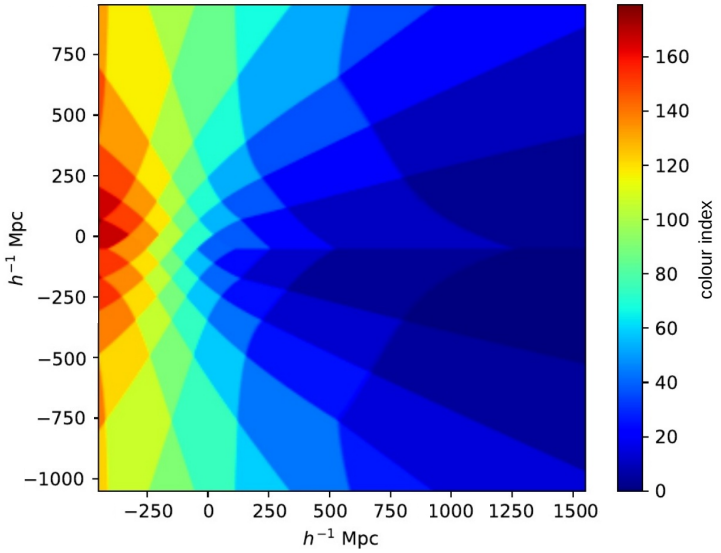


FIGURE 9.2: Slice through the 3D coloured box. The extrusion of the colour indexing scheme (cf. Fig. 9.1) onto a 3D grid yields a collection of patches, denoted by a given colour, with a group of voxels belonging to a particular patch, to be employed in the computation of the robust likelihood. The axes indicate the comoving distances to the observer, who is located at the origin (0,0,0).

The key contribution of this work is to develop a more robust likelihood than the standard Poissonian likelihood by marginalizing over the unknown large-scale foreground contamination amplitudes. We start with the assumption that there is a large-scale foreground modulation that can be considered to have a constant amplitude over a particular group of voxels. Assuming that A is the amplitude of this large-scale perturbation, we can write $\lambda_\alpha = A\bar{\lambda}_\alpha$, where the index α labels the voxels over which the perturbation is assumed to have constant amplitude. The likelihood consequently has the following form:

$$\mathcal{P}(N|\bar{\lambda}, A) = \prod_\alpha \frac{e^{-A\bar{\lambda}_\alpha} A^{N_\alpha} (\bar{\lambda}_\alpha)^{N_\alpha}}{N_\alpha!} \tag{9.3}$$

$$= e^{-A\sum_\alpha \bar{\lambda}_\alpha} A^{\sum_\alpha N_\alpha} \prod_\alpha \frac{(\bar{\lambda}_\alpha)^{N_\alpha}}{N_\alpha!}. \tag{9.4}$$

We can marginalize over the unknown foreground amplitude A as follows:

$$\mathcal{P}(N|\bar{\lambda}) = \int \mathrm{d}A\, \mathcal{P}(N, A|\bar{\lambda}) \tag{9.5}$$

$$= \int \mathrm{d}A\, \mathcal{P}(A|\bar{\lambda})\, \mathcal{P}(N|A, \bar{\lambda}) \tag{9.6}$$

$$= \int \mathrm{d}A\, \mathcal{P}(A)\, \mathcal{P}(N|A, \bar{\lambda}), \tag{9.7}$$

where, in the last step, we assumed conditional independence, $\mathcal{P}(A|\bar{\lambda}) = \mathcal{P}(A)$. This assumption is justified since the processes which generate the foregrounds are expected to be independent of the mechanisms involved in galaxy formation. As a result of this marginalization over the amplitude A, and using a power-law prior for A, $\mathcal{P}(A) = \kappa A^{-\gamma}$ where γ is the power-law exponent and κ is

an arbitrary constant, the likelihood simplifies to:

$$\mathcal{P}(N|\bar{\lambda}) = \kappa \frac{\left(\sum_\alpha N_\alpha\right)!}{\left(\sum_\beta \bar{\lambda}_\beta\right)^{\sum_\alpha N_\alpha + 1 - \gamma}} \prod_\alpha \frac{(\bar{\lambda}_\alpha)^{N_\alpha}}{N_\alpha!} \tag{9.8}$$

$$\propto \frac{1}{\left(\sum_\beta \bar{\lambda}_\beta\right)^{1-\gamma}} \prod_\alpha \left(\frac{\bar{\lambda}_\alpha}{\sum_\beta \bar{\lambda}_\beta}\right)^{N_\alpha}. \tag{9.9}$$

We employ a Jeffreys prior for the foreground amplitude A, which implies setting $\gamma = 1$. Jeffrey's prior is a solution to a measure invariant scale transformation (Jeffreys, 1946) and is therefore a scale independent prior, such that different scales have the same probability and there is no preferred scale. This scale invariant prior is optimal for inference problems involving scale measurements as this does not introduce any bias on a logarithmic scale. Moreover, this is especially interesting because this allows for a total cancellation of unknown amplitudes in Eq. (9.9), resulting in the following simplified form of our augmented likelihood:

$$\mathcal{P}(N|\bar{\lambda}) \propto \prod_\alpha \left(\frac{\bar{\lambda}_\alpha}{\sum_\beta \bar{\lambda}_\beta}\right)^{N_\alpha}. \tag{9.10}$$

9.3 Numerical implementation

9.3.1 Modified formalism for HMC sampling

We implement the robust likelihood in BORG (Bayesian Origin Reconstruction from Galaxies, Jasche and Wandelt, 2013a), a hierarchical Bayesian inference framework for the non-linear inference of large-scale structures. It encodes a physical description for non-linear dynamics via Lagrangian Perturbation Theory (LPT), resulting in a highly non-trivial Bayesian inverse problem. At the core, it employs a Hamiltonian Monte Carlo (HMC) method for the efficient sampling of a high-dimensional and non-linear parameter space of possible initial conditions at an earlier epoch, with typically $\mathcal{O}(10^7)$ free parameters, corresponding to the discretized volume elements of the observed domain. The HMC implementation is detailed in Jasche and Kitaura (2010) and Jasche and Wandelt (2013a). The essence of BORG is that it incorporates the joint inference of initial conditions, and consequently the corresponding non-linearly evolved density fields and associated velocity fields, from incomplete observations. An augmented variant, BORG-PM, employing a particle mesh model for gravitational structure formation, has recently been presented (Jasche and Lavaux, 2019). An extension to BORG has also been developed to constrain cosmological parameters via a novel application of the Alcock-Paczyński test (Kodi Ramanah et al., 2019).

For the implementation of the robust likelihood, the HMC method that constitutes the basis of the joint sampling framework requires the negative log-likelihood and its adjoint gradient, which

are given by

$$\begin{aligned}\Psi &\equiv -\log \mathcal{P}(N|\bar{\lambda}) \\ &= \sum_{\alpha} N_{\alpha} \log\left(\sum_{\beta} \bar{\lambda}_{\beta}\right) - \sum_{\alpha} N_{\alpha} \log \bar{\lambda}_{\alpha}\end{aligned} \tag{9.11}$$

and

$$\frac{\partial \Psi}{\partial \bar{\lambda}_{\gamma}} \frac{\partial \bar{\lambda}_{\gamma}}{\partial \rho} = \frac{\bar{\lambda}_{\gamma}}{\rho}\left(b + \epsilon \rho_g \rho^{-\epsilon}\right)\left[\frac{\sum_{\alpha} N_{\alpha}}{\sum_{\beta} \bar{\lambda}_{\beta}} - \frac{N_{\gamma}}{\bar{\lambda}_{\gamma}}\right]. \tag{9.12}$$

9.3.2 Colour indexing scheme

The labelling of voxels with the same foreground modulation is encoded via a colour indexing scheme that groups the voxels into a collection of angular patches. This requires the construction of a sky map which is divided into regions of a given angular scale, where each region is denoted by specific colour and stored in HEALPIX format (Górski et al., 2005), as illustrated in Fig. 9.1. An extrusion of the sky map onto a 3D grid subsequently yields a 3D distribution of patches, with a particular slice of this 3D coloured grid displayed in Fig. 9.2. The collection of voxels belonging to a particular patch is employed in the computation of the robust likelihood given by Eq. (9.11), where α corresponds to the colour index.

This is a maximally ignorant approach to deal with unknown systematics where we enforce that every modulation above a given angular scale is not known. Since the colouring scheme does not depend on any foreground information, the numerical implementation of the likelihood is therefore generic. Moreover, another advantage of our approach is that the other components in our forward modelling scheme do not require any adjustments to encode this data model. However, we have not considered additive contaminations typically emanating from stars. We defer the extension of our data model to account for such additive contaminants to a future investigation.

9.4 Mock generation with foreground contaminants

We provide a brief description of the generation of the mock data set, that will be used to test the effectiveness of our novel likelihood, essentially based on the procedure adopted in Jasche and Kitaura (2010) and Jasche and Wandelt (2013a). We first generate a realization for the initial density contrast δ_k^{i} from a zero-mean normal distribution with covariance corresponding to the cosmological power spectrum, such that we have a 3D Gaussian initial density field in a cubic equidistant grid with $N_{\mathrm{side}} = 256$, consisting of 256^3 voxels, where each voxel corresponds to a discretized volume element, and comoving box length of $2000h^{-1}$ Mpc. This 3D distribution of initial conditions must then be scaled to a cosmological scale factor of $a_{\mathrm{init}} = 0.001$ using a cosmological growth factor $D^{+}(a_{\mathrm{init}})$.

The underlying cosmological power spectrum, including baryonic acoustic oscillations, for the matter distribution is computed using the prescription described in Eisenstein and Hu (1998) and Eisenstein and Hu (1999). We assume a standard Λ cold dark matter (ΛCDM) cosmology with the set of cosmological parameters ($\Omega_{\mathrm{m}} = 0.3089$, $\Omega_{\Lambda} = 0.6911$, $\Omega_{\mathrm{b}} = 0.0486$, $h = 0.6774$, $\sigma_8 =$

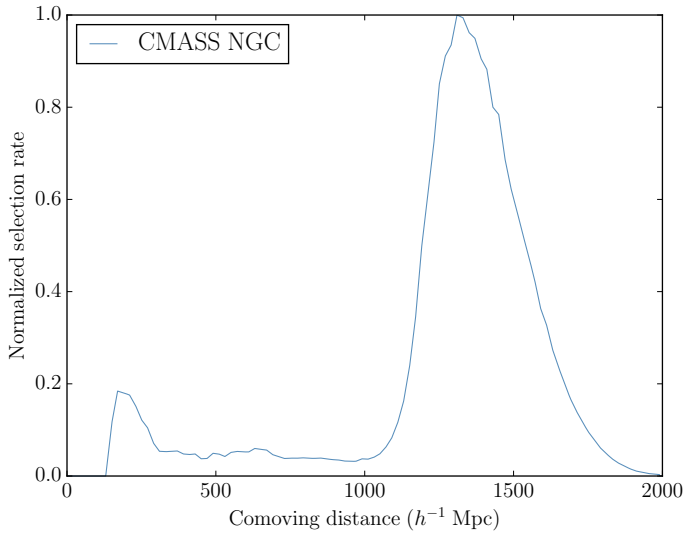


FIGURE 9.3: Radial selection function for the CMASS (north galactic cap) survey which is used to generate the mock data to emulate features of the actual SDSS-III BOSS data.

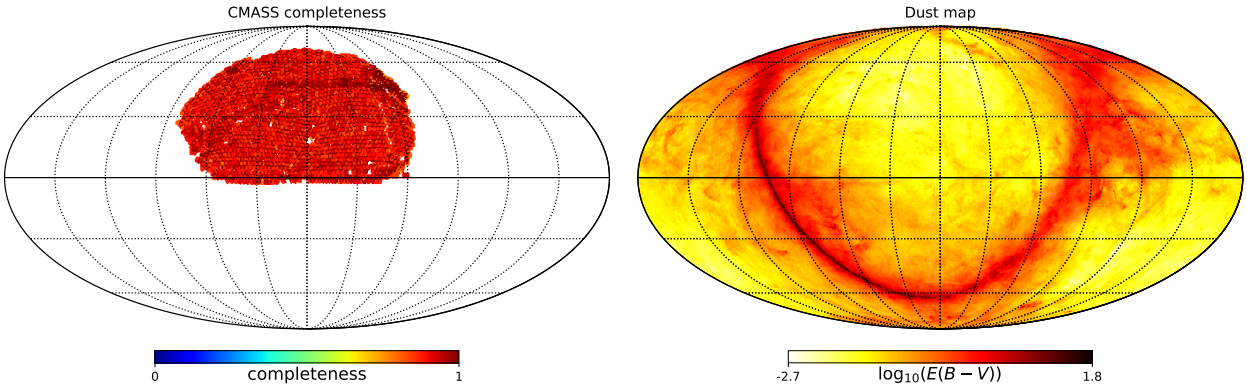


FIGURE 9.4: Observed sky completeness (*left panel*) of the CMASS component of the SDSS-III survey for the north galactic cap and dust extinction map (*right panel*) used to generate the large-scale contamination. This reddening map has been generated from the SFD maps (Schlegel, Finkbeiner, and Davis, 1998).

0.8159, $n_\mathrm{s} = 0.9667$) from Planck Collaboration et al. (2016b). We then employ LPT to transform the initial conditions into a non-linearly evolved field δ_k^f at redshift $z = 0$, which is subsequently constructed from the resulting particle distribution via the cloud-in-cell (CIC) method (e.g. Hockney and Eastwood, 1988).

Given the final density field δ_k^f, we generate a mock galaxy redshift catalogue subject to foreground contamination. For the test case considered in this work, we generate a data set that emulates the characteristics of the SDSS-III survey, in particular the highly structured survey geometry and selection effects. We use a numerical estimate of the radial selection function of the CMASS component of the SDSS-III survey, shown in Fig. 9.3, obtained by binning the corresponding distribution of tracers $N(d)$ in the CMASS sample (e.g. Ross et al., 2017), where d is the comoving distance from the observer. The CMASS radial selection function is therefore estimated from a histogram of

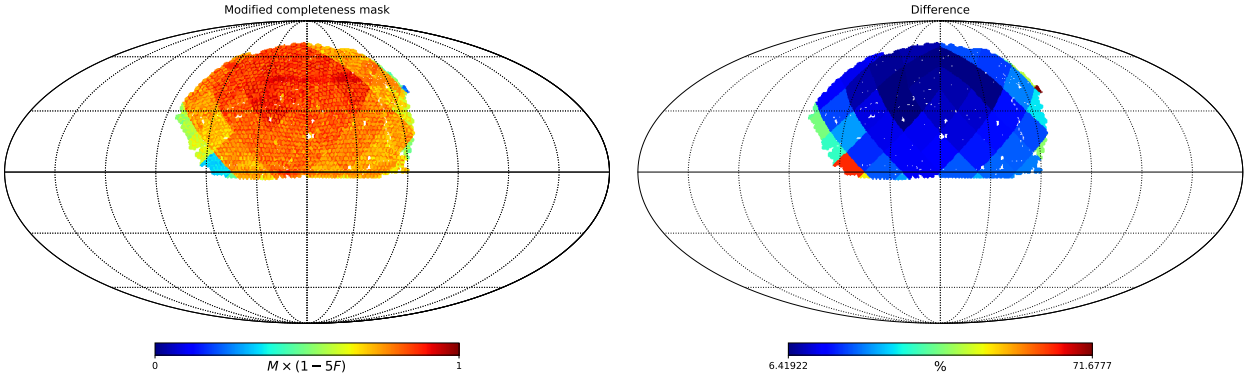


FIGURE 9.5: Contaminated completeness mask (*left panel*) and percentage difference compared to the original completeness mask (*right panel*). The contamination is introduced by multiplying the original mask by a factor of $(1 - 5F)$ where F is a foreground template, in this case, the dust extinction map downgraded to the angular resolution of the colour indexing map depicted in Fig. 9.1. The factor $\alpha = 5$ is chosen such that the mean contamination is 15%, an arbitrary choice to ensure that the contaminations are significant in the completeness mask. The difference between the original and contaminated mask shows that the effect is stronger on the edges of the survey.

galaxy distribution over redshift. The procedure to construct the CMASS sky completeness is, however, less trivial. We derive this CMASS mask, depicted in the left panel of Fig. 9.4, from the SDSS-III BOSS Data Release 12 (Alam et al., 2015) database by taking the ratio of spectroscopically confirmed galaxies to the target galaxies in each polygon from the mask.

In order to emulate a large-scale foreground contamination, we construct a reddening map that describes dust extinction, illustrated in the right panel of Fig. 9.4. This dust template is derived from the data provided by Schlegel, Finkbeiner, and Davis (1998) via straightforward interpolation, rendered in HEALPIX format (Górski et al., 2005).[1] The contamination is produced by multiplying the completeness mask of CMASS, shown in the left panel of Fig. 9.4, by a factor of $(1 - \eta F)$, where F is the foreground template rescaled to the angular resolution of the colour indexing scheme, and η controls the amplitude of this contamination. To obtain a mean contamination of 15% in the completeness, we chose $\eta = 5$ as an arbitrary choice to ensure that the foreground contaminations are significant. This mean value corresponds to the average contamination per element of the sky completeness. Fig. 9.5 shows the contaminated sky completeness and the percentage difference, with the edges of the survey being more affected by the contamination due to their proximity to the galactic plane where the dust is more abundant. The mock catalogue is produced by drawing random samples from the inhomogeneous Poissonian distribution described by Eq. (9.1) and using the modified completeness.

9.5 Results and discussion

In this section, we discuss results obtained by applying the BORG algorithm with the robust likelihood to contaminated mock data. We also compare the performance of our novel likelihood with that of the standard Poissonian likelihood typically employed in large-scale structure analyses. In

[1] The construction of this template is described in more depth in Section 3 of Jasche and Lavaux (2017).

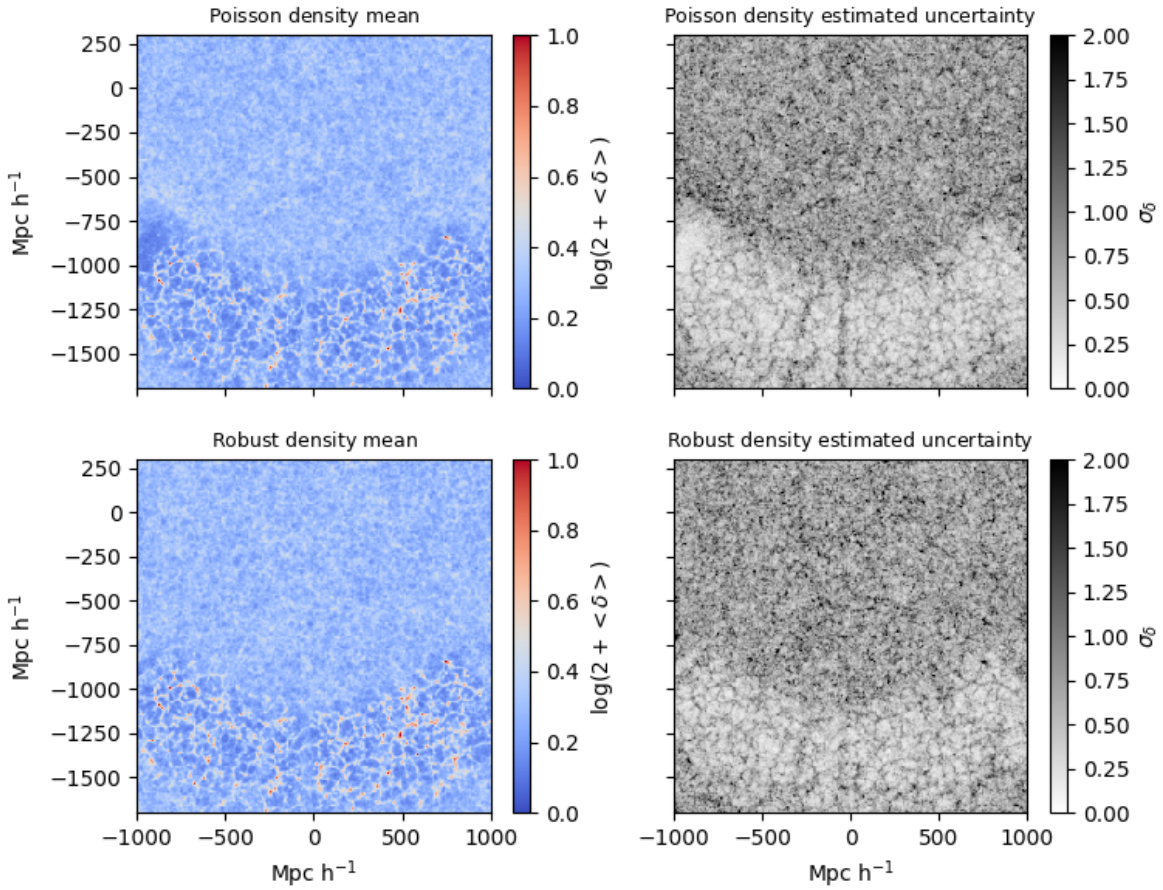
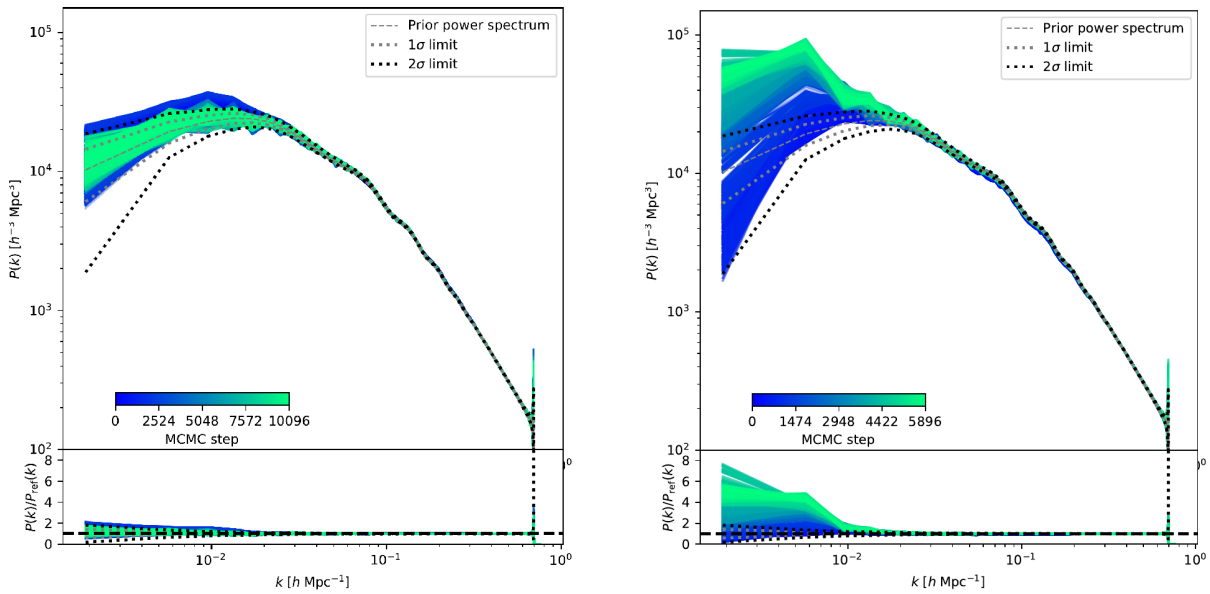


FIGURE 9.6: Mean and standard deviation of the inferred non-linearly evolved density fields, computed from the MCMC realizations, with the same slice through the 3D fields being depicted above for both the Poissonian (*upper panels*) and augmented (*lower panels*) likelihoods. The filamentary nature of the non-linearly evolved density field can be observed in the regions constrained by the data, with the unobserved or masked regions displaying larger uncertainty, as expected. Unlike our robust data model, the standard Poissonian analysis yields some artefacts in the reconstructed density field, particularly near the edges of the survey, where the foreground contamination is stronger.

order to test the effectiveness of our likelihood against unknown systematics and foreground contaminations, the algorithm is agnostic about the contamination and assumes the CMASS sky completeness depicted in the left panel of Fig. 9.4.

9.5.1 Non-linear density field inference

We first study the impact of the large-scale contamination on the inferred non-linearly evolved density field. To this end, we compare the ensemble mean density fields and corresponding standard deviations for the two Markov chains with the Poissonian and novel likelihoods, respectively, illustrated in the top and bottom panels of Fig. 9.6, for a particular slice of the 3D density field. As can be deduced from the top left panel of Fig. 9.6, the standard Poissonian analysis results in spurious effects in the density field, particularly close to the boundaries of the survey since these are the regions that are the most affected by the dust contamination. In contrast, our novel likelihood



(A) Robust likelihood

(B) Standard Poissonian likelihood

FIGURE 9.7: Reconstructed power spectra from the inferred initial conditions from a BORG analysis with unknown foreground contamination for the robust likelihood (*left panel*) and the Poissonian likelihood (*right panel*) over the full range of Fourier modes considered in this work. The σ limit corresponds to the cosmic variance $\sigma = \sqrt{1/k}$. The colour scale shows the evolution of the power spectrum with the sample number. The power spectra of the individual realizations, after the initial burn-in phase, from the robust likelihood analysis possess the correct power across all scales considered, demonstrating that the foregrounds have been properly accounted for. In contrast, the standard Poissonian analysis exhibits spurious power artefacts due to the unknown foreground contaminations, yielding excessive power on these scales.

analysis yields a homogeneous density distribution through the entire observed domain, with the filamentary nature of the present-day density field clearly seen. While we can recover well-defined structures in the observed regions, the ensemble mean density field tends towards the cosmic mean density in the masked or poorly observed regions, with the corresponding standard deviation being higher to reflect the larger uncertainty in these regions. From this visual comparison, it is evident that our novel likelihood is more robust against unknown large-scale contaminations.

9.5.2 Cosmological power spectra reconstruction

From the realizations of our inferred 3D initial density field, we can reconstruct the corresponding matter power spectra and compare them to the prior cosmological power spectrum adopted for the mock generation. The top panel of Fig. 9.7 illustrates the inferred power spectra for both likelihood analyses, with the bottom panel displaying the ratio of the *a posteriori* power spectra to the prior power spectrum. While the standard Poissonian analysis yields excessive power on the large scales due to the artefacts in the inferred density field, the analysis with our novel likelihood allows us to recover an unbiased power spectrum across the full range of Fourier modes.

In addition, we tested the combined effects of the foreground and unknown noise amplitudes by estimating the covariance matrix of the Fourier amplitudes of the reconstructed power spectra. As depicted in Fig. 9.8, our novel likelihood exhibits uncorrelated amplitudes of the Fourier modes,

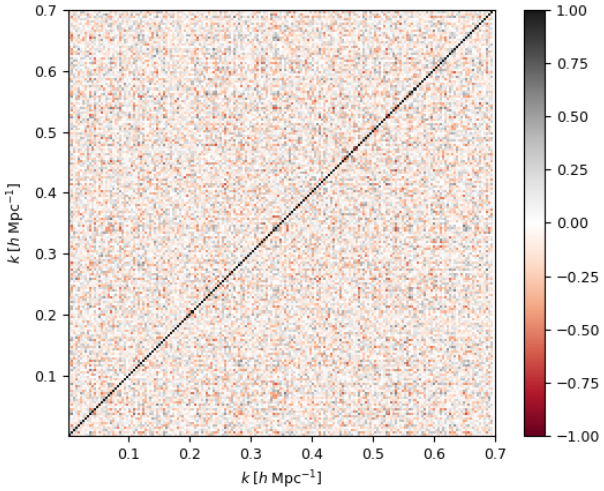


FIGURE 9.8: Correlation matrix of power spectrum amplitudes with respect to the mean value for the robust likelihood, normalized using the variance of amplitudes of the power spectrum modes. The correlation matrix shows that our augmented data model does not introduce any spurious correlation artefacts, thereby implying that it has properly accounted for the selection and foreground effects.

as expected from ΛCDM cosmology. The strong diagonal shape of the correlation matrix indicates that our proposed data model correctly accounted for any mode coupling introduced by survey geometry and foreground effects.

9.6 Summary and conclusions

The increasing requirement to control systematic and stochastic effects to high precision in next-generation deep galaxy surveys is one of the major challenges for the coming decade of surveys. If not accounted for, unknown foreground effects and target contaminations will yield significant erroneous artefacts and bias cosmological conclusions drawn from galaxy observations. A common spurious effect is an erroneous modulation of galaxy number counts across the sky, hindering the inference of 3D density fields and associated matter power spectra.

To address this issue, we proposed a novel likelihood to implicitly and efficiently account for unknown foreground and target contaminations in surveys. We described its implementation into a framework of non-linear Bayesian inference of large-scale structures. Our proposed data model is conceptually straightforward and easy to implement. We illustrated the application of our robust likelihood to a mock data set with significant foreground contaminations and evaluated its performance via a comparison with an analysis employing a standard Poissonian likelihood to showcase the contrasting physical constraints obtained with and without the treatment of foreground contamination. We have shown that foregrounds, when unaccounted for, lead to spurious and erroneous large-scale artefacts in density fields and corresponding matter power spectra. In contrast, our novel likelihood allows us to marginalize over unknown large-angle contamination amplitudes, resulting in a homogeneous inferred density field, thereby recovering the fiducial power spectrum amplitudes.

The results presented above in Section 9.5 clearly demonstrate the efficacy of our proposed likelihood in robustly dealing with unknown foreground contaminations for the inference of non-linearly evolved dark matter density fields and the underlying cosmological power spectra from deep galaxy redshift surveys. This method can be inverted to constrain foreground properties of the contamination. The inferred dark matter density allows building galaxy catalogues without contaminations. These can be compared to the actually observed number counts to reconstruct the foreground properties as the mismatch between the two catalogues.

We are convinced that our approach will contribute to optimizing the scientific returns of current and coming galaxy redshift surveys. We have demonstrated the effectiveness of our robust likelihood in the context of large-scale structure analysis. Our augmented data model remains nevertheless relevant for more general applications with other cosmological probes, with applications potentially extending even beyond the cosmological context.

Part IV

Deep learning for cosmological simulations

Chapter 10

Machine learning

10.1 Artificial neural networks

The development and application of machine learning algorithms to find patterns embedded in cosmological data sets is one of the key themes at the frontiers of research in the field (e.g. Ntampaka et al., 2019; Carleo et al., 2019). Machine learning techniques have proven to be extremely effective at extracting substantial information from data, identifying interesting outliers, data compression via dimensionality reduction and generating simulated or fake observations at minimal computational costs (Acquaviva, 2019).

A list of recent machine learning successes in cosmology is provided by Ntampaka et al. (2019). These compelling successful applications indicate that machine learning is now another crucial tool in our data analysis toolbox, with the potential to make significant contributions to cosmological research in the next decade. In particular, upcoming large galaxy surveys, such as Euclid (Laureijs et al., 2011; Racca et al., 2016) and LSST (Ivezic et al., 2008) will provide a plethora of opportunities to optimize and harness the capability of machine learning methods. Acquaviva (2019) puts forward a list of arguments and useful guidelines for the proper use of machine learning in solving the big data problems from future galaxy surveys. One of the salient points mentioned is the use of machine learning to model the currently unresolved and unknown physics involved in a given cosmic phenomenon, thereby providing a data-driven approach as an alternative to a simplistic (analytical) model. This is the rationale behind our development of a cutting-edge machine learning technique to map dark matter simulations to 3D halo distributions, presented in Chapter 12.

In the remainder of this chapter, we describe the basic concepts in machine learning and introduce the terminology. We outline the standard framework of the ubiquitous convolutional neural networks in Section 10.2, followed by a discussion of crucial practices for proper neural network implementation in Sections 10.3 and 10.4.

10.1.1 Perceptron

Artificial neural networks consist of simple interconnected adaptive units, analogous to the biological nervous system (Kohonen, Barna, and Chrisley, 1988; Kohonen, 1990), that combine to learn complex patterns. The fundamental building block of a neural network is the artificial *neuron*, which receives a set of weighted (scalar) inputs and outputs a single scalar, as illustrated in Fig. 10.1. This primitive artificial neuron is known as a *perceptron* (Goodfellow, Bengio, and Courville, 2016).

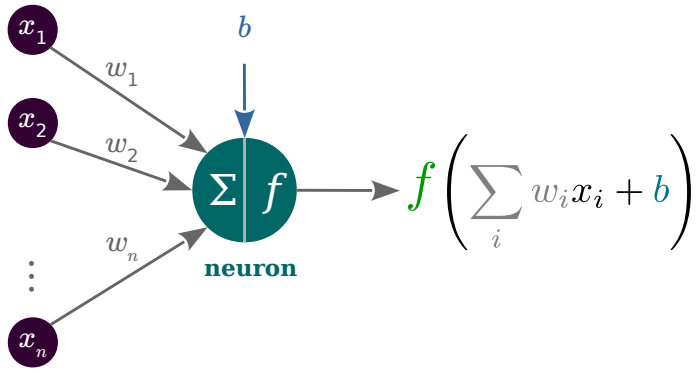


FIGURE 10.1: Basic unit of a neural network known as the perceptron. The neuron takes the sum of a set of weighted inputs and adds a bias parameter which is used to change the threshold at which the neuron is activated, followed by the application of an "activation" function. The latter is usually a non-linear function to encode a degree of non-linearity in the perceptron.

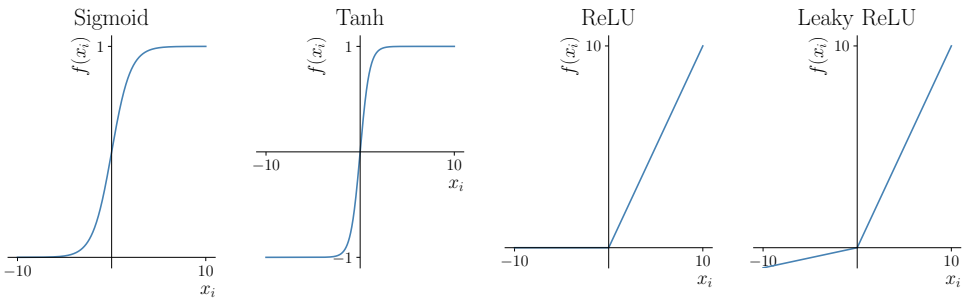


FIGURE 10.2: Examples of common activation functions.

The terminology of *deep* learning is usually attributed to neural networks having a large number of layers, typically exceeding four layers, and implies a certain degree of abstraction between the inputs and outputs which may not be within the limits of our understanding (Bengio, 2009). The first layer of the neural network is the *input layer* and is fed with the input, while the last layer is called the *output layer* and yields the result of the computation. All the other layers, sandwiched between the input and output layers, are referred to as the *hidden layers*. Stacking layers of neurons yields *multilayer perceptrons* (Nielsen, 2015), i.e. fully connected networks. In such a multilayer network, the composition of the straightforward operations described in Eq. (10.1) below, from the initial to the final layers, yields a highly non-linear mapping between the inputs and their corresponding outputs. An example of a multilayer perceptron with three input and one output neurons, and four neurons in each of the two intermediate hidden layers, is illustrated in Fig. 10.3.

In *feed-forward* neural networks, the information is propagated forward throughout the distinct layers of the network, from the inputs to their corresponding outputs. Every neuron in a given layer takes a linear combination of the outputs of a certain subset of neurons in the previous layer and applies a non-linear operation, commonly known as the *activation function* (Nielsen, 2015). If we define x_i^l and x_j^{l+1} as the i^{th} neuron of the l^{th} layer, j^{th} neuron of the $(l+1)^{\text{th}}$ layer, respectively, and w_{ij}^l, b_j^l as the trainable weights and bias of the l^{th} layer, then the outputs of the l^{th} layer can be written as

$$x_j^{l+1} = f\left[\sum_{i \in N^l} \left(w_{ji}^l x_i^l + b_j^l\right)\right] \equiv f(z_j^l), \tag{10.1}$$

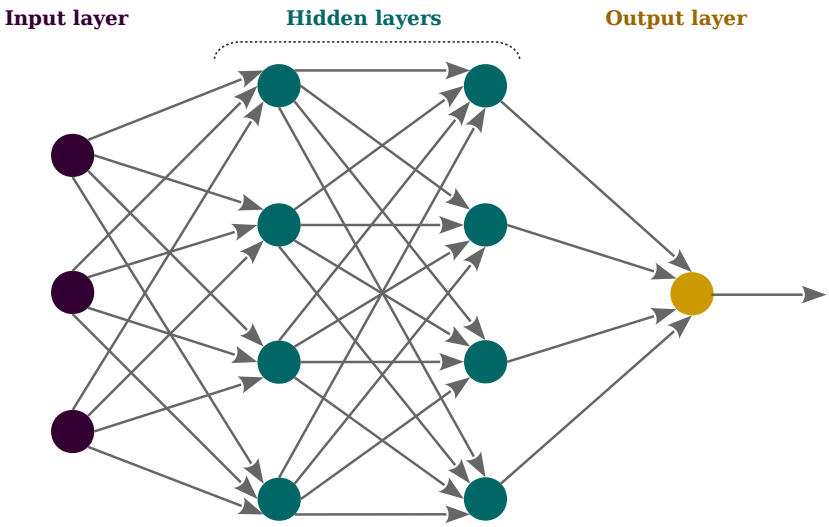


FIGURE 10.3: Example of a multilayer perceptron (fully connected network) with three inputs, one output and two hidden layers with four nodes in each layer.

where N^l is the number of neurons in the l^{th} layer and f is the activation function described in the following section. We will refer to z_j^l as the weighted, biased input in the sections below for simplicity. The purpose of the bias parameter associated to a neuron is to vary the threshold at the neuron is triggered (McCulloch and Pitts, 1943). The essence of neural networks is to model complex and highly non-trivial mappings between the inputs and the corresponding outputs by applying the non-linear activation function to linear combinations of their input elements (Goodfellow, Bengio, and Courville, 2016). The key advantage is that no *a priori* information of the unknown underlying mapping is required, as the network can learn it from a sufficient set of labelled data. This is known as the *training set* since it consists of a set of inputs for which the corresponding outputs are known, and this is the actual data set used to fit the model. This is known as *supervised learning* (Carleo et al., 2019) and the training procedure is detailed in Section 10.1.3.

10.1.2 Activation function

Some popular activation functions are the sigmoid, hyperbolic tangent, rectified linear activation (`ReLU`) (Nair and Hinton, 2010) and leaky rectified linear activation (`leaky ReLU`) (Maas, Hannun, and Ng, 2013), depicted in Fig. 10.2. We make use of the `ReLU` and `leaky ReLU` in our neural networks in Chapter 12, with the latter defined as follows:

$$f(z_i) = \begin{cases} \alpha z_i, & z_i < 0 \\ z_i, & z_i \geq 0. \end{cases} \tag{10.2}$$

We obtain the `ReLU` function simply by setting α to zero, where α is a pre-defined small negative slope. The class of the `ReLU` activation functions are usually preferred over the sigmoid and hyperbolic tangent functions since they mitigate the vanishing gradient issue (Nielsen, 2015) outlined in Section 10.1.3, and are computationally less expensive.

10.1.3 Backpropagation

The multiple network parameters in each layer, i.e. the weights and biases, are optimized using gradient descent applied to the error in the output. This procedure is known as *backpropagation*, due to the use of the chain rule by the gradient to propagate the information backwards from the respective outputs of each layer (Goodfellow, Bengio, and Courville, 2016). The quantity to be minimized is referred to as the *loss function*, denoted by $\mathcal{L}$, which is a metric that measures the distance between the network predicted and true known values for the elements in the training set.

The change in the output Δx_j^{l+1}, from node j can be expressed as a linear function of the change in the weights, Δw_{ij}, and change in the biases of node j, Δb_j^l, as follows[1] (Nielsen, 2015):

$$\Delta x_j^{l+1} \approx \sum_{i \in N^l} \left(\frac{\partial f(z_i^l)}{\partial w_{ij}^l} \Delta w_{ij}^l + \frac{\partial f(z_i^l)}{\partial b_{ij}^l} \Delta b_j^l \right), \tag{10.3}$$

which can be computed using gradient descent. The above can be reversed, such that the change in the weights and biases can be computed via the change in the outputs for each layer. In order for the network's output to match the known outputs from the training set, we can perform a minimization procedure on the error, i.e. loss function (Rumelhart, Hinton, and Williams, 1988). Although the choice of this loss function will depend on the specific task to be performed by the neural network, a suitable loss function, by design, should be sensitive to slight changes in the output induced by extremely small changes in the weights and biases (LeCun et al., 2012).

For optimal training of the weights and biases, the loss function should gradually be reduced to its minimum value, i.e. $\Delta\mathcal{L} < 0$. The change in the loss function, given a change in the weights and biases, can be written as (Rumelhart, Hinton, and Williams, 1988):

$$\Delta\mathcal{L} \approx \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}^l} \Delta \boldsymbol{w}^l, \quad \text{and} \quad \Delta\mathcal{L} \approx \frac{\partial \mathcal{L}}{\partial \boldsymbol{b}^l} \Delta \boldsymbol{b}^l, \tag{10.4}$$

respectively, where we employ a vector notation above for simplicity. For the change in the loss function to be negative, we choose

$$\Delta \boldsymbol{w}^l = -\eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}^l}, \quad \text{and} \quad \Delta \boldsymbol{b}^l = -\eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{b}^l}, \tag{10.5}$$

which would imply that

$$\Delta\mathcal{L} = -\eta \left|\left| \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}^l} \right|\right|^2, \quad \text{and} \quad \Delta\mathcal{L} = -\eta \left|\left| \frac{\partial \mathcal{L}}{\partial \boldsymbol{b}^l} \right|\right|^2, \tag{10.6}$$

respectively. η is the *learning rate* for the gradient descent algorithm, which should be set to a typically small value, for the approximations in Eq. (10.4) to hold, but sufficiently large for the weights to update (Nielsen, 2015). The weights and biases of the network are updated via backpropagation

[1] Note that this approximation is valid only for very small steps.

until convergence of the loss function, as follows:

$$\boldsymbol{w}_{n+1}^{l} = \boldsymbol{w}_{n}^{l} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{w}_{n}^{l}}, \quad \text{and} \quad \boldsymbol{b}_{n+1}^{l} = \boldsymbol{b}_{n}^{l} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{b}_{n}^{l}}, \tag{10.7}$$

with the above transformations yielding an optimal reduction in the loss function on every subsequent input of training data (Rumelhart, Hinton, and Williams, 1988).

The rate of change of the loss function, given the weighted, biased input $\boldsymbol{z}^l$ at any layer, with this gradient denoted by $\partial\mathcal{L}/\partial\boldsymbol{z}^l$, can be evaluated via the chain rule which relates the rate of change of the weights or biases between different layers (Rumelhart, Hinton, and Williams, 1988). For the output layer, this can be obtained as follows:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{z}^{\text{out}}} = \frac{\partial \mathcal{L}}{\partial f(\boldsymbol{z}^{\text{out}})} \odot \frac{\partial f(\boldsymbol{z}^{\text{out}})}{\partial \boldsymbol{z}^{\text{out}}}, \tag{10.8}$$

where $\odot$ implies the *Hadamard* or *Schur* product[2] (Davis, 1962). We can propagate the gradient of the loss function to any preceding layer l:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{z}^{l}} = \left(\boldsymbol{w}^{l+1}\right)^{T} \frac{\partial \mathcal{L}}{\partial \boldsymbol{z}^{l+1}} \odot \frac{\partial f(\boldsymbol{z}^{l})}{\partial \boldsymbol{z}^{l}}, \tag{10.9}$$

where $(\boldsymbol{w}^{l+1})^T$ indicates the transpose of the weights at the subsequent $(l+1)^{\text{th}}$ layer (Nielsen, 2015). As such, the error propagation through the activation function of the l^{th} layer is done via the Hadamard product, thereby yielding the rate of change of $\boldsymbol{z}^l$. The above Eqs. (10.8) and (10.9) allow us to compute the gradient of the loss function with respect to the weighted, biased output at any given layer. The gradients of the loss function with respect to the weights and biases, respectively, can then be determined using the chain rule and the relation (10.9), as follows:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}^{l}} = f\left(\boldsymbol{z}^{l-1}\right) \frac{\partial \mathcal{L}}{\partial \boldsymbol{z}^{l}}, \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{b}^{l}} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{z}^{l}}. \tag{10.10}$$

The final step in the backpropagation algorithm is to update the weight and biases using Eq. (10.7).

We can optimize the neural network hyperparameters, such as the number of layers and neurons, and learning rate, by a quantitative evaluation of the model's performance over the *validation set* with known output values. The purpose of the validation data set is to allow frequent model evaluations on occasionally seen data, but the network never learns from the latter. A common practice is to split the original data set into two portions for the training and validation. The final assessment of the mapping learnt by the network is done over a separate unseen *test set*, which provides an unbiased evaluation of the trained model (Carleo et al., 2019).

The vanishing gradient predicament arises when $\partial\mathcal{L}/\partial w_{ij} \approx 0$, such that $w_{n+1} \to w_n$ (cf. Eq. (10.7)), with an analogous argument applying to the biases. This implies that the neuron *saturates*, impeding its ability to learn (Rosasco et al., 2004). This is due to an activation function where $f(x_i) \approx 0$ or $f(x_i) \approx 1$ causing the gradient to tend to zero, as in the case of the sigmoid function. The latter is also non-zero centred, which is not conducive to the effectiveness of gradient descent

[2]This is simply the element-wise product. For instance, for two vectors of same dimension: $(s \odot t)_j \equiv s_j t_j$.

during training. These two factors are detrimental to training efficiency and are therefore important elements to consider in the choice of activation functions. The hyperbolic tangent function also saturates, but it is zero-centred, which alleviates some training issues, and is therefore preferred over the sigmoid function (LeCun et al., 2012).

10.2 Convolutional neural networks

Convolutional neural networks (hereafter CNNs) (LeCun and Bengio, 1995; LeCun et al., 1998) are a particular type of network, tailored for problems where spatial information is important. A schematic of a standard CNN is depicted in Fig. 10.4, where the input image is a 2D snapshot of the matter density field. A convolutional kernel, or *filter*, of a given size, encoding a set neurons, is applied to each pixel of the input image and its neighbourhood as it sweeps through the whole region. Each pixel in a given layer is only a function of the pixels in the previous layer which are contained within the window defined by the kernel, known as the *receptive field* of the layer. This produces a *feature map* which contains high values in the pixels which match the pattern encoded in the weights and biases of the corresponding neurons that are optimized during training. To probe different aspects of the input image, a convolutional layer typically employs several filters, yielding a set of feature maps which become the inputs to the subsequent layer. In the standard framework, the convolutional step is usually followed by a *pooling* layer for the purpose of subsampling or dimensionality reduction (Goodfellow, Bengio, and Courville, 2016). Alternating between these two types of layers will reduce the original input image to a compact representation of features, which can be flattened to a vector. This vectorized feature is subsequently fed to the final layer which is a fully connected layer, thereby generating an output (Lecun, Bengio, and Hinton, 2015).

Formally, the convolutional operation can be described as a specialized linear operation, with the discrete convolution implemented by matrix multiplication, such that a convolutional layer, labelled by l, can be computed via

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right), \tag{10.11}$$

where f is the activation function, k corresponds to the convolutional kernel, M_j represents the receptive field and b is the bias parameter (Goodfellow, Bengio, and Courville, 2016).

The architecture outlined above allows the network to autonomously extract significant spatial features from the input image. Stacking several convolutional layers creates a hierarchical internal representation of features, allowing the network to identify increasingly complex patterns as more layers are added. As the information propagates across distinct layers of neurons, the information encoded in the input image is transformed into an internal representation encapsulating the most relevant features from the image. Convolutional layers, therefore, provide a natural approach to take spatial context into consideration, with a stack of such layers increasing the sensitivity of subsequent layers to features on increasingly larger scales, i.e. the size of the receptive field grows the deeper we go in the network. Moreover, CNNs retain the local information while performing the

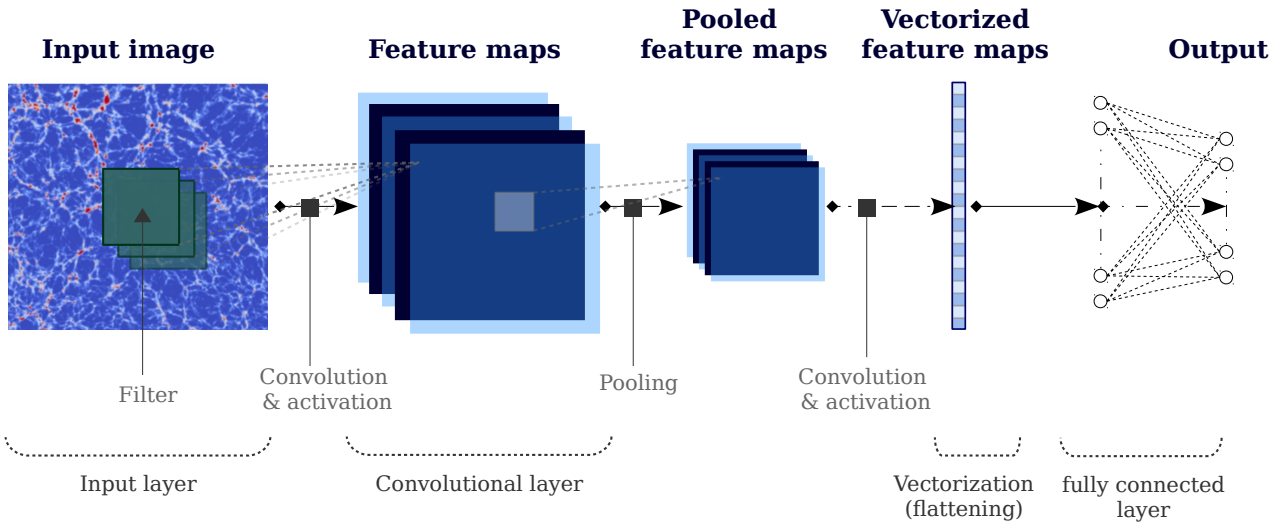


FIGURE 10.4: Schematic of a convolutional neural network. The input image, here a snapshot of a non-linearly evolved density field, is scanned using a convolutional kernel to extract spatial features. There is a set of such kernels in one convolutional layer to extract distinct features, yielding a set of feature maps. In the standard framework, a subsequent pooling step downgrades the resolution, and the convolutional and pooling steps are repeated until the input image is reduced to a compact representation of features, at which point it is flattened to a vector and fed to a fully connected layer, eventually resulting in a given output.

convolution on adjacent pixels, such that convolutional layers allow both local and global information to propagate through the network (Lecun, Bengio, and Hinton, 2015).

Such deep CNNs have been particularly useful for image processing and pattern recognition, and are therefore gaining in popularity in the cosmological community. Caldeira et al. (2018) exploited deep CNNs for the reconstruction of CMB lensing potential. Schmelzle et al. (2017) implemented a similar network architecture for cosmological model discrimination using noisy convergence maps, capturing non-Gaussian information from the data. Based on the work of Gupta et al. (2018), Ribli, Pataki, and Csabai (2019) implemented an improved CNN architecture, exploiting such non-Gaussian information in simulated weak lensing maps, which led to tighter cosmological constraints. The application of CNNs to fields on the sphere has recently been the subject of a few investigations. Cohen et al. (2018) used spherical convolution in Fourier space to implement spherical CNNs, while Perraudin et al. (2019) devised a graph-based representation of spherical convolutional layers, which they applied to a classification problem of weak lensing mass maps derived from distinct cosmological models. Krachmalnicoff and Tomasi (2019) developed a pixel-based algorithm to build spherical CNNs, showcasing its performance for the estimation of cosmological parameters from simulated CMB maps. Recent work involving the large-scale structures are outlined in Chapter 12.

10.3 Overfitting

During training, there is a risk that the neural network learns about specific recurring features present in the training set, which are not relevant or useful in practice, and this is known as *overfitting* (Draper and Smith, 2014). As a result, the network weights are tuned to detect these particular

features in the test set, such that they overlook the actual informative features in the data.

To alleviate the risk of overfitting, regularization methods (e.g. Nowlan and Hinton, 1992), where weight penalties are introduced in the loss function, are commonly employed. A conceptually straightforward solution to overfitting was recently proposed in the form of *dropout* (Srivastava et al., 2014). This technique entails switching off a given fraction of randomly chosen network weights by setting them to zero, such that only the fraction of non-zero weights are updated during training. This is equivalent to sampling different networks in the total number of possible networks for the given architecture, where each network can learn about specific features, with the ensemble of networks being insensitive to these features, thereby preventing overfitting. Dropout, therefore, allows the network to learn distinct features, without overlearning certain features present in the data, leading to significant improvement in the performance of deep networks (Srivastava et al., 2014). In the next section, we describe another alternative involving *data augmentation*, which we employ in training our neural network in Chapter 12.

10.4 Data augmentation

Training deep neural networks requires a large amount of data and it is well-known that a large data set is conducive to network performance due to the inherent model complexity. However, this may be a limitation for certain problems where it is computationally expensive to generate large training sets, as in the case of cosmological N-body simulations. Data augmentation provides a straightforward solution to this predicament by artificially augmenting the training set size. To this end, data augmentation takes an original piece of data and makes a particular modification to it, such that both the original and synthetically modified data can be used as inputs to the network (e.g. Witten et al., 2016). Convolutional layers, for instance, encode a certain degree of invariance with respect to translation, orientation and size, and this is essentially the premise of data augmentation. Some commonly used techniques are:

- Flipping the image across either the horizontal or vertical axis;
- Applying a random rotation to the image;
- Applying a certain translation to the image;
- Adding artificial (Gaussian) noise to the image pixels;
- Randomly cropping a certain portion of the image.

The above transformations generate a similar image which the neural network can learn from, without being biased by the salient features in the original image. As a result, this prevents the network from learning irrelevant patterns or features in the original data set. The effectiveness of the above standard data augmentation techniques for image classification was recently demonstrated by Perez and Wang (2017). In Chapter 12, we artificially augment our training set on the fly by performing a random rotation to our cosmological simulations, such that we extract 3D slices from a randomly oriented region.

Chapter 11

Generative adversarial networks

This chapter draws from Kodi Ramanah, Charnock, and Lavaux (2019).

11.1 Introduction

Generative adversarial networks were first proposed in the seminal work of Goodfellow et al. (2014), and have emerged as powerful generative models, albeit with some limitations, as described below, which have been addressed by recent developments. In the standard GAN framework, generative modelling is formulated as a game between two competing networks, trained in an adversarial setting. A generator network, parameterized by a vector θ, $\mathcal{G}_\theta$ produces some artificial data given some vector of random noise, and a discriminator network $\mathcal{D}$ differentiates between the synthetic output of the generator and the true data. Otherwise said, the generator network provides a way to map one distribution to another, and notably produces samples from the latter target distribution.

The game between the generator $\mathcal{G}_\theta$ and discriminator $\mathcal{D}_\epsilon$ can be formally expressed as the minimax objective of distance $\mathcal{V}(\mathbb{P}_\mathrm{r}, \mathbb{P}_\mathrm{g})$:

$$\mathcal{V}(\mathbb{P}_\mathrm{r}, \mathbb{P}_\mathrm{g}) = \min_{\theta \in \mathbb{R}^{N_\mathcal{G}}} \max_{\epsilon \in \mathbb{R}^{N_d}} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathbb{P}_\mathrm{r}} \left[\log\left(\mathcal{D}_\epsilon(\boldsymbol{x})\right)\right] + \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathbb{P}_\mathrm{z}} \left[\log\left(1 - \mathcal{D}_\epsilon(\mathcal{G}_\theta(\boldsymbol{z}))\right)\right], \tag{11.1}$$

where $\mathbb{P}_\mathrm{r}$ corresponds to the data distribution and $\mathbb{P}_\mathrm{g}$ is the model distribution defined by the distribution $\mathbb{P}_\mathrm{z}$ transformed by the generator $\mathcal{G}_\theta$, with θ corresponding to the network weights. The source distribution $\mathbb{P}_\mathrm{z}$ is often a uniform or Gaussian distribution. In this work, the vector $\boldsymbol{z}$, and the corresponding distribution $\mathbb{P}_\mathrm{z}$, are provided by another complex distribution, as discussed in Section 12.4. We note that the discriminator network must also be provided and optimized according to the weights ϵ.

The training phase is completed when a Nash equilibrium (Nash, 1951) is reached, i.e. when neither of the two opponents can improve by unilaterally adjusting their strategy, and the discriminator cannot distinguish between the true and artificial data. At this point, the generator, in principle, would have learned to output a sufficiently good representation of the real data probability distribution, i.e. $\mathbb{P}_\mathrm{g} \approx \mathbb{P}_\mathrm{r}$, and would therefore be able to map a known or latent probability distribution to the target data distribution.

Unfortunately, the standard GAN framework is vulnerable to training instabilities, often resulting from issues involving vanishing gradients or mode collapsing, where the generator output lies

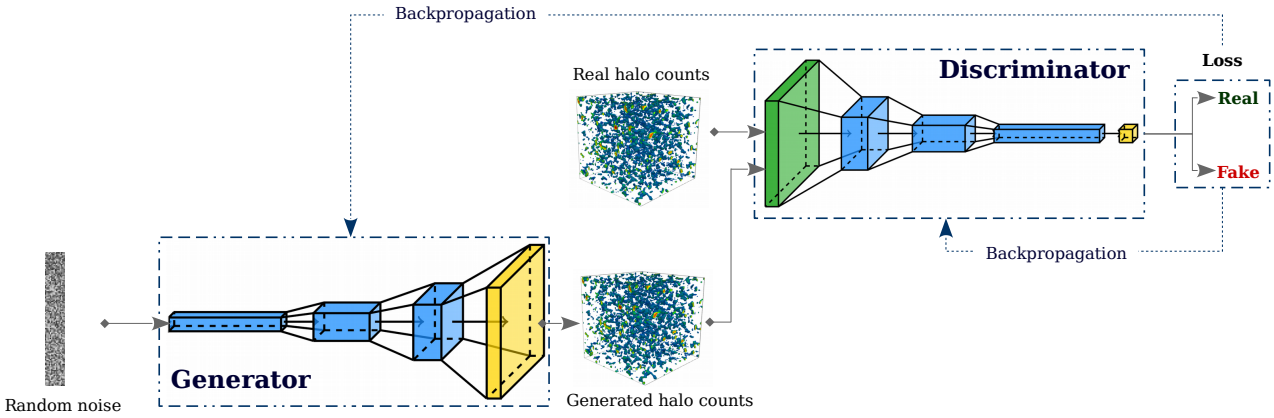


FIGURE 11.1: Schematic of a generative adversarial network, where two networks are trained in an adversarial setting. The task of the generator is to generate realistic samples from input random noise and the discriminator takes as input the generated or fake samples and classifies them as real or fake. The two networks compete against each other until the discriminator can no longer distinguish between the real and fake samples, at which point the generator has learned how to map random noise to extremely realistic samples.

in a restricted phase space, thereby producing incoherent results. In remainder of this chapter, we review some of the successes of the improved deep convolutional GAN and outline the underlying conceptual framework of the recently proposed variant of Wasserstein GAN.

11.2 Deep convolutional GANs

Despite the drawbacks outlined above, GANs have still been shown to be extremely successful. Radford, Metz, and Chintala (2015) developed an improved GAN architecture, known as deep convolutional GAN (DCGAN), by replacing the multilayer perceptrons in the generator and discriminator networks with convolutional layers (Lecun, Bengio, and Hinton, 2015). A schematic of the DCGAN is provided in Fig. 11.1. Other improvements, such as batch normalization (Ioffe and Szegedy, 2015), i.e. ensuring that the input to each unit is normalized with zero mean and unit variance, were also introduced to stabilize the learning and promote gradient flow in deeper networks. Such infrastructural upgrades result in overall improved training stability and render the network more robust to discrete-mode and manifold model collapse (Metz et al., 2016; Arjovsky and Bottou, 2017). Nevertheless, DCGANs remain susceptible to model instabilities.

11.3 Wasserstein GANs

Arjovsky, Chintala, and Bottou (2017) proposed another variant of GAN, which encodes an alternative loss function based on the Wasserstein-1 distance between a real and a generated distribution, also known as the Earth Mover's distance. This distance can be informally and intuitively interpreted as the minimum cost of transporting mass in order to transform a probability distribution into a given target distribution. This variant, referred to as Wasserstein GAN (or WGAN) despite there being no adversarial component to the training, has shown to be capable of learning arbitrarily complex probability distributions of a panoply of data sets (Arjovsky, Chintala, and Bottou, 2017), leading to extremely realistic results (e.g. Karras et al., 2017).

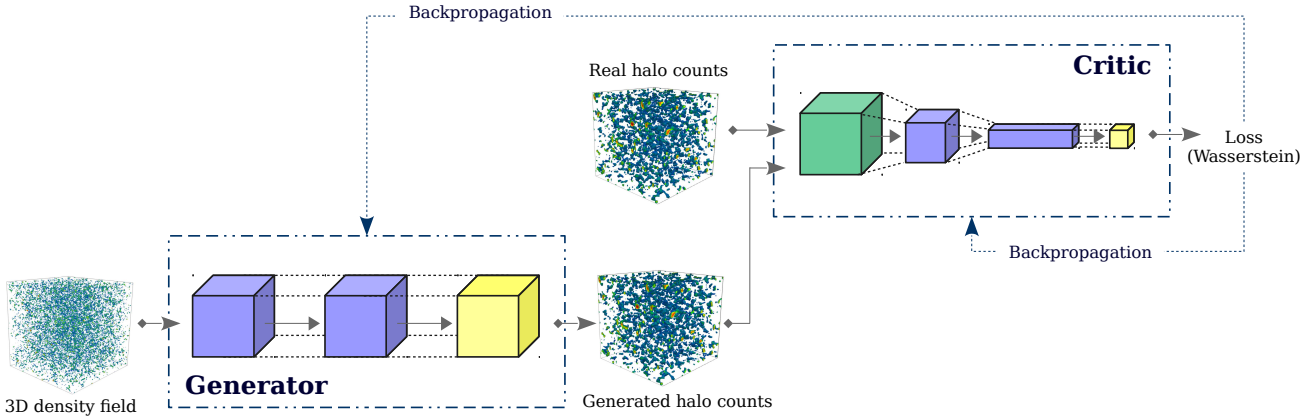


FIGURE 11.2: Schematic representation of Wasserstein halo painting network implemented in this work. The role of the generator is to learn the underlying non-linear relationship between the input 3D density field and the corresponding halo count distribution. The difference between the output of the critic for the real and predicted halo distributions is the approximately learnt Wasserstein distance and is used as the loss function which must be minimized to train the generator.

The Wasserstein distance, $\mathcal{W}(\mathbb{P}_r, \mathbb{P}_g)$, has the desired properties for the convergence of sequences of probability distributions (Arjovsky, Chintala, and Bottou, 2017). The WGAN value function can be expressed via the Kantorovich-Rubinstein duality (Villani, 2008) as

$$\mathcal{W}(\mathbb{P}_r, \mathbb{P}_g) = \sup_{\mathcal{C}\in\mathfrak{C}} \left\{ \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r} [\mathcal{C}_\epsilon(x)] - \mathop{\mathbb{E}}_{z\sim\mathbb{P}_z} [\mathcal{C}_\epsilon(\mathcal{G}_\theta(z))] \right\}, \tag{11.2}$$

where the supremum is over the set of 1-Lipschitz functions denoted by $\mathfrak{C}$, such that minimizing the above value function with respect to the generator parameters will also minimize $\mathcal{W}(\mathbb{P}_r, \mathbb{P}_g)$ for the case of an optimal discriminator. The discriminator network $\mathcal{D}(x)$ is now designated as the critic $\mathcal{C}(x)$, since it is not trained to differentiate or classify as in the standard GAN framework. To ensure that the critic is a 1-Lipschitz function, the form of the function that the network takes must be closed. This condition can be enforced by restricting the allowed weight space for the critic network. WGANs are, therefore, simply generative networks where the loss function is learned using a second network. This new distance mitigates the concept of adversarial training by broadening the concept further and allowing to measure differences between whole distributions. By relying on the earth-mover concept, it also ensures that the critic stays within a safe subspace contrary to GAN discriminatory network. A schematic representation of our generative network, used to establish a mapping between the 3D density field and its corresponding halo count distribution, is illustrated in Fig. 11.2. A detailed description of the training methodology is provided in Section 12.4.

11.4 Prior work involving GANs

GANs, and their variants mentioned above, are becoming increasingly popular among the astrophysical and cosmological community by virtue of their versatility and effectiveness to achieve impressive results. Mustafa et al. (2019) developed a DCGAN to generate cosmological weak lensing convergence maps with high statistical confidence, while the de-noising of such maps within a GAN framework was investigated by Shirasaki, Yoshida, and Ikeda (2018). GANs have also been

optimized for the efficient generation of realistic 2D realizations of the cosmic web, demonstrating their ability to capture the complexity of large-scale structures (Rodríguez et al., 2018). Tröster et al. (2019) employed a GAN to map dark matter density fields to gas pressure distributions, thereby augmenting N-body simulations with baryons. Other interesting applications of GANs involve denoising galaxy images to recover impressive detailed features, outperforming standard convolution methods (Schawinski et al., 2017), separating AGN from their host galaxy's light profile (Stark et al., 2018), deblending galaxy superpositions (Reiman and Göhre, 2018), atmospheric retrievals on exoplanets (Zingales and Waldmann, 2018) and generating physically realistic galaxy images (Fussell and Moews, 2018) and deep galaxy fields (Smith and Geach, 2019). The improved variant of WGAN has been used for the generation and refinement of signal patterns of particle detectors from simulations of cosmic-ray induced air showers (Erdmann et al., 2018). Zamudio-Fernandez et al. (2019) recently used a WGAN to generate 3D cosmic neutral hydrogen (HI) distributions with properties closely matching those from costly cosmological hydrodynamic simulations.

Chapter 12

Painting halos from 3D dark matter fields

The work presented in this chapter is based on Kodi Ramanah, Charnock, and Lavaux (2019).

12.1 Introduction

Investigating the formation and evolution of dark matter halos, as the key building blocks of cosmic large-scale structure, is essential for constraining various cosmological models and further understanding our Universe. The highly non-linear dynamics involved nevertheless renders this a complex problem, with N-body simulations currently the only tool to compute the non-linear gravitational evolution from initial conditions (e.g. Springel, 2005), yielding mock dark matter halo catalogues as the main output. The resulting catalogues of positions, velocities and masses of halos are necessary for cosmological inference from galaxy surveys. As an example, running very large simulations of pure dark matter, such as FUR-DEUS (Alimi et al., 2012), to generate fake observations of the full Universe several times is not feasible, and requires a large amount of memory and disk storage. A way to emulate such simulations, quickly and reliably, would be of use to a wide community as a new method for data analysis and light cone production for the next cosmological survey missions such as Euclid (Laureijs et al., 2011) and Large Synoptic Survey Telescope (LSST) (Ivezic et al., 2008).

With the recent developments in the field of machine learning, deep generative modelling techniques have emerged as a viable tool to construct emulators of expensive simulations. In this work, we present such a deep learning approach to generate the 3D halo distribution from dark matter simulations. Using our construction, the neural network is used to learn the mapping from the dark matter density to halo fields and therefore predicts the abundance of halos at a given position based on the large-scale density distribution. Once trained, the emulator is capable of rapidly predicting simulations of halo distribution based on a non-linearly evolved density field. Furthermore, by learning this mapping for different halo mass bins, we can also predict the mass distribution of the halos.

A key aspect of our approach is that the neural network is able to *paint* a halo count distribution from a numerically cheap non-linear density field, such as a realization obtained via Lagrangian Perturbation Theory (LPT), which requires negligible computational resources on modern machines relative to N-body simulations. The interest of this technique lies in the possibility that most of the cosmological dependence of the observed matter distribution in the N-body simulation is already encoded in the inexpensive LPT simulation. As a result, this approach would eliminate the need to

run a full particle mesh simulation, thereby allowing detailed analyses of state-of-the-art surveys to be feasible on regular computing facilities. The idea of painting complex astronomical objects has been implemented in the past, notably for galaxies with MOLUSC (Sousbie et al., 2008) and LYMAS (Peirani et al., 2014). Another related work is the PINOCCHIO algorithm (Monaco et al., 2002; Monaco, Theuns, and Taffoni, 2002) for identifying dark matter halos in a given numerical realization of the linear density field. However, our aim here is to build an automated model generator with even higher accuracy and insensitivity to the underlying cosmology.

We take inspiration from a recently proposed variant of generative models, known as generative adversarial networks (GANs) (Goodfellow et al., 2014), which have met considerable success with a range of applications, such as generating extremely realistic fake celebrities (Karras et al., 2017; Karras, Laine, and Aila, 2018) and artificial bedroom images (Radford, Metz, and Chintala, 2015). In particular, we will use the key ideas in training WGANs, i.e. GANs optimized using the Wasserstein distance (Arjovsky, Chintala, and Bottou, 2017), to ensure that our network is able to paint halos well. The GANs, and variants thereof, are described in depth in Chapter 11.

Neural networks have recently been employed for various aspects of large-scale structure analysis. He et al. (2019) devised a deep neural network to predict the non-linear cosmic structure formation from linear perturbation theory, with the network architecture based on the U-Net (Ronneberger, Fischer, and Brox, 2015) learning model. Zhang et al. (2019) constructed a two-phase convolutional neural network architecture to map 3D dark matter fields to the corresponding galaxy distribution in hydrodynamic simulations. Berger and Stein (2019) implemented a 3D deep convolutional neural network to generate mock halo catalogues by identifying protohalos directly from the cosmological initial conditions. This was preceded by a similar work by Lucie-Smith et al. (2018), where a random forest classifier was used to trace the halos formed in N-body simulations back to their initial conditions. Modi, Feng, and Seljak (2018) proposed a framework, with the end product being the converse, to reconstruct initial conditions from the halo fields using the multilayer perceptron, i.e. fully connected neural networks. Deep convolutional networks were also used to classify the distinct components of the cosmic web such as filaments and walls from N-body simulations (Aragon-Calvo, 2019).

This chapter is organized as follows. Section 12.2 describes several other network refinements which we implement in designing our network architecture. Section 12.3 illustrates the relevant aspects of the dark matter simulations used in the training and validation of our neural network, as described in Section 12.4. We follow up by investigating the performance of the algorithm in terms of various diagnostics, as detailed in Section 12.5. Finally, in Section 12.6, we summarize our main findings and discuss the areas of application where our halo painting network, tailored for mapping dark matter distributions to halo fields, can be optimized.

12.2 Network refinements

To optimize the performance of our halo painting network, we consider several architectural upgrades which have been recently presented in the literature. In this section, we briefly describe the network refinements which are implemented in our network architecture and training machinery.

12.2.1 Inception

A particular deep convolutional architecture was proposed by Szegedy et al. (2015) and Szegedy et al. (2016), code-named Inception, which achieved state-of-the-art performance for object classification and detection purposes. This is a novel level of organization which results in increased network depth and width, thereby improving the efficiency of deep neural network architecture.

The original Inception module (Szegedy et al., 2015) consists of a series of convolutions, on the same level, with kernel sizes of 5×5, 3×3 and 1×1 in each Inception module, which is diagrammatically represented in Fig. 12.1. In this diagram, the blue box represents convolution within each feature space, while the gray convolutions are taken across features. In the case of blue convolutions, the kernel size defines the receptive field of a layer, i.e. the size of the patch of the input which affects the connected output. The outputs are concatenated and fed to the subsequent components of the network, which may be another Inception module. This concatenation increases the amount of features available for the next layer to make more computations. Stacking successive Inception modules yields further depth, with each module being optimized to recognize features on various scales. The Inception module, as proposed in Szegedy et al. (2015), also introduces max pooling layers and 1×1 convolutions. A max pooling transformation downgrades the resolution of the input grid by some factor by taking local maxima. The two additional 1×1 convolutions are inserted in each branch before the 3×3 and 5×5 convolutions for dimensionality reduction to limit the computational resources required to a reasonable amount. Indeed, by convolving along the feature space, they may produce new compressed features with lower dimensions, though with the same size for each feature. This possibility of controlled dimensionality reduction is the crux of the Inception module.

The essence of Inception lies in increased network depth, while obviating the potential drawbacks of deep convolutional networks. Naively stacking large convolutional layers to build very deep networks is computationally expensive and renders the networks prone to overfitting. Moreover, gradient updates may not flow smoothly throughout such networks. The key advantage of Inception is therefore a remarkable gain in quality where the computational workload is not greatly increased compared to networks with lower depth and width.

12.2.2 Residual connections

A residual learning framework was proposed by He et al. (2016) to improve the training of deep neural networks. Conceptually, the framework relies on a reformulation of a given layer as learning the residual mapping with reference to its input, rather than directly learning the desired underlying mapping. He et al. (2016) empirically demonstrated that such residual networks mitigate the degradation issue, whereby very deep networks are susceptible to saturation and eventually degradation of the training accuracy. This is the result of higher training error with increasing network depth. Residual networks are commonly abbreviated as *ResNets*.

The desired underlying mapping can be formally denoted as $\mathcal{H}(x_i)$, such that the stack of nonlinear layers would fit the residual mapping of $\mathcal{F}(x_i) = \mathcal{H}(x_i) - x_i$, where i labels a given layer. The original mapping can therefore be reformulated as $\mathcal{H}(x_i) = \mathcal{F}(x_i) + x_i$. The hypothesis is that the

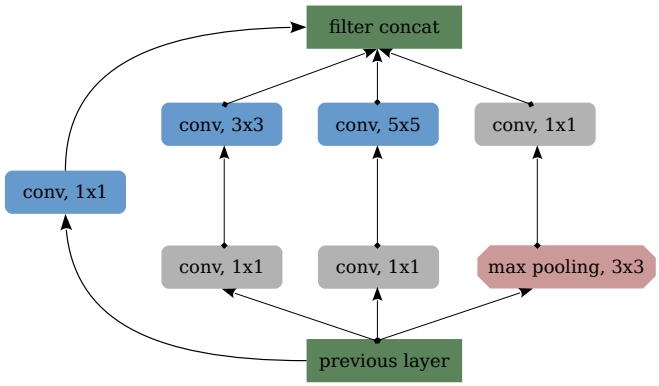


FIGURE 12.1: Original Inception block, with filter dimensionality reduction, as proposed by Szegedy et al. (2015). The grey convolutional ("conv") layers indicate the 1×1 convolutions introduced for the purpose of dimensionality reduction. We note that they are actually convolutions according along the feature space, of dimension F, of the previous layer, which is mono-dimensional, to produce $\tilde{F}$ new features but with the same physical size. As $\tilde{F} < F$, we achieve dimensionality reduction with respect to the number of features. The blue 1×1 convolutional layer corresponds to a normal convolution inside each feature. The output of the Inception module is the concatenation of the filters ("filter concat") from the respective convolutional layers. This increases the dimension of the feature space depending on the number of output of each of the top convolutions. With the exclusion of the max pooling path, we make use of this architecture in the Wasserstein halo painting network.

residual function of $\mathcal{F}(x_i)$ is easier to optimize than the desired function of $\mathcal{H}(x_i)$ (He et al., 2016). The formulation of $\mathcal{F}(x_i) + x_i$ can be implemented via feedforward neural networks with "shortcut connections". Such connections skip one or more layers and perform identity mapping.

The implementation of residual connections in typical deep neural networks is straightforward. The only modification involves adding an extra identity shortcut connection between the input to a given layer and the output of the subsequent layer (cf. Fig. 12.2). Another positive aspect of adding residual connections is that no extra parameters are required and therefore the computational workload is not greatly increased except for negligible element-wise addition. The residual learning framework, in a nutshell, yields substantial accuracy gains by allowing extremely deep networks to be trained since the gradient information can flow without degradation to the early layers, thereby alleviating the vanishing gradient problem. In a standard encoder-decoder structure, this especially facilitates the propagation of small-scale information as the size of the images is reduced gradually in the encoding phase (Isola et al., 2016).

In this work, we implement a combination of ResNets and Inception architecture, as depicted in Fig. 12.3 with the same color convention as in Fig. 12.1, yielding residual Inception blocks. Szegedy et al. (2017) have demonstrated that the introduction of residual connections within the Inception module leads to significant improvement in training speed. The state-of-the-art performance obtained by combining these two network refinements is a crucial factor behind our choice of network architecture for the halo painting network.

12.2.3 Gradient penalty

As discussed above, the weights of the critic network for WGANs must be restricted to ensure that this network is 1-Lipschitz. When first conceived, weight clipping was used to enforce this criterion.

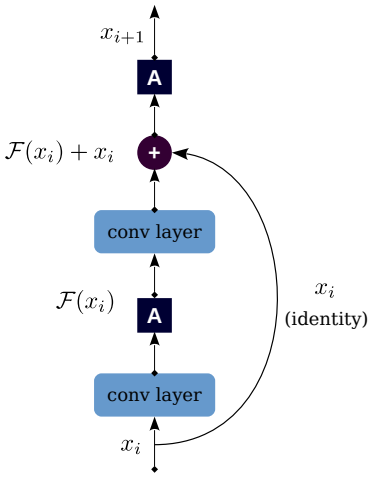


FIGURE 12.2: Residual learning block, as proposed by He et al. (2016). Activations are denoted by **A**, with element-wise addition indicated by the plus (+) symbol. In this case, the residual learning is implemented via an identity shortcut connection which skips two convolutional layers. As such, the output of the above ResNet is $x_{i+1} = \mathcal{F}(x_i) + x_i$, where $\mathcal{F}(x_i)$ represents the residual mapping. In the case where x_i and $\mathcal{F}(x_i)$ do not possess the same dimensions, an additional convolutional layer without any activation function is introduced before the sum.

However, Gulrajani et al. (2017) demonstrated that the use of weight clipping can lead to poor performance of WGANs in certain scenarios, such as optimization difficulties, which may be mitigated via batch normalization (Ioffe and Szegedy, 2015), although this does not guarantee convergence of very deep WGAN critics as illustrated in their work. Gulrajani et al. (2017) therefore came up with an alternative in the form of a gradient penalty in the loss function. This obviates the undesirable behaviour induced by weight clipping, while yielding substantial performance improvements.

Since a differentiable function is 1-Lipschitz if and only if the norm of its gradient is at most 1 everywhere, Gulrajani et al. (2017) proposed to directly constrain the gradient norm of the critic's output with respect to its input. The Lipschitz constraint is hence imposed by penalizing the gradient norm for random samples $\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}$, where $\hat{\boldsymbol{x}} = \epsilon \boldsymbol{x} + (1 - \epsilon)\tilde{\boldsymbol{x}}$ and ϵ is sampled randomly and uniformly, $\epsilon \in [0, 1]$, resulting in the following augmented objective:

$$\mathcal{L} = \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathbb{P}_z} \left[\mathcal{C}_\epsilon(\mathcal{G}_\theta(\boldsymbol{z})) \right] - \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathbb{P}_r} \left[\mathcal{C}_\epsilon(\boldsymbol{x}) \right] + \lambda \mathop{\mathbb{E}}_{\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}} \left[\left(||\nabla_{\hat{\boldsymbol{x}}} \mathcal{C}_\epsilon(\hat{\boldsymbol{x}})||_2 - 1 \right)^2 \right], \tag{12.1}$$

where λ is an arbitrary penalty coefficient and $\lambda = 10$ has been shown to work well for a range of architectures and data sets (Gulrajani et al., 2017). Essentially, we must introduce a gradient penalty term in the original critic loss, which forces the gradient of the critic network to remain close to unity. Alternatively, interpreting the loss function as the opposite of the log-likelihood, this term states that not more than a fluctuation of $\sim \sqrt{1/10}$ compared to one is allowed for the norm of the gradient.

A key advantage of the gradient penalty alternative is that it yields stable gradients which allows training of deep and complex networks without requiring *ad hoc* solutions such as batch normalization.

12.3 Cosmological simulations

In this work, we adopt the reference element of the VELMASS cosmological simulation suite. The VELMASS suite is comprised of 10 cosmological simulations, 9 of which are probing slightly different variations of a selection of cosmological parameters whilst using the same initial phases. The 10^{th} simulation has the same parameter values as the central simulation described below, but with different initial phases, such that it can be independent from the other simulations, allowing us to perform blind model comparison. The simulation that we use in this work assumes a Planck-like cosmology (Planck Collaboration et al., 2016b) with $\Omega_{\text{m}} = 0.315$, $\Omega_{\text{b}} = 0.049$, $H_0 = 68\,\text{km s}^{-1}\text{Mpc}^{-1}$, $\sigma_8 = 0.81$, $n_{\text{s}} = 0.97$ and $Y_{\text{He}} = 0.248$ (named "central" or Ω simulation). The power spectrum is obtained through the analytic prescription of Eisenstein and Hu (1999), and the initial conditions were generated by MUSIC (Hahn and Abel, 2011). This simulation suite is designed to test the robustness of analysis tools to acceptable variations in cosmology.

The cosmological simulation covers a volume of $2000h^{-1}$ Mpc with 2048^3 particles tracing dark matter. It was initialized at a redshift $z = 50$ and evolved to present time with GADGET2 (Springel, 2005), adopting a softening length for gravity equal to $48h^{-1}$ kpc corresponding to $1/20$ of the mean interparticle separation. The ROCKSTAR halo finder algorithm (Behroozi, Wechsler, and Wu, 2013) was subsequently employed to extract the halos from the simulation and generate the 3D halo field. The halo identification involves finding particles belonging to regions for which the local density is above a specific threshold, in our case as derived by the Friend-Of-Friend linking length, 45 times the mean density. Sub-halos are found by reducing recursively the linking length to find more compact structure. Each halo/sub-halo is pruned if it holds less than 10 particles as they are considered to be unstable. ROCKSTAR further performs a test for each particle to check if it is gravitationally bound to the structure. If this is not the case, the particles are removed from the halo. Once all these procedures are done, we are left with 23% of the total mass in structures considered as virialized by ROCKSTAR. We histogram the halo counts onto a grid of 512^3 which is the resolution we choose to work at, i.e. each voxel has a side length of $L \approx 4h^{-1}$ Mpc. The halos were selected in four equally spaced logarithmic bins in the mass range: $10^{12} - 10^{14}\ h^{-1}\ \text{M}_\odot$. Finally, we also produce the result of a pure second order Lagrangian perturbation theory (2LPT) simulation by setting the redshift for which MUSIC must create initial conditions to $z = 0$. We build the corresponding gridded density fluctuation field by applying the cloud-in-cell (CIC) algorithm (e.g. Hockney and Eastwood, 1988) to the particle distribution at the same resolution that we chose for the halo distribution, i.e. each voxel with length $L \approx 4h^{-1}$ Mpc.

We can test the cosmological dependence of the halo painting network by complementing the Ω simulation with other variants obtained assuming different cosmological parameters, for testing purposes. The initial random phases are kept the same and only the effects induced by different cosmological parameters are introduced. In this work, we concentrate on two additional 2048^3 cosmological simulations with $\Omega_{\text{m}} = 0.355$ and $\Omega_{\text{m}} = 0.275$. We repeat the entire procedure, above, for these two simulations, i.e. execution of the N-body simulation, identification of halos, construction of the 2LPT field and both resampling the halos and performing the CIC of the 2LPT field onto a 512^3 grid.

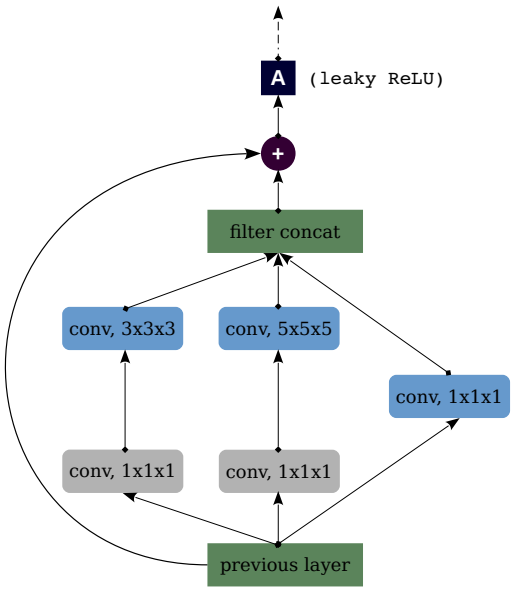


FIGURE 12.3: Residual Inception block, as implemented in this work, for the halo painting network architecture. Our Inception module is a slightly modified version of the original architecture (cf. Fig. 12.1). In particular, we replace all 2D convolutional layers with their 3D counterpart and the only dimensionality reduction comes from not padding the input when performing the convolutions. All convolutional layers employ 10 filters, with a `leaky ReLU` activation implemented after the residual connection.

12.4 Model architecture and training methodology

Our halo painting network is built to perform the map between the dark matter distribution and the halo count distribution. This allows us to use physical intuition to guide the model architecture. Considering the halo mapping network, we input a patch of the 2LPT field directly, unlike the usual conception of a GAN where a flat array of noise is normally used as the input. We are therefore physically mapping from a well-understood distribution of dark matter instead of a latent space. Although our halo painting network is not a generator in the classical machine learning sense, we refer to it as the "generator" below to make the analogy with the WGAN training routine. The "generator" terminology is also justified because the mapping network transforms a complex density field sampled from a Gaussian random field transformed by the 2LPT dynamics into a halo field with even less trivial statistical properties.

Knowing that information from the dark matter field is aggregated from a relatively local patch via some non-linear process allows us to choose a simple form for the generator network, a schematic of which is shown in Fig. 12.4. We connect the local region of the dark matter field using a 3D modification of the residual Inception block, shown in Fig. 12.3. Since the receptive field of 5×5 convolution kernel is $20h^{-1}$Mpc per side, then performing a second $5 \times 5 \times 5$ convolution on the output of the first increases the receptive field to $\sim 40h^{-1}$Mpc per side. In fact, since the kernels are 3D, the furthest distance that information can be propagated from the dark matter field to the halo distribution is $70h^{-1}$Mpc across the diagonal of such stacked kernels, whilst still learning the small scale features from the $1 \times 1 \times 1$ kernels and the mid-scale features with the $3 \times 3 \times 3$ kernels. This distance is far enough that the halo field should be insensitive to influences at such a scale. Since we use residual connections in the Inception blocks, we combine structure from distant patches, whilst

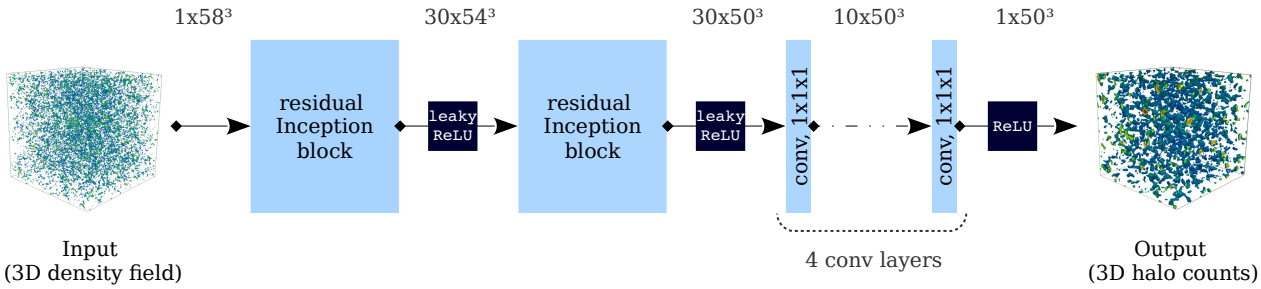


FIGURE 12.4: Schematic representation of the halo painting network implemented in this work. The network consists of two residual Inception blocks (cf. Fig. 12.3), followed by a series of four convolutional layers with $1 \times 1 \times 1$ kernel, across feature space as in the gray box in the Inception module. The activation function used is a leaky rectified linear (`leaky ReLU`) unit, except for the output layer where a rectified linear (`ReLU`) activation ensures non-negative halo counts in the generated 3D halo field. The input to the halo mapping network is a 3D realization of density field, with the output being the corresponding halo count distribution. The input is conveniently chosen to be larger to eliminate the need for padding. In the schematic, we indicate the size of the tensors used during training. We can use any input density fields sampled on a mesh with side $N_{\rm density} > 25$ which will predict a halo field sampled on a smaller mesh with side $N_{\rm halo} = N_{\rm density} - 8$.

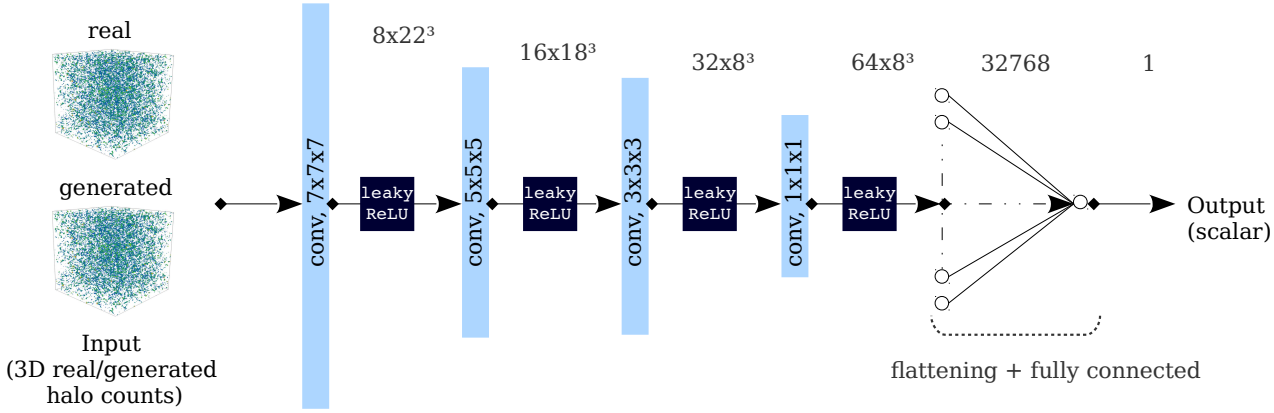


FIGURE 12.5: Schematic representation of the critic employed in this study. The network encodes a series of four convolutional layers, gradually reducing their respective kernel sizes from $7 \times 7 \times 7$ to $1 \times 1 \times 1$, and activated with `leaky ReLU`, with the corresponding output flattened and fed into a fully connected layer with linear activation. The critic reduces the input real and generated halo fields each to a compact scalar representation whose difference is an approximation to the Wasserstein distance.

still retaining a close relation to the density field itself. To learn the non-linear process from this receptive density field, we use four convolution layers with kernel size $1 \times 1 \times 1$ with no residual connection, which provides the non-linearity necessary to combine the local density distribution and perform the map to the distribution of halo counts. Since there is an enormous complexity in the 2LPT field, we use many filters in each layer to learn the wide variety of possible features. Every single kernel in our network has 10 filter channels to provide an extremely large path to build the complex non-linear map. For such a generator, we have 31,930 trainable parameters which is relatively few (in machine learning terms). The non-linearity is provided by the popular `leaky ReLU` activation function with a leaky parameter of $\alpha = 0.1$. To ensure the positivity of the halo count field, we use `ReLU` at the last layer.

In principle, we can train the generator network using 2LPT density patches as small as 9^3 voxels to predict a single halo count distribution voxel (due to using convolutions with no padding),

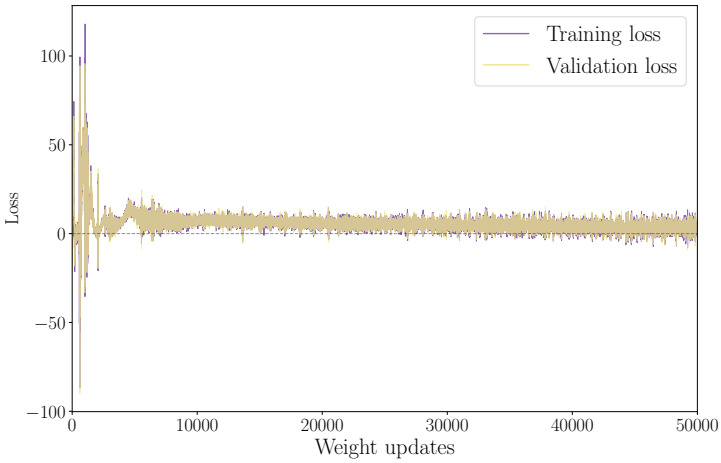


FIGURE 12.6: Training and validation loss for our physical mapping network for the first 5×10^4 weight updates. These respective losses are an approximation to the Wasserstein distance, as measured by the critic. As expected, this distance tends to zero as training proceeds and our halo painting network effectively learns the mapping from the dark matter to halo distribution. We note that the validation loss is closely tracking the training loss. As it can be seen, the two curves are mostly overlapping in the above plot, except at the very edge.

although the learned kernels, and therefore the predicted halo count field, would not be sensitive to any information from outside of this region. The maximum extent of the generator comes from the size of the kernels in the two residual Inception blocks, which is 25^3 voxels of the density field to predict a 17^3 halo patch. To increase the number of features to attempt to learn at once, we actually select 58^3 voxels from the gridded 2LPT field to predict a 50^3 voxel halo count distribution.

A major advantage to this prescription of building the generator in such a way is that we can train the network using small simulations and predict massive halo fields provided with large, cheap 2LPT fields. For example, we could, in parallel, make millions of 58^3 voxel cosmological simulations and run the halo finder on each for the training data, which is relatively quick in comparison to running an extremely large, say 2048^3 voxel simulation. Then, using the trained generator, we could predict *any* size halo field just by providing the 2LPT calculation, which is relatively cheap compared to performing the same sized simulation. Such a large 2LPT density field slab with $508 \times 508 \times 58$ voxels is used to predict a $500 \times 500 \times 50$ halo distribution slab and the projection is shown in Fig. 12.7 for the central slice of depth $\sim 100h^{-1}$ Mpc and side length of $\sim 2000h^{-1}$ Mpc.

Provided with our generator network which during training, as described above, will take 58^3 voxel 2LPT fields to paint 50^3 voxel halo count distributions, we now need to build a critic network to measure the distance between the painted halo counts and the corresponding real halo counts from the VELMASS simulation. Our critic, as depicted in Fig. 12.5, utilizes a series of four convolutional layers, while gradually reducing their respective kernel sizes from $7 \times 7 \times 7$ to $1 \times 1 \times 1$, and activated with `leaky ReLU` ($\alpha = 0.1$), with the output of the last convolutional layer flattened and fed into a fully connected layer with linear activation. The critic encodes relevant information from the input real or predicted halo fields into compact representations, thereby reducing the size of the 3D distributions. The output of the critic is a single scalar which is used to compute the approximately learned Wasserstein distance between the predicted and true halo distributions given a particular generative network. This output can therefore be used to compute the loss function

(11.2) which is minimized to train the generative network. We implement the above networks and the training routine outlined below in TENSORFLOW (Abadi et al., 2016).

During training, we load both the entire 2LPT density field and histogrammed halo count distribution from the Ω VELMASS simulation into the TENSORFLOW graph and select by index subvolume elements of size 58^3 and 50^3, respectively, which massively reduces computation time compared with passing the 3D slices of data at each weight update. These patches corresponds to side lengths of $L \approx 225h^{-1}$ Mpc and $L \approx 200h^{-1}$ Mpc, respectively. The input to the generator is randomly chosen and the corresponding true halo counts volume is selected. Note that here, the batch size is unity, such that the number of weight updates corresponds to the number of density field patches used to train the generator. We use a 512^3 simulation box for training, where we use a large portion of the box for training and the remaining section for validation, such that we utilize non-mutual parts of the box for the training and validation set. To encode some further symmetries through our training set, we also perform a rotation of the selected patches, thereby extracting the input 3D slice from a randomly oriented region. The generator employs the gradient of the Wasserstein loss function (11.2) with respect to its parameters θ for training.

The initial training step involves the optimization of the weights of the critic network to minimize the augmented loss function (12.1), while concurrently freezing the parameters of the generator. The weights of the critic must be updated $n_{\rm critic}$ times, where $n_{\rm critic}$ is sufficient for the critic to converge. The samples for this initial step are randomly selected (and rotated) 2LPT fields and corresponding true halo counts. In the subsequent step, the critic weights are temporarily anchored, and the generator parameters are adjusted. The training routine then proceeds in iterative fashion, until an overall convergence of the generator is achieved. The training rationale is to reduce the Wasserstein distance between the true halo counts and the halo counts mapped from the corresponding input density field such that the generator gradually learns the correct mapping. The training procedure outlined above is represented schematically in Fig. 11.2.

In this work, we use $n_{\rm critic} = 5$ and set the arbitrary coefficient for the gradient penalty to $\lambda = 10$, with ϵ having a randomly and uniformly drawn value $0 \leq \epsilon \leq 1$. We implement the popular *Adam* (Kingma and Ba, 2014) optimization algorithm, with a learning rate of 10^{-4} and first and second moment exponential decay rates of 0.5 and 0.999, respectively. We trained the network for $\sim 5 \times 10^5$ generator weight updates for the different mass bins, requiring around 30 hours on a NVIDIA Quadro P6000. The training and validation loss for the first 5×10^4 weight updates of our halo painting model is illustrated in Fig. 12.6. This is the approximately learnt Wasserstein distance which tends to zero as training proceeds and the generator effectively learns the mapping from the dark matter to halo distribution.

12.5 Results

Fig. 12.7 depicts the reference and predicted halo fields, for a 3D slice of depth $\sim 100h^{-1}$ Mpc and side length of $\sim 2000h^{-1}$ Mpc, and the difference between the reference and predicted fields, obtained from a completely unseen simulation. The corresponding 2LPT density field is also shown for the sake of completeness. Qualitative agreement is impressive, implying that the halo painting

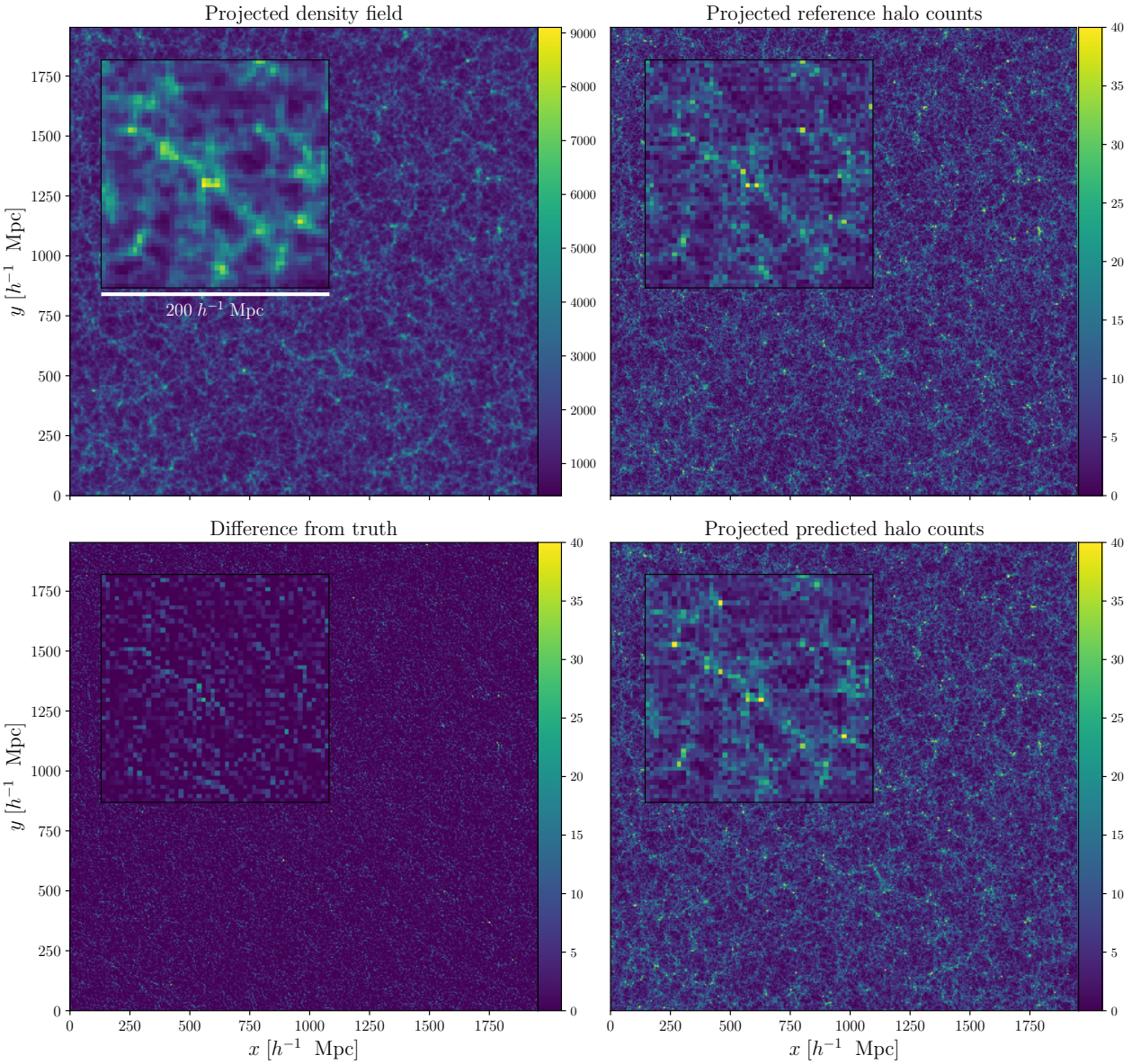


FIGURE 12.7: Prediction of 3D halo field by our halo painting model for a slice of depth $\sim 100h^{-1}$ Mpc and side length of $\sim 2000h^{-1}$ Mpc. A blind validation dataset is shown in the top right panel, with the predicted halo count depicted below it. The corresponding 2LPT density field is displayed in the top left panel, with the difference between the reference and predicted halo distributions depicted in the lower left panel. A visual comparison of the reference and predicted halo count distributions indicates qualitatively the efficacy of our halo painting network. Note that we did not normalize the 2LPT density field in this work, which renders the performance of our network even more remarkable.

network is capable of mapping the complex structures of the cosmic web, such as halos, filaments and voids, to the corresponding distribution of halo counts. We now assess and validate the performance of our halo painting model using quantitative diagnostics.

We illustrate the conditional probability distribution of the predicted halo count per voxel given the respective true (or reference) value in the left panel of Fig. 12.8, with the right panel depicting the distribution of the difference between the reference and predicted values. The conditional probability distribution, which accounts for the intrinsic halo distribution, indicates that our predictions fare well in lower and average density environments (counts less than 10) while it overshoots when

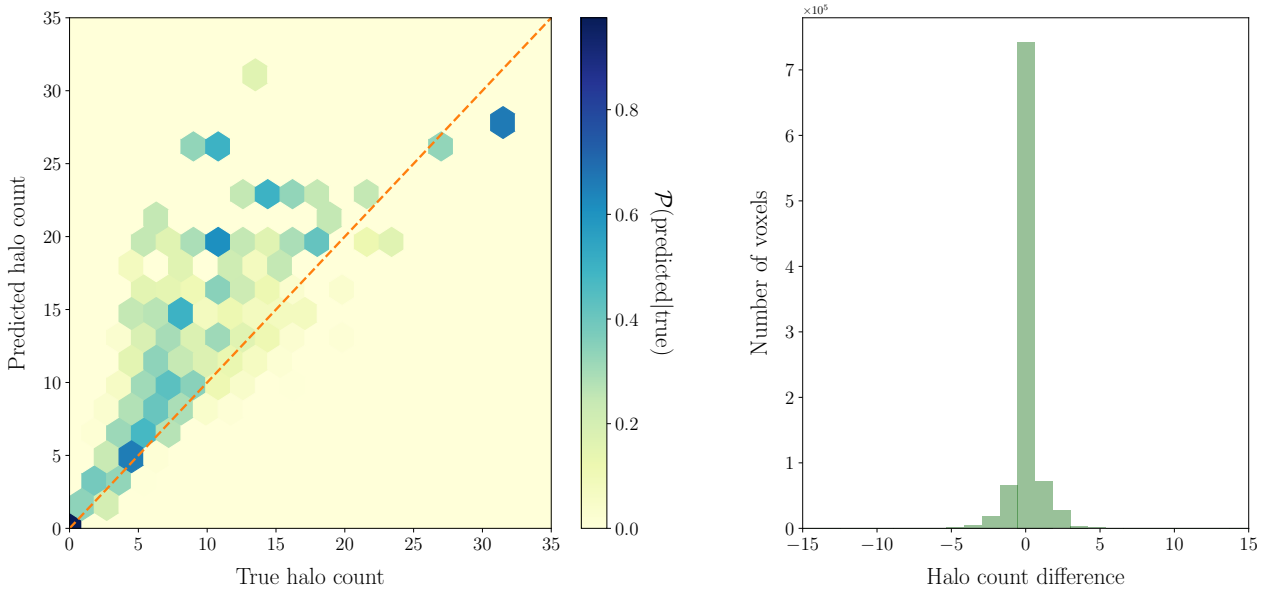


FIGURE 12.8: *Left panel:* The conditional probability distribution of the predicted halo count per voxel given the corresponding true (i.e. reference) value. This properly represents the error in our prediction, while accounting for the intrinsic halo distribution. Our network predictions fare well in lower and average density environments with halo counts less than 10, but overshoots when the halo count is higher. There are, however, very few regions which reach such high halo count ($\lesssim 0.03\%$), such that our predictions are not often skewed by such rare occurrences. *Right panel:* The distribution of the difference between the reference and predicted halo counts per voxel. This shows that the difference is close to zero for the majority of the voxels, with the difference being larger than 3 for only $\sim 1.5\%$ of the total number of voxels.

the halo count is high (counts greater than 10). But it is important to note that there are very few regions which reach such high halo count ($\lesssim 0.03\%$) and thus our network predictions are not often skewed by such rare occurrences. The distribution of the difference shows that our network predictions for most voxels closely match the corresponding true values, with the difference being larger than 3 for only $\sim 1.5\%$ of the total number of voxels. We now assess and validate the performance of our halo painting model using other quantitative diagnostics.

12.5.1 Two-point correlation and power spectrum

As quantitative assessment, we employ summary statistics, as per the standard practice in cosmology. These summary statistics provide a reliable metric to gauge our halo painting network in terms of their capacity to encode essential information.

The two-point correlation function, denoted by $\xi(r)$, is the quintessential measure employed by cosmologists, along with its Fourier transform, the power spectrum $P(k)$, defined as follows:

$$\xi(|\boldsymbol{r}|) = \langle \delta(\boldsymbol{r}')\delta(\boldsymbol{r}'+\boldsymbol{r})\rangle \tag{12.2}$$

$$P(|\boldsymbol{k}|) = \int \mathrm{d}^3\boldsymbol{r}\, \xi(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{12.3}$$

where the δ's correspond to the field relative contrast, i.e. $\delta(\boldsymbol{r}) = \rho(\boldsymbol{r})/\bar{\rho} - 1$, with $\rho(\boldsymbol{r})$ the matter density and $\bar{\rho}$ the mean matter density. Assuming the cosmological density field is approximately a

Gaussian random field, as is the case on the large scales or at earlier times, the above two statistics provide a sufficient description of the field.

We first consider the power spectra of the reference and predicted halo distributions, denoted by $P_{\rm hh}(k)$ and $P_{\rm pp}(k)$, as a standard measure for the description of the matter and halo distributions. The indices p and h label the predicted and true (reference) halo modes, respectively. The top left panel of Fig. 12.9 illustrates their respective mean and 1σ confidence regions for one thousand randomly selected patches with side lengths of $L \approx 400h^{-1}$ Mpc from a blind set of data, thereby quantitatively showcasing the remarkable performance of the halo painting model in learning the map from 2LPT density fields to halo number counts.

We also investigate some standard diagnostics at the level of two-point correlation functions in Fourier space to evaluate the model performance against the ground truth. The three metrics considered, each dependent on the mode k, are as follows:

1. Transfer function ($T(k)$)

$$T(k) \equiv \sqrt{\frac{P_{\rm pp}(k)}{P_{\rm hh}(k)}} \tag{12.4}$$

2. Cross correlation coefficient ($r_{\rm c}$)

$$r_{\rm c}(k) = \frac{P_{\rm hp}(k)}{\sqrt{P_{\rm hh}(k)P_{\rm pp}(k)}} \tag{12.5}$$

3. Stochasticity ($s(k)$)

$$s(k) = P_{\rm hh}(k)(1 - r_{\rm c}(k)^2), \tag{12.6}$$

where, as before, $P_{\rm hh}(k)$ and $P_{\rm pp}(k)$ are the auto power spectra of the reference and predicted halo fields, respectively, while $P_{\rm hp}(k)$ corresponds to the cross power spectrum of the reference and predicted halo field. The transfer function, defined as the square root of the ratio of the two auto power spectra, indicates the discrepancy in amplitudes, as a function of the Fourier modes, whilst the cross correlation coefficient characterizes the mismatch in phases between the real and predicted halo count fields. These two diagnostics quantify the predictive capability of our halo painting network, while the fraction of the variance that cannot be accounted for in the true model is encoded in the stochasticity.

Fig. 12.9 displays the scale dependence of the above summary statistics for the reference and predicted halo count fields with one thousand independent randomly selected patches, from an unseen simulation, in terms of their respective mean and 1σ confidence regions. The power spectra of the predicted halo fields match extremely closely that of the reference halo distributions. The cross correlation coefficient is close to unity on the large scales and drops to $r_{\rm c} \approx 0.7$ at $k \approx 0.4$. The turnover, at $k \approx 0.25$, is due to the grid resolution, and the prediction in the shaded area is not

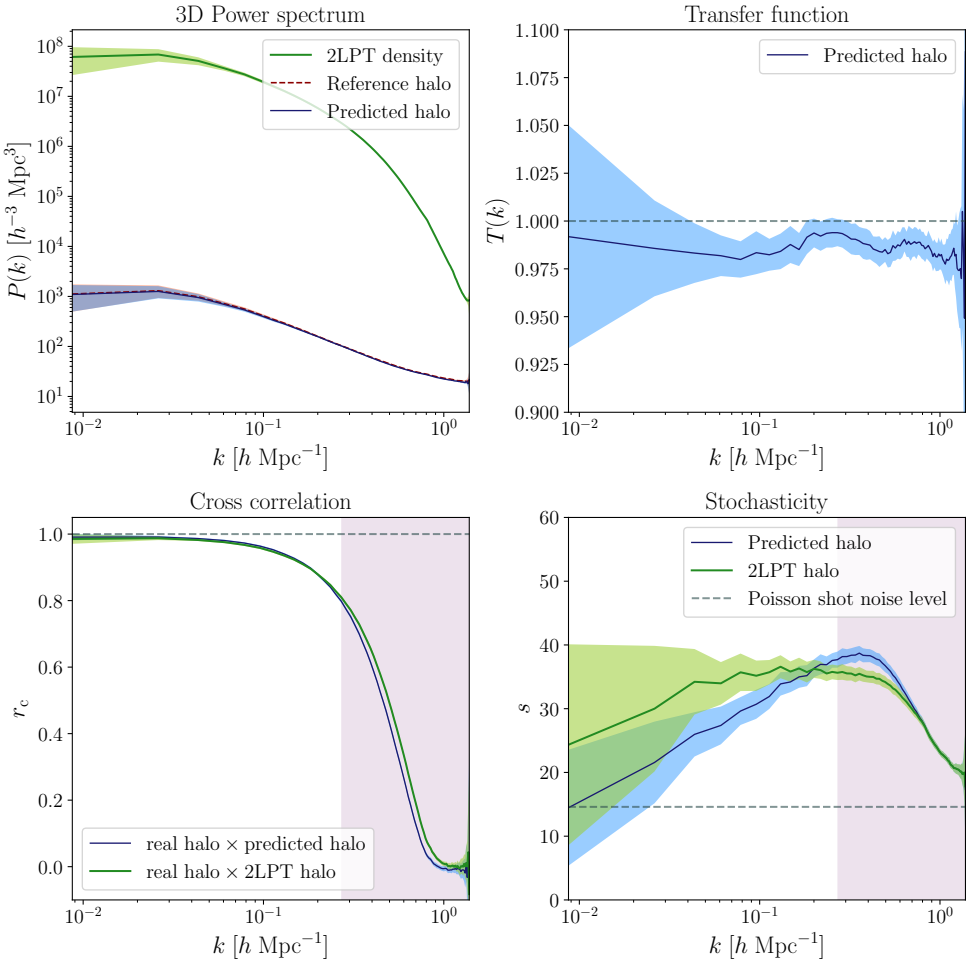


FIGURE 12.9: Summary statistics of the 3D power spectra of the density, reference and predicted halo fields for one thousand randomly selected patches. The solid lines indicate their respective means, while the shaded regions indicate their respective 1σ confidence regions, i.e. 68% probability volume. The above diagnostics demonstrate the ability of our halo painting model to reproduce the characteristic statistics of the reference halo fields and therefore provide substantial quantitative evidence for the performance of our neural network in mapping 3D density fields to their corresponding halo distributions. Note that the turnover in the cross-correlation coefficient is due to the grid resolution, and the prediction in the shaded area is redundant in practice. The behaviour of the cross-correlation coefficient and stochasticity of the halo field derived linearly from 2LPT, illustrated in the bottom panels, shows that our network predictions reproduce the reference halo field with overall higher fidelity, at least up to the regime limited by the grid resolution.

taken into account in practice. The above two-point summary statistics showcase the performance of our network in predicting the entire halo distribution. We also make a comparison with the corresponding diagnostics of the 2LPT halo field which corresponds to a statistical description of the halo distribution, derived from the 2LPT density field, which is valid, by construction, at the two-point level and on large scales. It is built by scaling the 2LPT density field by the adequate linear bias factor so that we match the power spectrum of the halo field on large scales. The behaviour of the cross-correlation coefficient and stochasticity of the 2LPT halo field, illustrated in the bottom panels of Fig. 12.9, shows that our network predictions reproduce the reference halo field with overall higher fidelity, at least up to the regime limited by the grid resolution. On large scales, the ensemble mean stochasticity is lower by typically a factor of two, and its variance is smaller

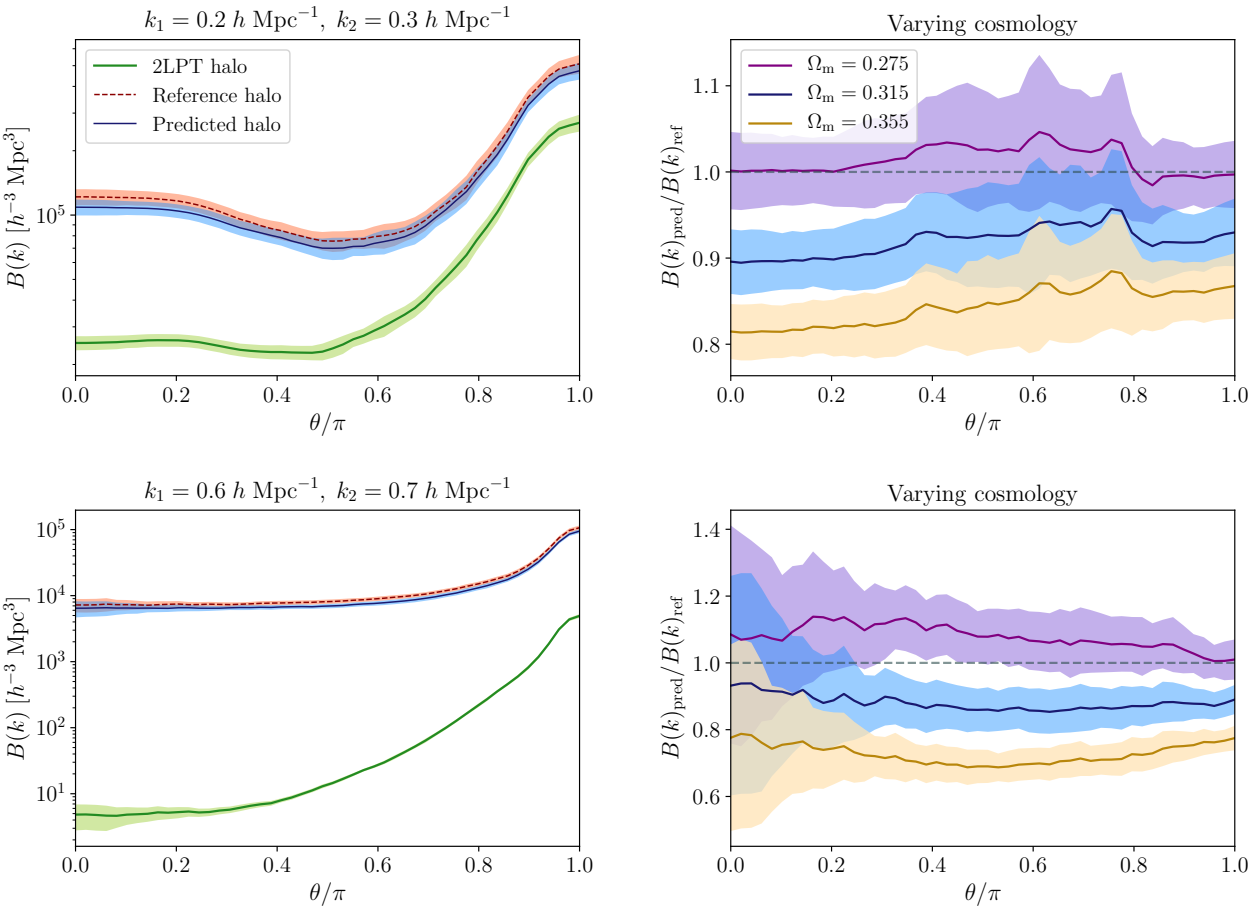


FIGURE 12.10: *Left panels:* Summary statistics of the 3D bispectra of the 2LPT, reference and predicted halo fields for a given small- and large-scale configurations, $k_1 = 0.6\ h\ \mathrm{Mpc}^{-1}$ and $k_2 = 0.7\ h\ \mathrm{Mpc}^{-1}$, $k_1 = 0.2\ h\ \mathrm{Mpc}^{-1}$ and $k_2 = 0.3\ h\ \mathrm{Mpc}^{-1}$, respectively. In both cases, there is a close agreement between the bispectra from the reference and predicted halo distributions. In particular, our network predictions are a significant improvement over the corresponding 2LPT halo fields. *Right panels:* The deviation from the 3D bispectra of the reference halo distributions of the corresponding predictions for the two cosmology variants. The above bispectrum diagnostics show that our network is more sensitive to the fiducial cosmology than at the level of the two-point correlation function. The 1σ confidence regions for five hundred randomly selected patches are depicted in each panel.

by about 30%. On small scales, which are not covered by our predictor, the two estimates behave mostly in a similar way. Over the entire scale range, our predictor fails to go below the level given by the mean Poisson shot noise. We will investigate in the future how to improve the current situation, particularly on large scales as shot noise should not be a limitation given that the entire procedure is fully deterministic, both on the N-body simulation and the neural network sides.

12.5.2 Three-point correlation and bispectrum

The non-linear dynamics involved in gravitational evolution of cosmic structures contributes to a certain degree of non-Gaussianity of the cosmic density field on the small scales. Higher-order statistics are therefore required to characterize this non-Gaussian field. We employ the bispectrum, i.e. the Fourier transform of the three-point correlation function, to quantify the spatial distribution

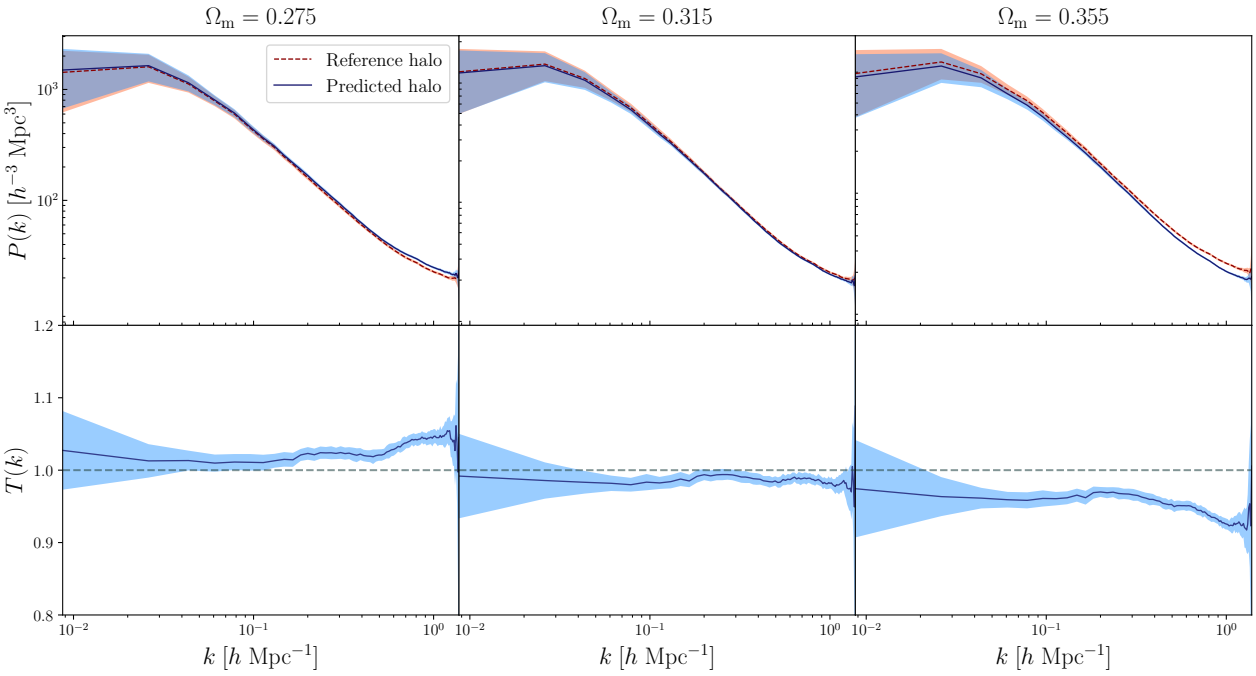


FIGURE 12.11: *Top panels:* The corresponding power spectra, as in the top left panel of Fig. 12.9, for the fiducial and two cosmology variants, with 1σ confidence regions for one thousand randomly selected patches. The top left and right panels demonstrate the capability of our halo painting network to reproduce the two-point summary statistics when applied to simulations generated with the fiducial cosmology. *Bottom panels:* The corresponding transfer functions highlight the consistency between the power spectra reconstructed from the predicted and real halo fields for the three cosmology variants, with the deviation from their respective reference spectra being below 10%. The above diagnostic therefore shows that our halo painting model is slightly sensitive to the underlying cosmology at the level of the power spectrum.

of the density field, defined as:

$$(2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)B(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) = \langle\hat{\delta}(\boldsymbol{k}_1)\hat{\delta}(\boldsymbol{k}_2)\hat{\delta}(\boldsymbol{k}_3)\rangle, \tag{12.7}$$

where δ_D is the Dirac delta. The bispectra reconstructed from the 2LPT, reference and predicted halo fields are displayed in Fig. 12.10. In particular, we show the bispectra for a given small- and large-scale configurations. The construction of the 2LPT halo field is described above in Section 12.5.1. This allows us to make a fair comparison between the clustering of the respective halo fields. The left panel demonstrates that our halo painting network reproduces the non-linear halo field both on the small and large scales, and is therefore capable of mapping the complex cosmic structures apparent in the reference halo field. The agreement between the reference and predicted halo bispectra on the largest scales is slightly suboptimal as both our halo painting and critic networks are not tailored to produce structure on this level. Our network predictions also show a significant improvement over the corresponding 2LPT halo fields. The above bispectrum computations were performed using the publicly available PYLIANS[1] code.

12.5.3 Dependence on cosmology

We investigate the influence of the fiducial cosmology adopted for the simulations on the efficacy of our halo mapping model. We depict in Fig. 12.11 the network predictions for two cosmology variants in terms of their respective two-point summary statistics. The corresponding transfer functions, depicted in the bottom panels, show a deviation of about 10% from the reference power spectra of their respective real halo distributions on the smallest and largest scales. The right panels of Fig. 12.10 illustrate the ratio of the predicted to reference bispectra for the two configurations displayed in the left panels. We find that there is a more significant dependence of our network on the fiducial cosmology at higher order statistics.

The above diagnostics therefore demonstrate a certain degree of sensitivity on the underlying cosmology, as expected, since the non-linearly evolved 2LPT density field does not capture completely the inherent cosmological dependence, in accordance with the results from He et al. (2019). However, we expect that this sensitivity to the cosmology may be reduced by training over a range of cosmologies, thereby rendering our halo painting network more resilient to cosmological priors. We defer such an investigation to a future work.

12.5.4 Individual mass bins

To verify whether the network has properly encoded the halo mass information, we investigate the mapping learned for the different mass bins. We depict in Figs. 12.12 and 12.13, respectively, the corresponding power and bispectrum diagnostics for the four individual mass bins. Their respective predicted power and bispectra demonstrates the exceptional performance of our halo painting network across the range of halo masses involved. The above diagnostics show that the network is also capable of predicting the mass distribution of halos.

12.6 Summary and conclusions

We have presented a novel halo painting network[2] for mapping 3D density fields to dark matter halo fields, inspired by an implementation of a recent variant of generative adversarial networks which employs the Wasserstein distance as a metric for training the network. Our network architecture encodes some of the recently proposed refinements to optimize its effectiveness and efficiency. The painting network employs residual Inception blocks, combining residual networks (ResNets) and Inception architecture, a choice which was heavily inspired by our physical notion of the true process. For improved training performance, we implement the gradient penalty method via the addition of a penalty term in the critic loss, as an alternative to the standard weight clipping, to enforce the Lipschitz-1 constraint on the critic. Our neural network, as a result of these upgrades, is not prone to training instability issues and does not exhibit any undesired behaviour.

We trained our halo painting network on 2LPT simulations to infer the relationship between the dark matter density field and the final halo distribution. We showcased the performance of our

[1] Available from https://github.com/franciscovillaescusa/Pylians

[2] The source code repository, including a JUPYTER notebook tutorial, is available at https://github.com/doogesh/halo_painting

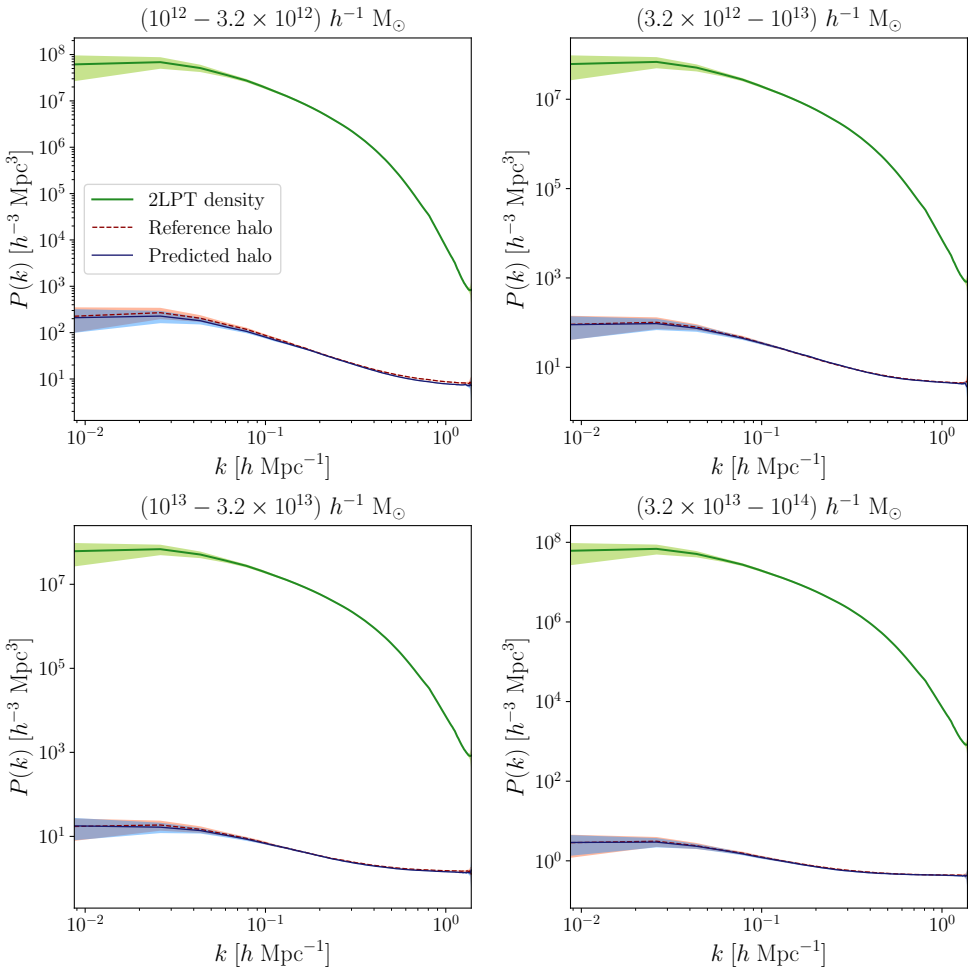


FIGURE 12.12: The corresponding power spectra, as in the top left panel of Fig. 12.9, for the individual mass bins, with 1σ confidence regions for one thousand randomly selected patches. The above diagnostic, at the level of two-point statistics, highlights the remarkable performance of our halo painting network across the range of halo masses involved. As such, it can also predict the mass distribution of halos.

network in predicting the 3D halo distribution via a series of diagnostics, at the level of two- and three-point summary statistics, demonstrating that this mapping can be learned to a sufficiently high level of accuracy and that it reduces substantially the stochasticity over the Poisson shot noise. This performance is especially remarkable given that our neural network has only $\sim \mathcal{O}(10^4)$ trainable parameters, which is relatively few parameters in machine learning terms. In essence, our halo painting model allows us to rapidly generate simulations of halo distribution based on a non-linearly evolved density field within a fraction of a second on modern GPUs. For instance, the network prediction for a 256^3 simulation size requires roughly one second on the NVIDIA Quadro P6000.

A crucial aspect of our halo painting network, in a nutshell, lies in its capability to paint a halo distribution from a computationally cheap non-linear density field. This, as a result, provides a deterministic transformation to statistically populate the density field with highly non-linear structures such as halos. The halo painting network, therefore, bypasses the need to run full particle

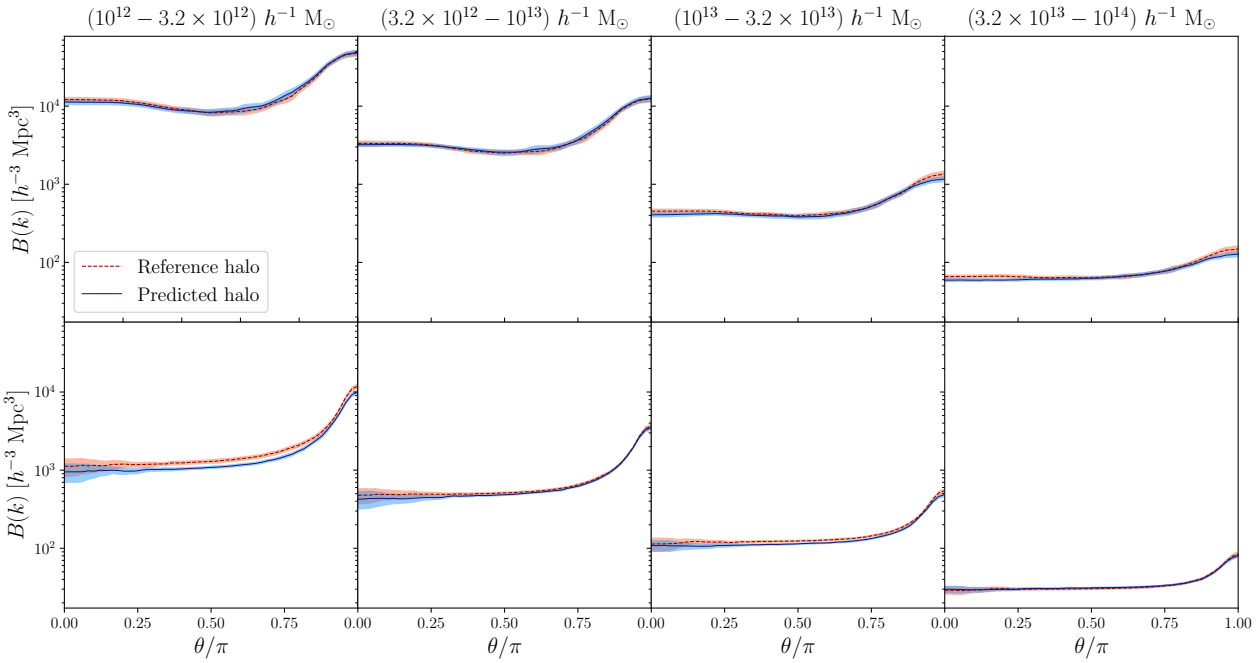


FIGURE 12.13: Bispectrum diagnostics for the individual mass bins, for five hundred randomly selected patches. The top and bottom panels depict the large and small configurations, respectively, as considered in the left panels of Fig. 12.10. At the level of three-point statistics, we find that our network performs remarkably well across the range of halo masses involved.

mesh simulations, thereby ensuring that detailed and high-resolution analyses of current and next-generation galaxy surveys, via the forward modelling approaches outlined below, are still feasible on regular computing clusters. Another interesting advantage of our approach is that our network can predict the 3D halo distribution for any arbitrary simulation box size due to the convolutional kernels being translationally invariant. With this method, a large simulation box does not require the tiling of smaller sub-elements, rendering our approach simple and elegant.

An immediate application of our halo painting network is that it can be employed for fast generation of mock halo catalogues and light cone production. This would be especially useful for the data analysis of upcoming large galaxy surveys of unprecedented sizes, such as Euclid and LSST. Another potential application of our network is that it can be utilized to fill in small-scale structure at a high resolution from low resolution large-scale simulations. It is also worth investigating whether we can further improve the performance of our painting model by training on the displacement field, rather than the density field, as carried out in He et al. (2019). We defer such an investigation to a future undertaking. We also intend to explore other avenues, in terms of network architecture and training methodology, to render our halo painting model robust to different cosmologies, such that it could be optimized to constrain cosmological parameters via hierarchical Bayesian models, such as the ALTAIR (ALcock-Paczyński consTrAIned Reconstruction) algorithm (Kodi Ramanah et al., 2019).

The technology developed here is relevant to Bayesian forward modelling approaches for large-scale structure inference. For instance, it may be incorporated as part of the forward model in the BORG (Bayesian Origin Reconstruction from Galaxies) (Jasche and Wandelt, 2013a; Jasche, Leclercq, and Wandelt, 2015; Lavaux and Jasche, 2016; Jasche and Lavaux, 2019) framework to generate

sharp features due to redshift space distortions, such as Fingers of God effects. Since dark matter dynamics is well-posed, such a neural network may be employed as an emulator to transform an approximate large-scale model, such as Lagrangian Perturbation Theory (LPT), into a 3D halo distribution. Halo masses can subsequently be remapped to galaxy masses via another deterministic function, while encoding a stochastic selection, yielding the galaxy distribution of interest.

Bayesian inference methods often require the adjoint gradient of the forward model in the sampling procedure, and as such, neural networks are perfectly suited for such tasks since they are, by their very nature, fully differentiable. The combination of statistical inference and machine learning is the plausible approach to accelerate high-resolution analyses of upcoming galaxy redshift surveys and provide statistically interpretable results, while maintaining the scientific rigour.

Part V

Summary, conclusions and outlook

Key aspects of doctoral work

In essence, this thesis presents the development of sophisticated Bayesian statistical inference and machine learning techniques to extract cosmological information from the CMB polarization and large-scale structures of the Universe. We showcased the efficacy and performance of the precise tools developed for optimal exploitation of next-generation CMB missions and galaxy redshift surveys to answer key unresolved questions at the frontiers of research in cosmology.

The Wiener filter is a powerful signal reconstruction tool that incorporates statistical information about the signal and noise, with widespread cosmological and astrophysical applications. This is nevertheless a highly non-trivial problem when dealing with large and complex data sets, as the solution requires the inversion of dense matrices. In Chapter 2, we developed a high-performance hierarchical method, known as the dual messenger algorithm, which guarantees numerical stability and convergence, while obviating matrix inversions. As a proof of concept, we applied the algorithm to maps of CMB temperature anisotropies in Chapter 3, showing that it outperforms a conventional preconditioned conjugate gradient (PCG) method in terms of execution time, while matching its efficacy of reconstruction.

In Chapter 4, we presented an extension to our dual messenger algorithm for the reconstruction of polarized CMB maps. Unlike standard PCG solvers, our method can deal with high-resolution joint temperature and polarization maps with relative ease. We showed how the algorithm can be modified to generate fluctuation maps, which, combined with the Wiener filter solution, yield unbiased constrained signal realizations, consistent with observed data. This generation of Gaussian constrained realizations of the CMB sky is an essential component of present-day CMB analyses. We demonstrated the performance of our algorithm for $\mathcal{E}/\mathcal{B}$ separation using various convergence diagnostics and illustrated the difficulties of standard PCG techniques in effectively treating this fundamentally ill-conditioned problem.

In Chapter 5, we presented an augmented version of our dual messenger algorithm for spin field reconstruction on the sphere, while accounting for highly non-trivial and realistic noise models such as anisotropic correlated noise. We also described an optimization method for the estimation of noise covariance from Monte Carlo simulations. Using simulated Planck polarized CMB maps as a showcase, we demonstrated the capabilities of the algorithm in reconstructing pure $\mathcal{E}$ and $\mathcal{B}$ maps, guaranteed to be free from ambiguous modes resulting from the leakage or coupling issue that plagues conventional methods of $\mathcal{E}/\mathcal{B}$ separation. Our algorithm, designated as DANTE (DuAl messeNger filTEr), therefore, has a potentially key role in the data analysis of high-resolution and high-sensitivity CMB data, especially with the range of upcoming CMB experiments tailored for the detection of the elusive primordial $\mathcal{B}$-mode signal.

In Chapter 8, we presented an extension to the Bayesian inference framework of BORG (Bayesian Origin Reconstruction from Galaxies), originally developed for the non-linear 3D reconstruction of large-scale structures, to constrain cosmological parameters using galaxy redshift surveys via a novel application of the Alcock-Paczyński (AP) test. Our inference machinery allows the joint reconstruction of the non-linearly evolved density field as well as the initial conditions and expansion history of the large-scale structures within the observed domain, while fully accounting for all uncertainties inherent to the analysis. Our augmented Bayesian hierarchical framework, designated

as ALTAIR (ALcock-Paczyński consTrAIned Reconstruction), exploits the high information content of the cosmic expansion, extracting several orders of magnitude more information compared to standard approaches. The novelty of this AP test lies in constraining the comoving-redshift transformation to infer the appropriate cosmology which yields isotropic correlations of the cosmic initial conditions, with the underlying assumption relying purely on the geometrical symmetries of the cosmological principle. Such an AP test does not rely explicitly on modelling the full statistics of the field, with the robustness of our test to model misspecification verified via simulations. This leads to another crucial advantage, namely that the cosmological parameters exhibit extremely weak dependence on the currently unresolved phenomenon of galaxy bias, thereby circumventing a potentially significant limitation. This is consequently among the first methods to extract a large fraction of information from statistics other than that of direct density contrast correlations, without being sensitive to the amplitude of density fluctuations. Using a mock SDSS-III survey as showcase, we demonstrated the performance of ALTAIR in inferring tight constraints on the matter density and dark energy equation of state.

In Chapter 9, we presented a novel likelihood which is robust to unknown systematic effects such as foreground and target contaminations. Conceptually, this robust likelihood marginalizes over the unknown large-scale contamination amplitudes. The primary motivation underlying this work is the fact that forthcoming surveys, such as LSST and Euclid, will be limited by systematic uncertainties. In order to make progress in recovering cosmological information from such surveys, it is essential to develop a data model to deal with such effects. We illustrated the effectiveness of our robust likelihood, implemented within the BORG framework, via an application on a mock SDSS-III catalogue subject to dust extinction contamination, while being maximally agnostic about the foregrounds. We demonstrated that it allows the recovery of unbiased power spectrum across all scales, unlike the standard Poissonian likelihood typically employed in large-scale structure analyses, which yields excessive power on large scales due to contamination from spurious modes. Robust likelihood approaches, as presented in this chapter, will be crucial to control unknown systematics and maximize the outcome of decadal galaxy surveys.

In Chapter 12, we presented a novel halo painting technique using physically motivated neural networks to map approximate 3D dark matter fields to realistic halo distributions. Using such an emulator, we can learn the non-trivial local relation between dark matter density field and halo distributions without relying on a physical model. In learning to paint halo distributions from computationally cheap, analytical and non-linear density fields, we bypass the need for full N-body simulations and halo finding algorithms. Furthermore, by design, our halo painting network needs only local patches of dark matter density to predict the halos, and as such, it can predict the 3D halo distribution for any arbitrary simulation box size. Our neural network can be trained using small simulations and used to predict large halo distributions, as long as the resolutions are equivalent. We assessed our model's ability to generate 3D halo count distributions which reproduce, to a high degree, summary statistics such as the power spectrum and bispectrum, of the reference realizations. Such complex dynamics emulators would substantially improve the efficiency of our above forward modelling approaches.

Outlook

There is scope for further development of the methods presented above, either to render them even more efficient or as preparation for actual data applications by incorporating other complexities not yet accounted for. We outline some potential extensions and scientific applications of our statistical inference and machine learning tools below. The ultimate goal inevitably involves the more interesting aspect of real data applications of these algorithms to obtain valuable scientific products from state-of-the-art CMB and galaxy redshift data sets, thereby contributing towards solving the unresolved mysteries at the frontiers of research in modern cosmology.

Exact global analyses of CMB polarization

Due to its high speed execution, coupled with lenient memory requirements, the dual messenger (DANTE) algorithm can be optimized in exact global Bayesian analyses, such as Gibbs sampling, of state-of-the-art CMB polarization data for a statistically optimal separation of pure $\mathcal{E}$ and $\mathcal{B}$ modes. The development of the Gibbs sampling pipeline is especially relevant with the latest and final release of Planck data (Planck Collaboration et al., 2018a) and the advent of next-generation CMB experiments, such as CMB-S4 and Simons Observatory. Gibbs sampling of full-resolution polarized CMB data will potentially yield scientific products of significant value and interest to the community of cosmologists. The reconstructed maps may subsequently be employed for various applications such as power spectrum reconstruction, estimation of lensing potential, tests for foreground contamination and searches for non-Gaussianity and statistical anisotropy such as hemispherical power anisotropy.

Constraints on early Universe physics via CMB polarization

The application of DANTE, encoded within a statistically optimal Gibbs sampling scheme, to polarization data from upcoming CMB missions with unprecedented sensitivity may lead to an eventual detection of the primordial $\mathcal{B}$ modes. This groundbreaking detection will unlock the gravitational wave window to the very early Universe and drastically improve our understanding of the currently obscure phases of primordial cosmology. The imprint of primordial gravitational waves on the degree-scale $\mathcal{B}$ modes of the CMB polarization would strongly validate the inflationary paradigm, while revealing the energy scale of cosmic inflation. This would potentially also yield some insights about the fundamental symmetries of nature and properties of quantum gravity.

Redshift space distortions and unknown systematic effects

The ALTAIR formalism can be augmented to account for redshift space distortions, which is essential to infer unbiased cosmological constraints. To render the inference from upcoming galaxy redshift surveys more robust, we may employ our novel likelihood, presented in Chapter 9, in ALTAIR to deal with unknown systematic effects such as foreground and target contaminations. With the control of systematics being a crucial limiting factor for modern galaxy surveys reaching unprecedented depths and sky coverages, our Bayesian forward modelling approach is prepared to cope with such challenges.

Dynamical dark energy, neutrino masses and test of gravity

Within the modular statistical programming scheme of ALTAIR, we may encode the parameter inference of the equation of state of distinct phenomenological dark energy models, which may shed some light on the nature of this obscure component driving the late-time cosmic accelerated expansion and yield a more precise picture of our cosmic expansion history. Another possibility worth exploring is the inference of neutrino masses, currently a major area of investigation. We may also constrain the cosmic growth rate, which would correspond to a test of gravity

Principled way of using machine learning

Amidst all the excitement pertaining to the unprecedented rise and ubiquity of machine learning in the scientific community, they are simply another addition to our cosmological toolbox. As such, we should harness the power of such algorithms to enhance our set of data analysis tools. However, given the standard of our current computing resources, machine learning may be considered as expedient means to tackle virtually any problem, which may not yield any deep insights and actually hinder the progress of our understanding. It is, therefore, essential to learn how to formalize and decipher the specific problem under consideration, and find the best possible route to the solution without incurring costs of interpretability. Scientific intuition and rigour remain, nevertheless, the pillars of advancement in science.

Fast complex dynamics emulators for Bayesian hierarchical models

In line with the above principles, we have designed a physically motivated neural network in Chapter 12 to model complex dynamics involved in cosmic structure evolution. An immediate extension of our network architecture is to augment low-resolution N-body simulations with high-resolution structures. The next step is to build our emulator in the forward model of ALTAIR, which would substantially improve the efficiency of our Bayesian hierarchical framework. The synergy between deep learning and Bayesian inference provides us with a principled approach to perform detailed and high-resolution analyses of upcoming galaxy surveys, which would yield statistically interpretable results with proper uncertainty quantification.

Appendix A

Convergence properties of the messenger methods

A.1 Cooling scheme for the messenger algorithm

In this section, we illustrate the rationale behind the cooling scheme for the scalar parameter λ adopted for the standard messenger technique. In the computations below, we assume invertibility of matrices everywhere. Using $\mathbf{T} = \alpha\mathbb{1}$, where $\alpha = \min(\mathrm{diag}(\mathbf{N}))$, the χ^2_T for the messenger scheme from Eq. (2.8) can be written as

$$\chi^2_T = (\boldsymbol{d} - \boldsymbol{s})^\dagger \left[\mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right]^{-1} (\boldsymbol{d} - \boldsymbol{s}) + \boldsymbol{s}^\dagger \mathbf{S}^{-1} \boldsymbol{s}. \tag{A.1}$$

Hence, in the messenger approach, we have $\mathbf{N} \to [\mathbf{N} + (\lambda - 1)\alpha\mathbb{1}]$ and plugging this into the Wiener filter given by Eq. (2.3) yields

$$\begin{aligned} \boldsymbol{s}_{\rm WF}(\lambda) &= \left\{ \left[\mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right]^{-1} + \mathbf{S}^{-1} \right\}^{-1} \left[\mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right]^{-1} \boldsymbol{d} \\ &= \mathbf{S} \left[\mathbf{S} + \mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right]^{-1} \boldsymbol{d}, \end{aligned} \tag{A.2}$$

where we assume that all matrices are invertible. As a consistency check, for $\lambda = 1$, Eq. (A.2) reduces to Eq. (2.3), the usual Wiener filter equation. The difference between the solutions at two consecutive values of λ is given by

$$\begin{aligned} \boldsymbol{s}_{\rm WF}(\lambda) - \boldsymbol{s}_{\rm WF}(\lambda') &= \mathbf{S} \left\{ \left[\mathbf{S} + \mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right]^{-1} - \left[\mathbf{S} + \mathbf{N} + (\lambda' - 1)\alpha\mathbb{1}\right]^{-1} \right\} \boldsymbol{d} \\ &= \mathbf{S}\left[\mathbf{S} + \mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right]^{-1} \left\{ \mathbf{S} + \mathbf{N} + (\lambda' - 1)\alpha\mathbb{1} - \left[\mathbf{S} + \mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right] \right\} \\ &\qquad \cdot \left[\mathbf{S} + \mathbf{N} + (\lambda' - 1)\alpha\mathbb{1}\right]^{-1} \boldsymbol{d} \\ &= \alpha(\lambda' - \lambda)\mathbf{S} \left[\mathbf{S} + \mathbf{N} + (\lambda - 1)\alpha\mathbb{1}\right]^{-1} \left[\mathbf{S} + \mathbf{N} + (\lambda' - 1)\alpha\mathbb{1}\right]^{-1} \boldsymbol{d}. \end{aligned} \tag{A.3}$$

Now, we consider limiting cases of Eq. (A.3). We assume a homogeneous noise distribution, i.e., $\mathbf{N} = \alpha\mathbb{1}$. From Eq. (A.2), we have

$$\|\boldsymbol{s}_{\rm WF}(\lambda)\|_\ell = C_\ell (C_\ell + \lambda\alpha)^{-1} \|\boldsymbol{d}\|_\ell. \tag{A.4}$$

Using Eqs. (A.3) and (A.4), the relative error can then be computed as follows:

$$\frac{\|\boldsymbol{s}_{\rm WF}(\lambda)-\boldsymbol{s}_{\rm WF}(\lambda')\|_\ell}{\|\boldsymbol{s}_{\rm WF}(\lambda)\|_\ell} \leq \left[|\lambda'-\lambda|\alpha\left(\frac{C_\ell}{C_\ell+\lambda\alpha}\right)\left(\frac{1}{C_\ell+\lambda'\alpha}\right)\|\boldsymbol{d}\|_\ell\right]\left[\frac{C_\ell}{(C_\ell+\lambda\alpha)}\|\boldsymbol{d}\|_\ell\right]^{-1}$$
$$\leq \frac{\alpha|\lambda'-\lambda|}{C_\ell+\lambda'\alpha}, \tag{A.5}$$

where, for an arbitrary matrix $\mathbf{M}$, $\|\mathbf{M}\|_\ell$ is the subset of $\|\mathbf{M}\|$ over subspace $\boldsymbol{\ell}$. From Eq. (2.11), we have the fractional error reduction, after one iteration, given by

$$1-\epsilon \leq \frac{C_\ell}{C_\ell+\lambda\alpha}\left(\frac{\bar{\mathbf{N}}}{\bar{\mathbf{N}}+\lambda\alpha}\right), \tag{A.6}$$

where the term in parentheses will always favour rapid convergence. The other term is strongly convergent on small scales, but to give bounds to the maximum convergence speed of "slow modes", we focus on the latter, leading to

$$\epsilon \leq 1-\frac{C_\ell}{C_\ell+\lambda\alpha}=\frac{\lambda\alpha}{C_\ell+\lambda\alpha}. \tag{A.7}$$

To favour convergence, from Eqs. (A.5) and (A.7), we require the two conditions:

$$\frac{\alpha|\lambda'-\lambda|}{C_\ell+\lambda'\alpha} \lesssim \begin{cases} \frac{\lambda\alpha}{C_\ell+\lambda\alpha} \\ \\ \frac{\lambda'\alpha}{C_\ell+\lambda'\alpha} \end{cases} \tag{A.8}$$

so that the error reduction in one iteration matches the change in solution arising from two consecutive values of λ. This results in the following two constraints: $|\lambda'-\lambda| \leq \lambda'$ and $|\lambda'-\lambda| \leq \lambda$. The former is automatically satisfied, while the latter leads to the following bounds: $\lambda'/2 \leq \lambda \leq \lambda'$. Therefore, for the cooling scheme, we can write $\lambda = \eta\lambda'$, where $1/2 \leq \eta \leq 1$, to improve convergence on all scales. This is the motivation behind our choice of $\eta = 3/4$ in this work.

A.2 Cooling scheme for the dual messenger algorithm

As in the previous section, we wish to analytically determine the cooling scheme that would improve convergence for the dual messenger algorithm. We first derive the Wiener filter solution for the dual messenger, i.e., the analogue of Eq. (A.2) for this scheme. We begin by writing down Eqs. (2.14) and (2.15), showing the dependence on the power spectrum truncation μ:

$$\left(\mathbf{N}^{-1}+\mathbf{U}_\mu^{-1}\right)\boldsymbol{s}_{\rm WF}(\mu) = \mathbf{N}^{-1}\boldsymbol{d}+\mathbf{U}_\mu^{-1}\boldsymbol{u}_\mu \tag{A.9}$$
$$\left(\mathbf{U}_\mu^{-1}+\bar{\mathbf{S}}_\mu^{-1}\right)\boldsymbol{u}_\mu = \mathbf{U}_\mu^{-1}\boldsymbol{s}_{\rm WF}(\mu). \tag{A.10}$$

Solving for $\boldsymbol{s}_{\rm WF}(\mu)$ via the following steps:

$$\begin{aligned}\mathbf{N}^{-1}\boldsymbol{d} &= \left[(\mathbf{N}^{-1}+\mathbf{U}_\mu^{-1}) - \mathbf{U}_\mu^{-1}(\mathbf{U}_\mu^{-1}+\bar{\mathbf{S}}_\mu^{-1})^{-1}\mathbf{U}_\mu^{-1}\right]\boldsymbol{s}_{\rm WF}(\mu)\\ &= \left[(\mathbf{N}^{-1}+\mathbf{U}_\mu^{-1})(\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu) - \bar{\mathbf{S}}_\mu\mathbf{U}_\mu^{-1}\right](\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)^{-1}\boldsymbol{s}_{\rm WF}(\mu)\\ &= \left[\mathbf{N}^{-1}(\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)+\mathbb{1}\right](\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)^{-1}\boldsymbol{s}_{\rm WF}(\mu)\\ &= \left[\mathbf{N}^{-1}+(\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)^{-1}\right]\boldsymbol{s}_{\rm WF}(\mu), \end{aligned} \tag{A.11}$$

so that finally, we have

$$\boldsymbol{s}_{\rm WF}(\mu) = \left[\mathbf{N}^{-1}+(\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)^{-1}\right]\mathbf{N}^{-1}\boldsymbol{d}. \tag{A.12}$$

In the limit $\mu \to \nu$, where $\nu \equiv \min(\mathrm{diag}(\mathbf{S}))$, $\bar{\mathbf{S}}+\mathbf{U} \to \mathbf{S}$, reducing Eq. (A.12) to the usual Wiener filter Eq. (2.3), implying consistency. When we change the truncation of the spectrum, $\mu \to \tilde{\mu}$, for $\tilde{\mu} < \mu$, we obtain a new solution $\boldsymbol{s}_{\rm WF}(\tilde{\mu})$. The counterpart of Eq. (A.5) is then

$$\begin{aligned}\boldsymbol{s}_{\rm WF}(\tilde{\mu}) - \boldsymbol{s}_{\rm WF}(\mu) &= (\bar{\mathbf{S}}_{\tilde{\mu}}+\mathbf{U}_{\tilde{\mu}})\left[(\bar{\mathbf{S}}_{\tilde{\mu}}+\mathbf{U}_{\tilde{\mu}})+\mathbf{N}\right]^{-1}\boldsymbol{d} - (\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)\left[(\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)+\mathbf{N}\right]^{-1}\boldsymbol{d}\\ &= \left[(\bar{\mathbf{S}}_{\tilde{\mu}}+\mathbf{U}_{\tilde{\mu}})(\mathbf{N}+\bar{\mathbf{S}}_{\tilde{\mu}}+\mathbf{U}_{\tilde{\mu}})^{-1} - (\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)(\mathbf{N}+\bar{\mathbf{S}}_\mu+\mathbf{U}_\mu)^{-1}\right]\boldsymbol{d}. \end{aligned} \tag{A.13}$$

Here, we consider Fourier space, so that $\mathbf{U}_{\tilde{\mu}} \to \tilde{\mu}$ and $\mathbf{U}_\mu \to \mu$. We can write the truncated signal covariance matrix as

$$\bar{\mathbf{S}}_{\tilde{\mu}} = \bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu}\mathbb{1}, \tag{A.14}$$

where $\Delta_{\tilde{\mu},\mu} = \tilde{\mu} - \mu$ is the portion of the power spectrum bounded by $\tilde{\mu}$ and μ, while the corresponding truncated signal covariances can be represented by Heaviside functions as follows:

$$\bar{\mathbf{S}}_\mu = \Theta(\mathbf{S}-\mu) \tag{A.15}$$

$$\bar{\mathbf{S}}_{\tilde{\mu}} = \Theta(\mathbf{S}-\tilde{\mu}), \tag{A.16}$$

where, for a matrix $\mathbf{M} = \mathbf{P}\Lambda\mathbf{P}^{-1}$, after applying a basis transformation, with Λ being diagonal,

$$\Theta(\mathbf{M}) = \mathbf{P}\Theta(\Lambda)\mathbf{P}^{-1}, \tag{A.17}$$

and

$$\Theta(\Lambda)_{ii} = \begin{cases} 0, & \Lambda_{ii} \leq 0 \\ \Lambda_{ii}, & \Lambda_{ii} > 0. \end{cases} \tag{A.18}$$

Using Eq. (A.14) in Eq. (A.13) results in

$$\begin{aligned}
s_{\mathrm{WF}}(\tilde{\mu}) - s_{\mathrm{WF}}(\mu) &= \left[(\bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu})(\mathbf{N} + \bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu})^{-1} - (\bar{\mathbf{S}}_\mu - \mu)(\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}\right] \boldsymbol{d} \\
&= \Big[\bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu} - (\bar{\mathbf{S}}_\mu + \mu)(\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}(\mathbf{N} + \bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu})\Big] \\
&\qquad \cdot (\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu} - \mu)^{-1}\boldsymbol{d} \\
&= \Big[\bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu} - \bar{\mathbf{S}}_\mu - \mu - (\bar{\mathbf{S}}_\mu + \mu) + (\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}(\Delta_{\tilde{\mu},\mu} + \tilde{\mu} - \mu)\Big] \\
&\qquad \cdot (\mathbf{N} + \bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu})^{-1}\boldsymbol{d} \\
&= \left[(\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu) - \bar{\mathbf{S}}_\mu - \mu\right](\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}(\Delta_{\tilde{\mu},\mu} + \tilde{\mu} - \mu) \\
&\qquad \cdot (\mathbf{N} + \bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu})^{-1}\boldsymbol{d} \\
&= \mathbf{N}(\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}(\Delta_{\tilde{\mu},\mu} + \tilde{\mu} - \mu)(\mathbf{N} + \bar{\mathbf{S}}_\mu + \Delta_{\tilde{\mu},\mu} + \tilde{\mu})^{-1}\boldsymbol{d}. \qquad \text{(A.19)}
\end{aligned}$$

Using Eq. (A.14) and substituting the following form of Eq. (A.12),

$$s_{\mathrm{WF}}(\tilde{\mu}) = \bar{\mathbf{S}}_{\tilde{\mu}}\left[\mathbf{N} + (\bar{\mathbf{S}}_{\tilde{\mu}} + \tilde{\mu})^{-1}\right]^{-1}\boldsymbol{d}, \qquad \text{(A.20)}$$

in Eq. (A.19) leads to

$$s_{\mathrm{WF}}(\tilde{\mu}) - s_{\mathrm{WF}}(\mu) = \mathbf{N}(\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}(\Delta_{\tilde{\mu},\mu} + \tilde{\mu} - \mu)(\bar{\mathbf{S}}_{\tilde{\mu}} + \tilde{\mu})^{-1}s_{\mathrm{WF}}(\tilde{\mu}), \qquad \text{(A.21)}$$

such that finally we obtain the following equation for the relative error:

$$\frac{\|s_{\mathrm{WF}}(\tilde{\mu}) - s_{\mathrm{WF}}(\mu)\|}{\|s_{\mathrm{WF}}(\tilde{\mu})\|} \leq \left\|\mathbf{N}(\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}(\Delta_{\tilde{\mu},\mu} + \tilde{\mu} - \mu)(\bar{\mathbf{S}}_{\tilde{\mu}} + \tilde{\mu})^{-1}\right\|. \qquad \text{(A.22)}$$

Now, for $\mu \to \tilde{\mu}$, there are three distinct regimes of relevance and we consider each case below. To investigate the convergence behaviour in the different regimes, we make use of the following equation, obtained by plugging Eqs. A.15) and (A.16) in Eq. (A.14),

$$\Theta(\mathbf{S} - \tilde{\mu}) = \Theta(\mathbf{S} - \mu) + \Delta_{\tilde{\mu},\mu}. \qquad \text{(A.23)}$$

The three regimes are:

- $\mathbf{S}_\ell > \mu\mathbb{1}$,
  where $\Theta(\mathbf{S} - \tilde{\mu}) = \mathbf{S}_\ell - \tilde{\mu}\mathbb{1}$, and $\Theta(\mathbf{S} - \mu) = \mathbf{S}_\ell - \mu\mathbb{1}$, so that $\Delta_{\tilde{\mu},\mu} = \mu - \tilde{\mu}$, and this causes the second term in Eq. (A.22) to vanish, implying that the relative error is not affected by our choice of $\tilde{\mu}$. This shows that the Wiener filter solution is naturally computed in a hierarchical fashion using the dual messenger algorithm.

- $\tilde{\mu}\mathbb{1} < \mathbf{S}_\ell < \mu\mathbb{1}$,
  where $\Theta(\mathbf{S} - \tilde{\mu}) = \mathbf{S}_\ell - \tilde{\mu}\mathbb{1}$, and $\Theta(\mathbf{S} - \mu) = 0$, so that $\Delta_{\tilde{\mu},\mu} = \mathbf{S}_\ell - \mu\mathbb{1}$, leading to the following relative error:

$$\frac{\|s_{\mathrm{WF}}(\tilde{\mu}) - s_{\mathrm{WF}}(\mu)\|}{\|s_{\mathrm{WF}}(\tilde{\mu})\|} \leq \left\|\mathbf{N}(\mathbf{N} + \bar{\mathbf{S}}_\mu + \mu)^{-1}\right\| \left\|(\mathbf{S}_\ell - \mu\mathbb{1})(\bar{\mathbf{S}}_{\tilde{\mu}} + \tilde{\mu})^{-1}\right\|_{\tilde{\mu},\mu}. \qquad \text{(A.24)}$$

The first term behaves as a constant, i.e., $\left\|\mathbf{N}(\mathbf{N}+\bar{\mathbf{S}}_{\mu}+\mu)^{-1}\right\| \sim \alpha'$, so that the relative error can be approximated as

$$\begin{aligned}\frac{\|\boldsymbol{s}_{\mathrm{WF}}(\tilde{\mu})-\boldsymbol{s}_{\mathrm{WF}}(\mu)\|}{\|\boldsymbol{s}_{\mathrm{WF}}(\tilde{\mu})\|} &\leq \alpha' \left|\frac{C_\ell-\mu}{C_\ell}\right|_{\tilde{\mu},\mu} \\ &= \alpha' \left|1-\frac{\mu}{C_\ell}\right|_{\tilde{\mu},\mu},\end{aligned} \tag{A.25}$$

since $\bar{\mathbf{S}}_{\tilde{\mu}}+\tilde{\mu} \sim C_\ell$. Also, $\alpha' \sim 1$, and to favour convergence, we want this change to be as large as possible but sufficiently small such that iterating the solution results in rapid decay of the error, i.e., the change in the solution due to changing μ should be matched to the change when iterating the solution. If $\mu = \beta C_\ell$, choosing $0 < \beta < 1$ would therefore improve convergence by avoiding the early freeze of modes in the iteration scheme. This served as the basis for our choice of $\beta = 3/4$ in this work.

- $\mathbf{S}_\ell < \tilde{\mu}\mathbb{1}$,
  where $\Theta(\mathbf{S}-\tilde{\mu}) = 0 = \Theta(\mathbf{S}-\mu)$, so that $\Delta_{\tilde{\mu},\mu} = 0$, and again using the approximation $\bar{\mathbf{S}}_{\tilde{\mu}}+\tilde{\mu} \sim C_\ell$ yields
  $$\frac{\|\boldsymbol{s}_{\mathrm{WF}}(\tilde{\mu})-\boldsymbol{s}_{\mathrm{WF}}(\mu)\|}{\|\boldsymbol{s}_{\mathrm{WF}}(\tilde{\mu})\|} \leq \alpha' \left|\frac{\tilde{\mu}-\mu}{C_\ell}\right|_{\tilde{\mu},\mu}. \tag{A.26}$$

So, the overall convergence behaviour can be quantitatively described by:

$$\frac{\|\boldsymbol{s}_{\mathrm{WF}}(\tilde{\mu})-\boldsymbol{s}_{\mathrm{WF}}(\mu)\|}{\|\boldsymbol{s}_{\mathrm{WF}}(\tilde{\mu})\|} \leq 0 + \alpha' \left|1-\frac{\mu}{C_\ell}\right|_{\tilde{\mu},\mu} + \alpha' \left|\frac{\tilde{\mu}-\mu}{C_\ell}\right|_{\tilde{\mu},\mu}, \tag{A.27}$$

and this leads to the following interpretation: For higher values of μ on large scales, the third term in Eq. (A.27) dominates since $|\tilde{\mu}-\mu|$ is large. But for the final truncations, $\tilde{\mu} \sim \mu$, so this regime is saturated, while on small scales, the second term dominates as $\mu \gg C_\ell$, so the relative error continues to drop till the final truncation. The cooling scheme described above applies naturally to the hybrid version of the dual messenger method.

Appendix B

Supplementary information for spin field reconstruction

B.1 Preconditioner for conjugate gradient descent

For the PCG computation in Chapter 4, we implement a straightforward generalization of the diagonal preconditioner adopted in KLW17 (cf. Chapter 2.4). Essentially, the Wiener filter Eq. (2.3) is reformulated as $\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{y}$, where $\boldsymbol{\mathcal{A}} = \mathbb{1} + \mathbf{S}^{1/2}\mathbf{N}^{-1}\mathbf{S}^{1/2}$, $\boldsymbol{x} = \mathbf{S}^{-1/2}\boldsymbol{s}_{\rm WF}$ and $\mathbf{y} = \mathbf{S}^{1/2}\mathbf{N}^{-1}\boldsymbol{d}$. Solving for $\boldsymbol{x}$ requires the inverse of $\boldsymbol{\mathcal{A}}$, which can be approximated by a preconditioner $\boldsymbol{\mathcal{M}}$, i.e. $\boldsymbol{\mathcal{M}} \approx \boldsymbol{\mathcal{A}}^{-1}$.

Here, the preconditioner $\boldsymbol{\mathcal{M}}$ has the following block-diagonal form, for all multipole moments ℓ:

$$\mathcal{M}_\ell = \begin{pmatrix} 1 + \psi C_\ell^{TT} & 1 + \psi C_\ell^{TE} & 0 \\ 1 + \psi C_\ell^{TE} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{B.1}$$

with $\psi = L^{-4} N_{\rm pix} \left(\mathcal{F}^{-1} \mathbf{N}_{\mathcal{I}}^{-1} \right)_{\ell=0}$, where $\mathbf{N}_{\mathcal{I}}$ is the noise covariance associated with the Stokes parameter $\mathcal{I}$.

B.2 Spherical harmonic transforms

We provide a brief description of the transformation between pixel and spherical harmonic domain in order to be precise about the notation employed in Chapter 5.

Assuming the primary CMB fluctuations to be an isotropic Gaussian random field, the CMB signal can be described as a vector of spherical harmonic coefficients, with the associated signal covariance $\mathbf{S}$ given by:

$$S_{\ell m, \ell' m'} = \langle a_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell, \tag{B.2}$$

where C_ℓ is the CMB power spectrum. The proper basis to represent isotropic Gaussian random fields on the sphere is described by spherical harmonics. Given a grid on the sphere, i.e. a set of pixel positions $\hat{n}_p$, we can transform a field expressed in spherical harmonic (SH) basis, with coefficients $s_{\ell m}$, to one sampled on the sphere via SH synthesis, as follows:

$$s(\hat{n}_p) = \sum_{\ell=0}^{\ell_{\rm max}} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n}_p) = \sum_{\ell=0}^{\ell_{\rm max}} \sum_{m=-\ell}^{\ell} \mathcal{Y}_p^{\ell m} a_{\ell m}. \tag{B.3}$$

Formally, the SH synthesis may be expressed as a matrix product, $\boldsymbol{s}_{(p)} = \boldsymbol{\mathcal{Y}}\boldsymbol{a}_{(\ell m)}$, where $\boldsymbol{\mathcal{Y}}$ is the synthesis operator that encodes the value of the SHs evaluated at each $\hat{n}_p$ of the grid.

Conversely, the transformation from pixel to harmonic basis is referred to as SH analysis, with the analysis operator being an integral related to the synthesis operator:

$$\boldsymbol{a}_{(\ell m)} \simeq \sum_p Y^*_{\ell m}(\hat{n}_p) s(\hat{n}_p) \delta\Omega_p = \boldsymbol{\mathcal{Y}}^{-1}\boldsymbol{s}_{(p)} = \sum_p \frac{4\pi}{N_{\rm pix}} \mathcal{Y}^{\dagger,\ell m}_p s(\hat{n}_p) = \frac{4\pi}{N_{\rm pix}} \boldsymbol{\mathcal{Y}}^\dagger \boldsymbol{s}_{(p)}. \tag{B.4}$$

It is important to emphasize the scaling operation above and note that the last equalities are valid only for an equal-area pixelization such as HEALPix (Górski et al., 2005). These spherical harmonic transforms are the spherical analogue of Fourier transforms.

B.3 Differential operators

We define the following complex differential operators acting on a spin-s quantity f_s on the sphere:

$$\eth f_s \equiv -(\sin\theta)^s \left(\partial_\theta + i\frac{\partial_\varphi}{\sin\theta}\right)(\sin\theta)^{-s} f_s(\theta,\varphi) \tag{B.5}$$

$$\bar{\eth} f_s \equiv -(\sin\theta)^{-s} \left(\partial_\theta - i\frac{\partial_\varphi}{\sin\theta}\right)(\sin\theta)^{s} f_s(\theta,\varphi), \tag{B.6}$$

where $\eth$ and $\bar{\eth}$ are the spin raising and lower operators, respectively, such that $\eth f_s$ is a spin-$(s+1)$ quantity and $\bar{\eth} f_s$ is a spin-$(s-1)$ quantity. The general definition of the spin-weighted spherical harmonics with spin $s \geq 0$ is as follows:

$${}_sY_{\ell m} \equiv \beta_{\ell,s}\eth^s Y_{\ell m},\ {}_{-s}Y_{\ell m} \equiv (-1)^s \beta_{\ell,s}\bar{\eth}^s Y_{\ell m}, \tag{B.7}$$

where $\beta_{\ell,s} \equiv \sqrt{(\ell-s)!/(\ell+s)!}$, with $Y_{\ell m}$ being the standard spherical harmonics. The property of $({}_sY_{\ell m})^* = (-1)^{(s+m)}_{\ \ -s} Y_{\ell -m}$ is satisfied by these functions.

The spherical harmonic vectors ${}_s\mathbf{Y}^{\mathcal{E},\mathcal{B}}_{\ell m}$ for the $\mathcal{E}$ and $\mathcal{B}$ modes, and their associated differential operators $\mathbf{D}^{\mathcal{E},\mathcal{B}}_s$, respectively, are defined as

$${}_s\mathbf{Y}^{\mathcal{E}}_{\ell m} \equiv \mathbf{D}^{\mathcal{E}}_s Y_{\ell m} \equiv -\frac{\beta_{\ell,s}}{2}\left(\frac{\eth^s + \bar{\eth}_s}{-i(\eth^s - \bar{\eth}_s)}\right) Y_{\ell m} = -\frac{1}{2}\left(\frac{{}_sY_{\ell m} + (-1)^s\ {}_{-s}Y_{\ell m}}{-i[{}_sY_{\ell m} - (-1)^s\ {}_{-s}Y_{\ell m}]}\right) \tag{B.8}$$

$${}_s\mathbf{Y}^{\mathcal{B}}_{\ell m} \equiv \mathbf{D}^{\mathcal{B}}_s Y_{\ell m} \equiv -\frac{\beta_{\ell,s}}{2}\left(\frac{i(\eth^s + \bar{\eth}_s)}{\eth^s + \bar{\eth}_s}\right) Y_{\ell m} = -\frac{1}{2}\left(\frac{i[{}_sY_{\ell m} - (-1)^s\ {}_{-s}Y_{\ell m}]}{{}_sY_{\ell m} + (-1)^s\ {}_{-s}Y_{\ell m}}\right). \tag{B.9}$$

For a spin-0 quantity ($s = 0$), the above functions simplify to: ${}_0\mathbf{Y}^{\mathcal{E}}_{\ell m} = (Y_{\ell m}, 0)$ and ${}_0\mathbf{Y}^{\mathcal{B}}_{\ell m} = (0, Y_{\ell m})$.

B.4 Dual messenger formalism for modulated correlated noise covariance

We provide a more in-depth derivation of the two dual messenger Eqs. (5.21) and (5.22) required for the treatment of modulated correlated noise covariance in Chapter 5. The third Eq. (5.20) can be

derived from Eq. (5.19) in straightforward fashion via linear algebraic simplifications.

Eq. (5.2) can be written in its explicit form as:

$$\boldsymbol{u} = \left[\mathbf{B}^\dagger \boldsymbol{\mathcal{Y}}^\dagger \mathbf{T}^{-1} \boldsymbol{\mathcal{Y}} \mathbf{B} + (\bar{\mathbf{S}} + \mathbf{U})^{-1}\right]^{-1} \mathbf{B}^\dagger \boldsymbol{\mathcal{Y}}^\dagger \mathbf{T}^{-1} \boldsymbol{t}, \tag{B.10}$$

with the covariance of the messenger field $\boldsymbol{t}$ being $\mathbf{T} = \mathbf{D}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)\mathbf{D}$. This equation bears a striking resemblance to the standard Wiener filter Eq. (2.3) and can therefore be solved via the introduction of an extra messenger field $\boldsymbol{v}$ with covariance $\mathbf{V} = \omega(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)\mathbb{1}$, where $\omega \equiv \min(\mathrm{diag}(\mathbf{D}^2))$, resulting in the following χ^2:

$$\chi^2_V = (\boldsymbol{t} - \boldsymbol{v})^\dagger \left[\mathbf{D}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)\mathbf{D} - \mathbf{V}\right](\boldsymbol{t} - \boldsymbol{v}) + (\boldsymbol{v} - \boldsymbol{\mathcal{Y}}\mathfrak{b}\boldsymbol{u})^\dagger \mathbf{V}^{-1}(\boldsymbol{v} - \boldsymbol{\mathcal{Y}}\mathfrak{b}\boldsymbol{u}) + \boldsymbol{u}^\dagger(\bar{\mathbf{S}} + \mathbf{U})^{-1}\boldsymbol{u}. \tag{B.11}$$

Minimizing the above χ^2 with respect to $\boldsymbol{v}$ and $\boldsymbol{u}$ yields the following set of equations:

$$\boldsymbol{v} = \omega(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)\mathbf{D}^{-1}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1}\mathbf{D}^{-1}\left[\boldsymbol{t} + \omega^{-1}\mathbf{D}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)\mathbf{D}(\boldsymbol{\mathcal{Y}}\phi\boldsymbol{\mathcal{Y}}^\dagger)^{-1} - \mathbb{1}\right]\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u} \tag{B.12}$$

$$\boldsymbol{u} = \left[\phi\omega(\bar{\mathbf{S}} + \mathbf{U})^{-1} + \mathbf{B}^\dagger\mathbf{B}\right]^{-1}\mathbf{B}^\dagger\boldsymbol{\mathcal{Y}}^\dagger\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{v}, \tag{B.13}$$

where, as before, we employ the definition of the coupling matrix $\boldsymbol{\mathcal{M}} \equiv \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^\dagger$. It is more convenient to work with $\tilde{\boldsymbol{t}} \equiv \mathbf{D}^{-1}\boldsymbol{t}$, such that we can rewrite Eq. (B.12) as:

$$\boldsymbol{v} = \omega\boldsymbol{\mathcal{M}}\mathbf{D}^{-1}\boldsymbol{\mathcal{M}}^{-1}\tilde{\boldsymbol{t}} + \left[\mathbb{1} - \omega\boldsymbol{\mathcal{M}}\mathbf{D}^{-1}\boldsymbol{\mathcal{M}}^{-1}\mathbf{D}^{-1}\right]\boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u}. \tag{B.14}$$

The preference for the above form is that masked regions do not pose any numerical issue, as $\mathbf{D}^{-1}|_{\mathrm{mask}} = \mathbf{0}$, such that $\boldsymbol{v}|_{\mathrm{mask}} \to \boldsymbol{\mathcal{Y}}\mathbf{B}\boldsymbol{u}$.

Bibliography

Abadi, M. et al. (2016). "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems". In: *ArXiv e-prints*. arXiv: 1603.04467 [cs.DC].

Abazajian, K. N. et al. (2009). "The Seventh Data Release of the Sloan Digital Sky Survey". In: ApJS 182, pp. 543–558. DOI: 10.1088/0067-0049/182/2/543. arXiv: 0812.0649.

Abazajian, K. N. et al. (2016). "CMB-S4 Science Book, First Edition". In: *ArXiv e-prints*. arXiv: 1610.02743.

Acquaviva, Viviana (2019). "Pushing the Technical Frontier: From Overwhelmingly Large Data Sets to Machine Learning". In: *arXiv e-prints*. arXiv: 1901.05978 [astro-ph.IM].

Ade, P. A. R. et al. (2014). "Measurement of the Cosmic Microwave Background Polarization Lensing Power Spectrum with the POLARBEAR Experiment". In: Phys. Rev. Lett. 113.2, p. 021301. DOI: 10.1103/PhysRevLett.113.021301. arXiv: 1312.6646 [astro-ph.CO].

Ade, Peter et al. (2019). "The Simons Observatory: science goals and forecasts". In: J. Cosmology Astropart. Phys. 2, p. 056. DOI: 10.1088/1475-7516/2019/02/056. arXiv: 1808.07445 [astro-ph.CO].

Ahn, C. P. et al. (2014). "The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment". In: ApJS 211, p. 17. DOI: 10.1088/0067-0049/211/2/17. arXiv: 1307.7735 [astro-ph.IM].

Alam, S. et al. (2015). "The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III". In: ApJS 219, p. 12. DOI: 10.1088/0067-0049/219/1/12. arXiv: 1501.00963 [astro-ph.IM].

Alam, S. et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". In: MNRAS 470, pp. 2617–2652. DOI: 10.1093/mnras/stx721. arXiv: 1607.03155.

Albrecht, A. and P. J. Steinhardt (1982). "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking". In: Phys. Rev. Lett. 48, pp. 1220–1223. DOI: 10.1103/PhysRevLett.48.1220.

Alcock, C. and B. Paczynski (1979). "An evolution free test for non-zero cosmological constant". In: Nature 281, p. 358. DOI: 10.1038/281358a0.

Alimi, Jean-Michel et al. (2012). "First-Ever Full Observable Universe Simulation". In: *2012 International Conference for High Performance Computing, Networking, Storage and Analysis*. Salt Lake City, UT: IEEE, pp. 1–11. ISBN: 978-1-4673-0805-2 978-1-4673-0806-9. DOI: 10.1109/SC.2012.58.

Alonso, D. et al. (2015). "Homogeneity and isotropy in the Two Micron All Sky Survey Photometric Redshift catalogue". In: MNRAS 449.1, pp. 670–684. DOI: 10.1093/mnras/stv309. arXiv: 1412.5151 [astro-ph.CO].

Alonso, David et al. (2019). "A unified pseudo-C_ℓ framework". In: MNRAS 484.3, pp. 4127–4151. DOI: 10.1093/mnras/stz093. arXiv: 1809.09603 [astro-ph.CO].

Alsing, J. et al. (2016). "Hierarchical cosmic shear power spectrum inference". In: MNRAS 455, pp. 4452–4466. DOI: 10.1093/mnras/stv2501. arXiv: 1505.07840.

Alsing, Justin, Alan Heavens, and Andrew H. Jaffe (2017). "Cosmological parameters, shear maps and power spectra from CFHTLenS using Bayesian hierarchical inference". In: MNRAS 466.3, pp. 3272–3292. DOI: 10.1093/mnras/stw3161. arXiv: 1607.00008 [astro-ph.CO].

Amendola, L. et al. (2016). "Cosmology and Fundamental Physics with the Euclid Satellite". In: *ArXiv e-prints*. arXiv: 1606.00180.

Anderes, E., B. D. Wandelt, and G. Lavaux (2015). "Bayesian Inference of CMB Gravitational Lensing". In: ApJ 808, p. 152. DOI: 10.1088/0004-637X/808/2/152. arXiv: 1412.4079.

Appleby, Stephen and Arman Shafieloo (2014). "Testing isotropy in the local Universe". In: J. Cosmology Astropart. Phys. 10, p. 070. DOI: 10.1088/1475-7516/2014/10/070. arXiv: 1405.4595 [astro-ph.CO].

Aragon-Calvo, M. A. (2019). "Classifying the large-scale structure of the universe with deep neural networks". In: MNRAS 484.4, pp. 5771–5784. DOI: 10.1093/mnras/stz393. arXiv: 1804.00816 [astro-ph.CO].

Arjovsky, M. and L. Bottou (2017). "Towards Principled Methods for Training Generative Adversarial Networks". In: *ArXiv e-prints*. arXiv: 1701.04862 [stat.ML].

Arjovsky, M., S. Chintala, and L. Bottou (2017). "Wasserstein GAN". In: *ArXiv e-prints*. arXiv: 1701.07875 [stat.ML].

Avila, F. et al. (2019). "The angular scale of homogeneity in the Local Universe with the SDSS blue galaxies". In: *arXiv e-prints*. arXiv: 1906.10744 [astro-ph.CO].

Avila, Felipe et al. (2018). "The scale of homogeneity in the local Universe with the ALFALFA catalogue". In: J. Cosmology Astropart. Phys. 12, p. 041. DOI: 10.1088/1475-7516/2018/12/041. arXiv: 1806.04541 [astro-ph.CO].

Ballinger, W. E., J. A. Peacock, and A. F. Heavens (1996). "Measuring the cosmological constant with redshift surveys". In: MNRAS 282, p. 877. DOI: 10.1093/mnras/282.3.877. eprint: astro-ph/9605017.

Barkana, R. (2006). "Separating out the Alcock-Paczyński effect on 21-cm fluctuations". In: MNRAS 372, pp. 259–264. DOI: 10.1111/j.1365-2966.2006.10882.x. eprint: astro-ph/0508341.

Bartelmann, M. and P. Schneider (2001). "Weak gravitational lensing". In: Phys. Rep. 340, pp. 291–472. DOI: 10.1016/S0370-1573(00)00082-X. eprint: astro-ph/9912508.

Baumann, Daniel (2011). "The physics of inflation". In: *ICTS course*.

Behroozi, P. S., R. H. Wechsler, and H.-Y. Wu (2013). "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores". In: ApJ 762, p. 109. DOI: 10.1088/0004-637X/762/2/109. arXiv: 1110.4372 [astro-ph.CO].

Bengio, Yoshua et al. (2009). "Learning deep architectures for AI". In: *Foundations and trends® in Machine Learning* 2.1, pp. 1–127.

Berger, Philippe and George Stein (2019). "A volumetric deep Convolutional Neural Network for simulation of mock dark matter halo catalogues". In: MNRAS 482, pp. 2861–2871. DOI: 10.1093/mnras/sty2949. arXiv: 1805.04537 [astro-ph.CO].

Bernardeau, F. et al. (2002). "Large-scale structure of the Universe and cosmological perturbation theory". In: Phys. Rep. 367, pp. 1–248. DOI: 10.1016/S0370-1573(02)00135-7. arXiv: astro-ph/0112551 [astro-ph].

Bertschinger, E. and A. Dekel (1989). "Recovering the full velocity and density fields from large-scale redshift-distance samples". In: ApJ 336, pp. L5–L8. DOI: 10.1086/185348.

— (1991). "Mapping large-scale flows in three dimensions: method." In: *Large-scale Structures and Peculiar Motions in the Universe*. Ed. by D. W. Latham and L. A. N. da Costa. Vol. 15. Astronomical Society of the Pacific Conference Series, pp. 67–82.

BICEP2 Collaboration et al. (2014). "Detection of B-Mode Polarization at Degree Angular Scales by BICEP2". In: Phys. Rev. Lett. 112.24, p. 241101. DOI: 10.1103/PhysRevLett.112.241101. arXiv: 1403.3985 [astro-ph.CO].

BICEP2 Collaboration et al. (2018). "Constraints on Primordial Gravitational Waves Using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season". In: Phys. Rev. Lett. 121.22, p. 221301. DOI: 10.1103/PhysRevLett.121.221301. arXiv: 1810.05216 [astro-ph.CO].

BICEP2/Keck Collaboration et al. (2015). "Joint Analysis of BICEP2/Keck Array and Planck Data". In: Phys. Rev. Lett. 114.10, p. 101301. DOI: 10.1103/PhysRevLett.114.101301. arXiv: 1502.00612 [astro-ph.CO].

Bilicki, Maciej et al. (2014). "Two Micron All Sky Survey Photometric Redshift Catalog: A Comprehensive Three-dimensional Census of the Whole Sky". In: ApJS 210.1, p. 9. DOI: 10.1088/0067-0049/210/1/9. arXiv: 1311.5246 [astro-ph.CO].

Bistolas, V. and Y. Hoffman (1998). "Nonlinear Constrained Realizations of the Large-Scale Structure". In: ApJ 492, pp. 439–451. DOI: 10.1086/305080. eprint: astro-ph/9707243.

Blake, C. and K. Glazebrook (2003). "Probing Dark Energy Using Baryonic Oscillations in the Galaxy Power Spectrum as a Cosmological Ruler". In: ApJ 594, pp. 665–673. DOI: 10.1086/376983. eprint: astro-ph/0301632.

Blake, C. et al. (2011). "The WiggleZ Dark Energy Survey: measuring the cosmic expansion history using the Alcock-Paczynski test and distant supernovae". In: MNRAS 418, pp. 1725–1735. DOI: 10.1111/j.1365-2966.2011.19606.x. arXiv: 1108.2637.

Bond, J. R. and G. Efstathiou (1984). "Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter". In: ApJ 285, pp. L45–L48. DOI: 10.1086/184362.

Bond, J. R., A. H. Jaffe, and L. Knox (1998). "Estimating the power spectrum of the cosmic microwave background". In: Phys. Rev. D 57, pp. 2117–2137. DOI: 10.1103/PhysRevD.57.2117. eprint: astro-ph/9708203.

Bond, J Richard, Lev Kofman, and Dmitri Pogosyan (1996). "How filaments are woven into the cosmic web". In: *Nature* 380, pp. 603–606.

Bouchet, F. R., S. Prunet, and S. K. Sethi (1999). "Multifrequency Wiener filtering of cosmic microwave background data with polarization". In: MNRAS 302, pp. 663–676. DOI: 10.1046/j.1365-8711.1999.02118.x.

Bouchet, F. R. et al. (1995). "Perturbative Lagrangian approach to gravitational instability." In: A&A 296, p. 575. arXiv: astro-ph/9406013 [astro-ph].

Bovy, J. et al. (2012). "Photometric Redshifts and Quasar Probabilities from a Single, Data-driven Generative Model". In: ApJ 749, p. 41. DOI: 10.1088/0004-637X/749/1/41. arXiv: 1105.3975.

Bowyer, J., A. H. Jaffe, and D. I. Novikov (2011). "MasQU: Finite Differences on Masked Irregular Stokes Q,U Grids". In: *ArXiv e-prints*. arXiv: 1101.0520.

Buchert, T. (1989). "A class of solutions in Newtonian cosmology and the pancake theory". In: A&A 223, pp. 9–24.

Buchert, T., A. L. Melott, and A. G. Weiss (1994). "Testing higher-order Lagrangian perturbation theory against numerical simulations I. Pancake models". In: A&A 288, pp. 349–364.

Bunn, E. F. (2002a). "Detectability of microwave background polarization". In: Phys. Rev. D 65.4, p. 043003. DOI: 10.1103/PhysRevD.65.043003. eprint: astro-ph/0108209.

— (2002b). "Erratum: Detectability of microwave background polarization [Phys. Rev. D 65, 043003 (2002)]". In: Phys. Rev. D 66.6, p. 069902. DOI: 10.1103/PhysRevD.66.069902.

— (2011). "Efficient decomposition of cosmic microwave background polarization maps into pure E, pure B, and ambiguous components". In: Phys. Rev. D 83.8, p. 083003. DOI: 10.1103/PhysRevD.83.083003. arXiv: 1008.0827 [astro-ph.CO].

Bunn, E. F. and B. Wandelt (2017). "Pure E and B polarization maps via Wiener filtering". In: Phys. Rev. D 96.4, p. 043523. DOI: 10.1103/PhysRevD.96.043523. arXiv: 1610.03345.

Bunn, E. F. et al. (1994). "Wiener filtering of the COBE Differential Microwave Radiometer data". In: ApJ 432, pp. L75–L78. DOI: 10.1086/187515. eprint: astro-ph/9404007.

Bunn, E. F. et al. (2003). "E/B decomposition of finite pixelized CMB maps". In: Phys. Rev. D 67.2, p. 023501. DOI: 10.1103/PhysRevD.67.023501. eprint: astro-ph/0207338.

Caldeira, J. et al. (2018). "DeepCMB: Lensing Reconstruction of the Cosmic Microwave Background with Deep Neural Networks". In: *arXiv e-prints*. arXiv: 1810.01483 [astro-ph.CO].

Cao, L. and L.-Z. Fang (2009). "A Wavelet-Galerkin Algorithm of the E/B Decomposition of Cosmic Microwave Background Polarization Maps". In: ApJ 706, pp. 1545–1555. DOI: 10.1088/0004-637X/706/2/1545. arXiv: 0910.4697.

Carleo, Giuseppe et al. (2019). "Machine learning and the physical sciences". In: *arXiv e-prints*. arXiv: 1903.10563 [physics.comp-ph].

Carroll, Sean M (2004). *Spacetime and geometry. An introduction to general relativity*. Vol. 1. Addison-Wesley.

Casella, George and Edward I George (1992). "Explaining the Gibbs sampler". In: *The American Statistician* 46.3, pp. 167–174.

Challinor, A. and G. Chon (2005). "Error analysis of quadratic power spectrum estimates for cosmic microwave background polarization: sampling covariance". In: MNRAS 360, pp. 509–532. DOI: 10.1111/j.1365-2966.2005.09076.x. eprint: astro-ph/0410097.

Challinor, Anthony and Hiranya Peiris (2009). "Lecture notes on the physics of cosmic microwave background anisotropies". In: *American Institute of Physics Conference Series*. Ed. by Mario Novello and Santiago Perez. Vol. 1132. American Institute of Physics Conference Series, pp. 86–140. DOI: 10.1063/1.3151849. arXiv: 0903.5158 [astro-ph.CO].

Chevallier, M. and D. Polarski (2001). "Accelerating Universes with Scaling Dark Matter". In: *International Journal of Modern Physics D* 10, pp. 213–223. DOI: 10.1142/S0218271801000822. eprint: gr-qc/0009008.

Chluba, Jens (2014). "Science with CMB spectral distortions". In: *arXiv e-prints*. arXiv: 1405.6938 [astro-ph.CO].

Cohen, Taco S. et al. (2018). "Spherical CNNs". In: *arXiv e-prints*. arXiv: 1801.10130 [cs.LG].

Coles, P. and B. Jones (1991). "A lognormal model for the cosmological mass distribution". In: MNRAS 248, pp. 1–13. DOI: 10.1093/mnras/248.1.1.

Cox, Richard T (1946). "Probability, frequency and reasonable expectation". In: *American journal of physics* 14.1, pp. 1–13.

Davis, Chandler (1962). "The norm of the Schur product operation". In: *Numerische Mathematik* 4.1, pp. 343–344.

de Laix, A. A. and G. Starkman (1998). "Sensitivity of Redshift Distortion Measurements to Cosmological Parameters". In: ApJ 501, pp. 427–441. DOI: 10.1086/305828. eprint: astro-ph/9707008.

Debono, Ivan and George F. Smoot (2016). "General Relativity and Cosmology: Unsolved Questions and Future Directions". In: *Universe* 2.4, p. 23. DOI: 10.3390/universe2040023. arXiv: 1609.09781 [gr-qc].

Desjacques, V., D. Jeong, and F. Schmidt (2016). "Large-Scale Galaxy Bias". In: *ArXiv e-prints*. arXiv: 1611.09787.

Dodelson, Scott (2003). *Modern cosmology*. Academic press.

Draper, Norman R and Harry Smith (2014). *Applied regression analysis*. Vol. 326. John Wiley & Sons.

Duane, S. et al. (1987). "Hybrid Monte Carlo". In: *Physics Letters B* 195, pp. 216–222. DOI: 10.1016/0370-2693(87)91197-X.

Dunkley, J. et al. (2009). "Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters from the WMAP Data". In: ApJS 180, pp. 306–329. DOI: 10.1088/0067-0049/180/2/306. arXiv: 0803.0586.

Efstathiou, G. et al. (1985). "Numerical techniques for large cosmological N-body simulations". In: *The Astrophysical Journal Supplement Series* 57, pp. 241–260. DOI: 10.1086/191003.

Einstein, A. (1915). "Die Feldgleichungen der Gravitation". In: *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847.*

Eisenstein, D. J. and W. Hu (1998). "Baryonic Features in the Matter Transfer Function". In: ApJ 496, pp. 605–614. DOI: 10.1086/305424. eprint: astro-ph/9709112.

— (1999). "Power Spectra for Cold Dark Matter and Its Variants". In: ApJ 511, pp. 5–15. DOI: 10.1086/306640. eprint: astro-ph/9710252.

Eisenstein, Daniel J. et al. (2005). "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". In: ApJ 633.2, pp. 560–574. DOI: 10.1086/466512. arXiv: astro-ph/0501171 [astro-ph].

Elsner, F., B. Leistedt, and H. V. Peiris (2017). "Unbiased pseudo-C_ℓ power spectrum estimation with mode projection". In: MNRAS 465, pp. 1847–1855. DOI: 10.1093/mnras/stw2752. arXiv: 1609.03577.

Elsner, F. and B. D. Wandelt (2009). "Improved Simulation of Non-Gaussian Temperature and Polarization Cosmic Microwave Background Maps". In: ApJS 184, pp. 264–270. DOI: 10.1088/0067-0049/184/2/264. arXiv: 0909.0009.

— (2010). "Local Non-Gaussianity in the Cosmic Microwave Background the Bayesian Way". In: ApJ 724, pp. 1262–1269. DOI: 10.1088/0004-637X/724/2/1262. arXiv: 1010.1254 [astro-ph.CO].

— (2012a). "Fast calculation of the Fisher matrix for cosmic microwave background experiments". In: A&A 540, p. L6. DOI: 10.1051/0004-6361/201218985. arXiv: 1202.4898 [astro-ph.CO].

Elsner, F. and B. D. Wandelt (2012b). "Fast Wiener filtering of CMB maps". In: *Proc. Big Bang, Big Data, Big Computers*. arXiv: 1211.0585.

— (2012c). "Likelihood, Fisher information, and systematics of cosmic microwave background experiments". In: A&A 542, A60. DOI: 10.1051/0004-6361/201219293. arXiv: 1205.0810 [astro-ph.CO].

— (2013). "Efficient Wiener filtering without preconditioning". In: A&A 549, A111. DOI: 10.1051/0004-6361/201220586. arXiv: 1210.4931 [astro-ph.CO].

Erdmann, M. et al. (2018). "Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks". In: *ArXiv e-prints*. arXiv: 1802.03325 [astro-ph.IM].

Erdoğdu, P. et al. (2004). "The 2dF Galaxy Redshift Survey: Wiener reconstruction of the cosmic web". In: MNRAS 352, pp. 939–960. DOI: 10.1111/j.1365-2966.2004.07984.x. eprint: astro-ph/0312546.

Erdoğdu, P. et al. (2006). "Reconstructed density and velocity fields from the 2MASS Redshift Survey". In: MNRAS 373, pp. 45–64. DOI: 10.1111/j.1365-2966.2006.11049.x. eprint: astro-ph/0610005.

Eriksen, H. K. et al. (2004). "Power Spectrum Estimation from High-Resolution Maps by Gibbs Sampling". In: ApJS 155, pp. 227–241. DOI: 10.1086/425219. eprint: astro-ph/0407028.

Eriksen, H. K. et al. (2008). "The Joint Large-Scale Foreground-CMB Posteriors of the 3 Year WMAP Data". In: ApJ 672, p. L87. DOI: 10.1086/526545. arXiv: 0709.1037.

Ferté, A. et al. (2013). "Efficiency of pseudospectrum methods for estimation of the cosmic microwave background B-mode power spectrum". In: Phys. Rev. D 88.2, p. 023524. DOI: 10.1103/PhysRevD.88.023524. arXiv: 1305.7441 [astro-ph.CO].

Fisher, K. B. et al. (1995). "Wiener reconstruction of density, velocity and potential fields from all-sky galaxy redshift surveys". In: MNRAS 272, pp. 885–908. DOI: 10.1093/mnras/272.4.885. eprint: astro-ph/9406009.

Flauger, Raphael, J. Colin Hill, and David N. Spergel (2014). "Toward an understanding of foreground emission in the BICEP2 region". In: J. Cosmology Astropart. Phys. 8, p. 039. DOI: 10.1088/1475-7516/2014/08/039. arXiv: 1405.7351 [astro-ph.CO].

Fornengo, N. (2016). "Dark matter overview". In: *arXiv e-prints*. arXiv: 1701.00119 [astro-ph.HE].

Fowler, David and Eleanor Robson (1998). "Square root approximations in Old Babylonian mathematics: YBC 7289 in context". In: *Historia Mathematica* 25.4, pp. 366–378.

Freese, Katherine (2017). "Status of dark matter in the universe". In: *International Journal of Modern Physics D* 26.6, pp. 1730012–223. DOI: 10.1142/S0218271817300129. arXiv: 1701.01840 [astro-ph.CO].

Friberg, Jöran (2007). *A Remarkable Collection of Babylonian Mathematical Texts: Manuscripts in the Schøyen Collection: Cuneiform Texts I*. Springer Science & Business Media.

Friedmann, A. (1922). "Über die Krümmung des Raumes". In: *Zeitschrift fur Physik* 10, pp. 377–386. DOI: 10.1007/BF01332580.

— (1924). "Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes". In: *Zeitschrift fur Physik* 21, pp. 326–332. DOI: 10.1007/BF01328280.

Frigo, Matteo and Steven G Johnson (2005). "The design and implementation of FFTW3". In: *Proc. IEEE* 93.2, pp. 216–231.

Fu, H. et al. (2018). "Perception Oriented Haze Image Definition Restoration by Basing on Physical Optics Model". In: *IEEE Photonics Journal* 10, p. 2837010. DOI: 10.1109/JPHOT.2018.2837010.

Fussell, L. and B. Moews (2018). "Forging new worlds: high-resolution synthetic galaxies with chained generative adversarial networks". In: *ArXiv e-prints*. arXiv: 1811.03081 [astro-ph.IM].

Galassi, M et al. (2009). *GNU Scientific Library reference manual Network Theory*. Bristol, UK.

Galli, S. et al. (2014). "CMB polarization can constrain cosmology better than CMB temperature". In: Phys. Rev. D 90.6, p. 063504. DOI: 10.1103/PhysRevD.90.063504. arXiv: 1403.5271.

Gauss, Carl Friedrich (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss*. sumtibus Frid. Perthes et IH Besser.

Gaztanaga, E. and J. Yokoyama (1993). "Probing the statistics of primordial fluctuations and their evolution". In: ApJ 403, pp. 450–465. DOI: 10.1086/172216.

Gelfand, Alan E and Adrian FM Smith (1990). "Sampling-based approaches to calculating marginal densities". In: *Journal of the American statistical association* 85.410, pp. 398–409.

Gelman, Andrew et al. (2008). "Objections to Bayesian statistics". In: *Bayesian Analysis* 3.3, pp. 445–449.

Geman, Stuart and Donald Geman (1987). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". In: *Readings in computer vision*. Elsevier, pp. 564–584.

Glazebrook, K. and C. Blake (2005). "Measuring the Cosmic Evolution of Dark Energy with Baryonic Oscillations in the Galaxy Power Spectrum". In: ApJ 631, pp. 1–20. DOI: 10.1086/432497. eprint: astro-ph/0505608.

Gleser, L., A. Nusser, and A. J. Benson (2008). "Decontamination of cosmological 21-cm maps". In: MNRAS 391, pp. 383–398. DOI: 10.1111/j.1365-2966.2008.13897.x. arXiv: 0712.0497.

Golub, Gene H and Charles F Van Loan (1996). "Matrix computations. 1996". In: *Johns Hopkins University, Press, Baltimore, MD, USA*, pp. 374–426.

Goodfellow, I. J. et al. (2014). "Generative Adversarial Networks". In: *ArXiv e-prints*. arXiv: 1406.2661 [stat.ML].

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep learning*. MIT press.

Górski, K. M. et al. (2005). "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere". In: ApJ 622, pp. 759–771. DOI: 10.1086/427976. eprint: astro-ph/0409513.

Grain, J., M. Tristram, and R. Stompor (2009). "Polarized CMB power spectrum estimation using the pure pseudo-cross-spectrum approach". In: Phys. Rev. D 79.12, p. 123515. DOI: 10.1103/PhysRevD.79.123515. arXiv: 0903.2350 [astro-ph.CO].

Gulrajani, I. et al. (2017). "Improved Training of Wasserstein GANs". In: *ArXiv e-prints*. arXiv: 1704.00028 [cs.LG].

Gupta, A. et al. (2018). "Non-Gaussian information from weak lensing data via deep learning". In: *ArXiv e-prints*. arXiv: 1802.01212.

Guth, A. H. and S. Y. Pi (1982). "Fluctuations in the New Inflationary Universe". In: Phys. Rev. Lett. 49, pp. 1110–1113. DOI: 10.1103/PhysRevLett.49.1110.

Guth, Alan H. (1981). "Inflationary universe: A possible solution to the horizon and flatness problems". In: Phys. Rev. D 23, pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

Guzzetti, M. C. et al. (2016). "Gravitational waves from inflation". In: *ArXiv e-prints*. arXiv: 1605.01615.

Hahn, O. and T. Abel (2011). "Multi-scale initial conditions for cosmological simulations". In: MNRAS 415, pp. 2101–2121. DOI: 10.1111/j.1365-2966.2011.18820.x. arXiv: 1103.6031.

Hamaus, N. et al. (2014). "Testing cosmic geometry without dynamic distortions using voids". In: *J. Cosmol. Astropart. Phys.* 12, p. 013. DOI: 10.1088/1475-7516/2014/12/013. arXiv: 1409.3580.

— (2015). "Probing cosmology and gravity with redshift-space distortions around voids". In: *J. Cosmol. Astropart. Phys.* 11, p. 036. DOI: 10.1088/1475-7516/2015/11/036. arXiv: 1507.04363.

Hamaus, N. et al. (2016). "Constraints on Cosmology and Gravity from the Dynamics of Voids". In: *Physical Review Letters* 117.9, p. 091302. DOI: 10.1103/PhysRevLett.117.091302. arXiv: 1602.01784.

Hanson, D. et al. (2013). "Detection of B-Mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope". In: *Physical Review Letters* 111.14, p. 141301. DOI: 10.1103/PhysRevLett.111.141301. arXiv: 1307.5830 [astro-ph.CO].

Harrison, Edward R (1970). "Fluctuations at the threshold of classical cosmology". In: Phys. Rev. D 1.10, p. 2726.

Hastings, W Keith (1970). "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1, pp. 97–109.

He, Kaiming et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

He, Siyu et al. (2019). "Learning to predict the cosmological structure formation". In: *Proceedings of the National Academy of Science* 116.28, pp. 13825–13832. DOI: 10.1073/pnas.1821458116. arXiv: 1811.06533 [astro-ph.CO].

Heavens, Alan (2009). "Statistical techniques in cosmology". In: *arXiv e-prints*. arXiv: 0906.0664 [astro-ph.CO].

Henning, J. W. et al. (2018). "Measurements of the Temperature and E-mode Polarization of the CMB from 500 Square Degrees of SPTpol Data". In: ApJ 852.2, p. 97. DOI: 10.3847/1538-4357/aa9ff4. arXiv: 1707.09353 [astro-ph.CO].

Hinshaw, G. et al. (2007). "Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Temperature Analysis". In: ApJS 170, pp. 288–334. DOI: 10.1086/513698. eprint: astro-ph/0603451.

Hirata, C. M. and U. Seljak (2003). "Reconstruction of lensing from the cosmic microwave background polarization". In: Phys. Rev. D 68.8, p. 083002. DOI: 10.1103/PhysRevD.68.083002. eprint: astro-ph/0306354.

Hirata, C. M. et al. (2004). "Cross-correlation of CMB with large-scale structure: Weak gravitational lensing". In: Phys. Rev. D 70.10, p. 103501. DOI: 10.1103/PhysRevD.70.103501. eprint: astro-ph/0406004.

Ho, S. et al. (2015). "Sloan Digital Sky Survey III photometric quasar clustering: probing the initial conditions of the Universe". In: J. Cosmology Astropart. Phys. 5, p. 040. DOI: 10.1088/1475-7516/2015/05/040. arXiv: 1311.2597.

Ho, Shirley et al. (2012). "Clustering of Sloan Digital Sky Survey III Photometric Luminous Galaxies: The Measurement, Systematics, and Cosmological Implications". In: ApJ 761, p. 14. DOI: 10.1088/0004-637X/761/1/14. arXiv: 1201.2137 [astro-ph.CO].

Hockney, R. W. and J. W. Eastwood (1988). *Computer simulation using particles*.

Hoffman, Y. (1994). "Wiener Reconstruction of the Large-Scale Structure in the Zone of Avoidance". In: *Unveiling Large-Scale Structures Behind the Milky Way*. Ed. by C. Balkowski and R. C. Kraan-Korteweg. Vol. 67. Astronomical Society of the Pacific Conference Series, p. 185.

Hoffman, Y. and E. Ribak (1991). "Constrained realizations of Gaussian fields - A simple algorithm". In: ApJ 380, pp. L5–L8. DOI: 10.1086/186160.

Hogg, David W. et al. (2005). "Cosmic Homogeneity Demonstrated with Luminous Red Galaxies". In: ApJ 624.1, pp. 54–58. DOI: 10.1086/429084. arXiv: astro-ph/0411197 [astro-ph].

Horowitz, Benjamin, Uros Seljak, and Grigor Aslanyan (2018). "Efficient Optimal Reconstruction of Linear Fields and Band-powers from Cosmological Data". In: *arXiv e-prints*. arXiv: 1810.00503 [astro-ph.CO].

Hu, W. (2003). "CMB temperature and polarization anisotropy fundamentals". In: *Annals of Physics* 303, pp. 203–225. DOI: 10.1016/S0003-4916(02)00022-2. eprint: astro-ph/0210696.

Hu, W. and S. Dodelson (2002). "Cosmic Microwave Background Anisotropies". In: ARA&A 40, pp. 171–216. DOI: 10.1146/annurev.astro.40.060401.093926. eprint: astro-ph/0110414.

Hu, W. and M. White (1997). "A CMB polarization primer". In: New Astron. 2, pp. 323–344. DOI: 10.1016/S1384-1076(97)00022-5. eprint: astro-ph/9706147.

Hubble, E. (1934). "The Distribution of Extra-Galactic Nebulae". In: ApJ 79, p. 8. DOI: 10.1086/143517.

Huffenberger, K. M. and S. K. Næss (2018). "Cosmic Microwave Background Mapmaking with a Messenger Field". In: ApJ 852, p. 92. DOI: 10.3847/1538-4357/aa9c7d. arXiv: 1705.01893 [astro-ph.IM].

Huffenberger, Kevin M. (2018). "Preconditioner-free Wiener filtering with a dense noise matrix". In: MNRAS 476.3, pp. 3425–3431. DOI: 10.1093/mnras/sty232. arXiv: 1704.00865 [astro-ph.CO].

Huffenberger, Kevin M. and Benjamin D. Wandelt (2010). "Fast and Exact Spin-s Spherical Harmonic Transforms". In: *The Astrophysical Journal Supplement Series* 189.2, pp. 255–260. DOI: 10.1088/0067-0049/189/2/255. arXiv: 1007.3514 [astro-ph.IM].

Hui, L., A. Stebbins, and S. Burles (1999). "A Geometrical Test of the Cosmological Energy Contents Using the Lyα Forest". In: ApJ 511, pp. L5–L8. DOI: 10.1086/311826. eprint: astro-ph/9807190.

Huterer, Dragan, Carlos E. Cunha, and Wenjuan Fang (2013). "Calibration errors unleashed: effects on cosmological parameters and requirements for large-scale structure surveys". In: MNRAS 432, pp. 2945–2961. DOI: 10.1093/mnras/stt653. arXiv: 1211.1015 [astro-ph.CO].

Ioffe, S. and C. Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *ArXiv e-prints*. arXiv: 1502.03167 [cs.LG].

Isola, Phillip et al. (2016). "Image-to-Image Translation with Conditional Adversarial Networks". In: *arXiv e-prints*. arXiv: 1611.07004 [cs.CV].

Ivezic, Z. et al. (2008). "LSST: from Science Drivers to Reference Design and Anticipated Data Products". In: *ArXiv e-prints*. arXiv: 0805.2366.

Jacobi, Carl GJ (1845). "Ueber eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden lineären Gleichungen". In: *Astronomische Nachrichten* 22.20, pp. 297–306.

Jasche, J. and F. S. Kitaura (2010). "Fast Hamiltonian sampling for large-scale structure inference". In: MNRAS 407, pp. 29–42. DOI: 10.1111/j.1365-2966.2010.16897.x. arXiv: 0911.2496.

Jasche, J. and G. Lavaux (2015). "Matrix-free large-scale Bayesian inference in cosmology". In: MNRAS 447, pp. 1204–1212. DOI: 10.1093/mnras/stu2479. arXiv: 1402.1763.

— (2017). "Bayesian power spectrum inference with foreground and target contamination treatment". In: A&A 606, A37. DOI: 10.1051/0004-6361/201730909. arXiv: 1706.08971.

— (2019). "Physical Bayesian modelling of the non-linear matter distribution: New insights into the nearby universe". In: A&A 625, A64. DOI: 10.1051/0004-6361/201833710. arXiv: 1806.11117 [astro-ph.CO].

Jasche, J., F. Leclercq, and B. D. Wandelt (2015). "Past and present cosmic structure in the SDSS DR7 main sample". In: J. Cosmology Astropart. Phys. 1, p. 036. DOI: 10.1088/1475-7516/2015/01/036. arXiv: 1409.6308.

Jasche, J. and B. D. Wandelt (2012). "Bayesian inference from photometric redshift surveys". In: MNRAS 425, pp. 1042–1056. DOI: 10.1111/j.1365-2966.2012.21423.x. arXiv: 1106.2757.

— (2013a). "Bayesian physical reconstruction of initial conditions from large-scale structure surveys". In: MNRAS 432, pp. 894–913. DOI: 10.1093/mnras/stt449. arXiv: 1203.3639.

— (2013b). "Methods for Bayesian Power Spectrum Inference with Galaxy Surveys". In: ApJ 779, p. 15. DOI: 10.1088/0004-637X/779/1/15. arXiv: 1306.1821.

Jasche, J. et al. (2010a). "Bayesian non-linear large-scale structure inference of the Sloan Digital Sky Survey Data Release 7". In: MNRAS 409, pp. 355–370. DOI: 10.1111/j.1365-2966.2010.17313.x. arXiv: 0911.2498.

Jasche, J. et al. (2010b). "Bayesian power-spectrum inference for large-scale structure data". In: MNRAS 406, pp. 60–85. DOI: 10.1111/j.1365-2966.2010.16610.x. arXiv: 0911.2493.

Jaynes, Edwin T (1957). "Information theory and statistical mechanics". In: *Physical review* 106.4, p. 620.

— (2003). *Probability theory: The logic of science*. Cambridge university press.

Jeffreys, Harold (1946). "An invariant form for the prior probability in estimation problems". In: *Proc. R. Soc. Lond. A* 186.1007, pp. 453–461.

Jewell, J., S. Levin, and C. H. Anderson (2004). "Application of Monte Carlo Algorithms to the Bayesian Analysis of the Cosmic Microwave Background". In: ApJ 609, pp. 1–14. DOI: 10.1086/383515. eprint: astro-ph/0209560.

Jones, D. H. et al. (2009). "The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures". In: MNRAS 399, pp. 683–698. DOI: 10.1111/j.1365-2966.2009.15338.x. arXiv: 0903.5451.

Kaiser, N. (1984). "On the spatial correlations of Abell clusters". In: ApJ 284, pp. L9–L12. DOI: 10.1086/184341.

Kalman, Rudolph Emil (1960). "A new approach to linear filtering and prediction problems". In: *Journal of basic Engineering* 82.1, pp. 35–45.

Kamionkowski, M., A. Kosowsky, and A. Stebbins (1997a). "A Probe of Primordial Gravity Waves and Vorticity". In: *Physical Review Letters* 78, pp. 2058–2061. DOI: 10.1103/PhysRevLett.78.2058. eprint: astro-ph/9609132.

— (1997b). "Statistics of cosmic microwave background polarization". In: Phys. Rev. D 55, pp. 7368–7388. DOI: 10.1103/PhysRevD.55.7368. eprint: astro-ph/9611125.

Kamionkowski, M. and E. D. Kovetz (2016). "The Quest for B Modes from Inflationary Gravitational Waves". In: ARA&A 54, pp. 227–269. DOI: 10.1146/annurev-astro-081915-023433. arXiv: 1510.06042.

Karras, T. et al. (2017). "Progressive Growing of GANs for Improved Quality, Stability, and Variation". In: *ArXiv e-prints*. arXiv: 1710.10196.

Karras, Tero, Samuli Laine, and Timo Aila (2018). "A Style-Based Generator Architecture for Generative Adversarial Networks". In: *arXiv e-prints*. arXiv: 1812.04948 [cs.NE].

Kayo, I., A. Taruya, and Y. Suto (2001). "Probability Distribution Function of Cosmological Density Fluctuations from a Gaussian Initial Condition: Comparison of One-Point and Two-Point Lognormal Model Predictions with N-Body Simulations". In: ApJ 561, pp. 22–34. DOI: 10.1086/323227. eprint: astro-ph/0105218.

Kendall, Maurice G and Alan Stuart (1968). "The Advanced Theory of Statistics, Vol. 3, Charles Griffin & Co". In: *London, 1946Kendall1The Advanced Theory of Statistics1946*.

Kim, J. (2011). "How to make a clean separation between CMB E and B modes with proper foreground masking". In: A&A 531, A32. DOI: 10.1051/0004-6361/201116733. arXiv: 1010.2636 [astro-ph.CO].

Kim, J. and P. Naselsky (2010). "E/B decomposition of CMB polarization pattern of incomplete sky: a pixel space approach". In: A&A 519, A104. DOI: 10.1051/0004-6361/201014739. arXiv: 1003.2911 [astro-ph.CO].

Kingma, D. P. and J. Ba (2014). "Adam: A Method for Stochastic Optimization". In: *ArXiv e-prints*. arXiv: 1412.6980.

Kitaura, F.-S. (2012). "Non-Gaussian gravitational clustering field statistics". In: MNRAS 420, pp. 2737–2755. DOI: 10.1111/j.1365-2966.2011.19680.x. arXiv: 1012.3168.

Kitaura, F. S. and T. A. Enßlin (2008). "Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization". In: MNRAS 389, pp. 497–544. DOI: 10.1111/j.1365-2966.2008.13341.x. arXiv: 0705.0429.

Kitaura, F.-S., S. Gallerani, and A. Ferrara (2012). "Multiscale inference of matter fields and baryon acoustic oscillations from the Lyα forest". In: MNRAS 420, pp. 61–74. DOI: 10.1111/j.1365-2966.2011.19997.x. arXiv: 1011.6233.

Kitaura, F.-S., J. Jasche, and R. B. Metcalf (2010). "Recovering the non-linear density field from the galaxy distribution with a Poisson-lognormal filter". In: MNRAS 403, pp. 589–604. DOI: 10.1111/j.1365-2966.2009.16163.x. arXiv: 0911.1407.

Kitaura, F. S. et al. (2009). "Cosmic cartography of the large-scale structure with Sloan Digital Sky Survey data release 6". In: MNRAS 400, pp. 183–203. DOI: 10.1111/j.1365-2966.2009.15470.x. arXiv: 0906.3978 [astro-ph.CO].

Klypin, A. A. and S. F. Shandarin (1983). "Three-dimensional numerical model of the formation of large-scale structure in the Universe." In: MNRAS 204, pp. 891–907. DOI: 10.1093/mnras/204.3.891.

Klypin, Anatoly and Jon Holtzman (1997). "Particle-Mesh code for cosmological simulations". In: *arXiv e-prints*, astro–ph/9712217. arXiv: astro-ph/9712217 [astro-ph].

Kodi Ramanah, D., G. Lavaux, and B. D. Wandelt (2017). "Wiener filter reloaded: fast signal reconstruction without preconditioning". In: MNRAS 468, pp. 1782–1793. DOI: 10.1093/mnras/stx527. arXiv: 1702.08852.

— (2018). "Optimal and fast E/B separation with a dual messenger field". In: MNRAS 476, pp. 2825–2834. DOI: 10.1093/mnras/sty341. arXiv: 1801.05358.

Kodi Ramanah, Doogesh, Tom Charnock, and Guilhem Lavaux (2019). "Painting halos from cosmic density fields of dark matter with physically motivated neural networks". In: Phys. Rev. D 100 (4), p. 043515. DOI: 10.1103/PhysRevD.100.043515.

Kodi Ramanah, Doogesh, Guilhem Lavaux, and Benjamin D. Wandelt (2019). "Wiener filtering and pure E/B decomposition of CMB maps with anisotropic correlated noise". In: *arXiv e-prints*. arXiv: 1906.10704 [astro-ph.CO].

Kodi Ramanah, Doogesh et al. (2019). "Cosmological inference from Bayesian forward modelling of deep galaxy redshift surveys". In: A&A 621, A69. DOI: 10.1051/0004-6361/201834117. arXiv: 1808.07496 [astro-ph.CO].

Kohonen, Teuvo (1990). "The self-organizing map". In: *Proceedings of the IEEE* 78.9, pp. 1464–1480.

Kohonen, Teuvo, Gyorgy Barna, and Ronald Chrisley (1988). "Statistical pattern recognition with neural networks: Benchmarking studies". In: *IEEE International Conference on Neural Networks*. Vol. 1, pp. 61–68.

Kolb, Edward W and Michael Stanley Turner (1990). "The early universe." In: *Front. Phys., Vol. 69*, 1.

Komatsu, E., D. N. Spergel, and B. D. Wandelt (2005). "Measuring Primordial Non-Gaussianity in the Cosmic Microwave Background". In: ApJ 634, pp. 14–19. DOI: 10.1086/491724. eprint: astro-ph/0305189.

Komatsu, E. et al. (2011). "Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation". In: ApJS 192, p. 18. DOI: 10.1088/0067-0049/192/2/18. arXiv: 1001.4538 [astro-ph.CO].

Kovac, J. M. et al. (2002). "Detection of polarization in the cosmic microwave background using DASI". In: Nature 420.6917, pp. 772–787. DOI: 10.1038/nature01269. arXiv: astro-ph/0209478 [astro-ph].

Krachmalnicoff, Nicoletta and Maurizio Tomasi (2019). "Convolutional Neural Networks on the HEALPix sphere: a pixel-based algorithm and its application to CMB data analysis". In: *arXiv e-prints*. arXiv: 1902.04083 [astro-ph.IM].

Lahav, O. et al. (1994). "Wiener Reconstruction of All-Sky Galaxy Surveys in Spherical Harmonics". In: ApJ 423, p. L93. DOI: 10.1086/187244. eprint: astro-ph/9311059.

Lam, Siu Kwan, Antoine Pitrou, and Stanley Seibert (2015). "Numba: A LLVM-based Python JIT Compiler". In: *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM '15. Austin, Texas: ACM, 7:1–7:6. ISBN: 978-1-4503-4005-2. DOI: 10.1145/2833157.2833162.

Larson, D. L. et al. (2007). "Estimation of Polarized Power Spectra by Gibbs Sampling". In: ApJ 656, pp. 653–660. DOI: 10.1086/509802. eprint: astro-ph/0608007.

Lattner, Chris and Vikram Adve (2004). "LLVM: A compilation framework for lifelong program analysis & transformation". In: *Proceedings of the international symposium on Code generation and optimization: feedback-directed and runtime optimization*. IEEE Computer Society, p. 75.

Laureijs, R. et al. (2011). "Euclid Definition Study Report". In: *ArXiv e-prints*. arXiv: 1110.3193 [astro-ph.CO].

Lavaux, G. (2016). "Bayesian 3D velocity field reconstruction with VIRBIUS". In: MNRAS 457, pp. 172–197. DOI: 10.1093/mnras/stv2915. arXiv: 1512.04534.

Lavaux, G. and J. Jasche (2016). "Unmasking the masked Universe: the 2M++ catalogue through Bayesian eyes". In: MNRAS 455, pp. 3169–3179. DOI: 10.1093/mnras/stv2499. arXiv: 1509.05040.

Lavaux, G. and B. D. Wandelt (2012). "Precision Cosmography with Stacked Voids". In: ApJ 754, p. 109. DOI: 10.1088/0004-637X/754/2/109. arXiv: 1110.0345 [astro-ph.CO].

Layzer, D. (1956a). "A new model for the distribution of galaxies in space". In: AJ 61, p. 383. DOI: 10.1086/107366.

— (1956b). "A new model for the distribution of galaxies in space". In: AJ 61, p. 383. DOI: 10.1086/107366.

Leclercq, Florent (2015). "Bayesian large-scale structure inference and cosmic web analysis". In: *arXiv e-prints*. arXiv: 1512.04985 [astro-ph.CO].

Leclercq, Florent, Alice Pisani, and Benjamin D. Wandelt (2014). "Cosmology: from theory to data, from data to theory". In: *arXiv e-prints*. arXiv: 1403.1260 [astro-ph.CO].

Lecun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning". In: Nature 521, pp. 436–444. DOI: 10.1038/nature14539.

LeCun, Yann, Yoshua Bengio, et al. (1995). "Convolutional networks for images, speech, and time series". In: *The handbook of brain theory and neural networks* 3361.10, p. 1995.

LeCun, Yann et al. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

LeCun, Yann A et al. (2012). "Efficient backprop". In: *Neural networks: Tricks of the trade*. Springer, pp. 9–48.

Leistedt, B. and H. V. Peiris (2014). "Exploiting the full potential of photometric quasar surveys: optimal power spectra through blind mitigation of systematics". In: MNRAS 444, pp. 2–14. DOI: 10.1093/mnras/stu1439. arXiv: 1404.6530.

Leistedt, B. et al. (2017). "Wavelet reconstruction of E and B modes for CMB polarization and cosmic shear analyses". In: MNRAS 466, pp. 3728–3740. DOI: 10.1093/mnras/stw3176. arXiv: 1605.01414.

Lekien, F. and J. Marsden (2005). "Tricubic Interpolation in Three Dimensions". In: *Journal of Numerical Methods and Engineering* 63, pp. 455–471.

Lemaître, G. (1927). "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques". In: *Annales de la Société Scientifique de Bruxelles* 47, pp. 49–59.

Lemaître, G. (1931). "Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae". In: MNRAS 91, pp. 483–490. DOI: 10.1093/mnras/91.5.483.

— (1933). "L'Univers en expansion". In: *Annales de la Société Scientifique de Bruxelles* 53.

Levi, Michael et al. (2013). "The DESI Experiment, a whitepaper for Snowmass 2013". In: *arXiv e-prints*. arXiv: 1308.0847 [astro-ph.CO].

Lewis, A. (2003). "Harmonic E/B decomposition for CMB polarization maps". In: Phys. Rev. D 68.8, p. 083509. DOI: 10.1103/PhysRevD.68.083509. eprint: astro-ph/0305545.

Lewis, A., A. Challinor, and N. Turok (2002). "Analysis of CMB polarization on an incomplete sky". In: Phys. Rev. D 65.2, p. 023505. DOI: 10.1103/PhysRevD.65.023505. eprint: astro-ph/0106536.

Lewis, Antony, Anthony Challinor, and Anthony Lasenby (2000). "Efficient computation of CMB anisotropies in closed FRW models". In: *Astrophys. J.* 538, pp. 473–476. DOI: 10.1086/309179. arXiv: astro-ph/9911177 [astro-ph].

Linde, Andrei (2008). *Inflationary cosmology*. Springer, Berlin, Heidelberg, pp. 1–54.

Linde, Andrei D (1982). "A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems". In: *Physics Letters B* 108.6, pp. 389–393.

Linder, E. V. (2003). "Exploring the Expansion History of the Universe". In: *Physical Review Letters* 90.9, p. 091301. DOI: 10.1103/PhysRevLett.90.091301. eprint: astro-ph/0208512.

Linder, E. V. and A. Jenkins (2003). "Cosmic structure growth and dark energy". In: MNRAS 346, pp. 573–583. DOI: 10.1046/j.1365-2966.2003.07112.x. arXiv: astro-ph/0305286 [astro-ph].

López-Corredoira, M. (2014). "Alcock-Paczyński Cosmological Test". In: ApJ 781, p. 96. DOI: 10.1088/0004-637X/781/2/96. arXiv: 1312.0003.

Louis, T. et al. (2017). "The Atacama Cosmology Telescope: two-season ACTPol spectra and parameters". In: J. Cosmology Astropart. Phys. 6, p. 031. DOI: 10.1088/1475-7516/2017/06/031. arXiv: 1610.02360.

Lucie-Smith, Luisa et al. (2018). "Machine learning cosmological structure formation". In: MNRAS 479.3, pp. 3405–3414. DOI: 10.1093/mnras/sty1719. arXiv: 1802.04271 [astro-ph.CO].

Maas, Andrew L, Awni Y Hannun, and Andrew Y Ng (2013). "Rectifier nonlinearities improve neural network acoustic models". In: *Proceedings of the 30th International Conference on Machine Learning*. Vol. 28. ICML'13. Atlanta, Georgia: PMLR.

Mangilli, A. and L. Verde (2009). "Non-Gaussianity and the CMB bispectrum: Confusion between primordial and lensing-Rees-Sciama contribution?" In: Phys. Rev. D 80.12, p. 123007. DOI: 10.1103/PhysRevD.80.123007. arXiv: 0906.2317 [astro-ph.CO].

Mangilli, A. et al. (2013). "Optimal bispectrum estimator and simulations of the CMB lensing-integrated Sachs Wolfe non-Gaussian signal". In: A&A 555, A82. DOI: 10.1051/0004-6361/201321433. arXiv: 1303.1722.

Manzotti, A. et al. (2017). "CMB Polarization B-mode Delensing with SPTpol and Herschel". In: ApJ 846, p. 45. DOI: 10.3847/1538-4357/aa82bb. arXiv: 1701.04396.

Martínez, Vicent J. and Enn Saar (2003). "Statistics of galaxy clustering". In: *In: Statistical challenges in astronomy. Third Statistical Challenges in Modern Astronomy (SCMA III) Conference, University*

Park, PA, USA, July 18 - 21 2001. Eric D. Feigelson, G. Jogesh Babu (eds.). New York: Springer, ISBN 0-387-95546-1, 2003, p. 143 - 160. Ed. by Eric D. Feigelson and G. Jogesh Babu, pp. 143–160.

Massey, Richard et al. (2007). "Dark matter maps reveal cosmic scaffolding". In: *Nature* 445.7125, pp. 286–290.

Matsubara, T. and Y. Suto (1996). "Cosmological Redshift Distortion of Correlation Functions as a Probe of the Density Parameter and the Cosmological Constant". In: ApJ 470, p. L1. DOI: 10.1086/310290. eprint: astro-ph/9604142.

Matsumoto, Makoto and Takuji Nishimura (1998). "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator". In: *ACM Trans. Modeling Comput. Simulation* 8.1, pp. 3–30.

McCulloch, Warren S and Walter Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133.

McEwen, Jason D. and Yves Wiaux (2011). "A Novel Sampling Theorem on the Sphere". In: *IEEE Transactions on Signal Processing* 59.12, pp. 5876–5887. DOI: 10.1109/TSP.2011.2166394. arXiv: 1110.6298 [cs.IT].

Melott, A. L., T. Buchert, and A. G. Weiss (1995). "Testing higher-order Lagrangian perturbation theory against numerical simulations. II. Hierarchical models." In: A&A 294, pp. 345–365. arXiv: astro-ph/9404018 [astro-ph].

Metropolis, Nicholas et al. (1953). "Equation of state calculations by fast computing machines". In: *The journal of chemical physics* 21.6, pp. 1087–1092.

Metz, L. et al. (2016). "Unrolled Generative Adversarial Networks". In: *ArXiv e-prints*. arXiv: 1611.02163 [cs.LG].

Millea, M., E. Anderes, and B. D. Wandelt (2017). "Bayesian delensing of CMB temperature and polarization". In: *ArXiv e-prints*. arXiv: 1708.06753.

Modi, C., Y. Feng, and U. Seljak (2018). "Cosmological Reconstruction From Galaxy Light: Neural Network Based Light-Matter Connection". In: *ArXiv e-prints*. arXiv: 1805.02247.

Monaco, P., T. Theuns, and G. Taffoni (2002). "The pinocchio algorithm: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field". In: MNRAS 331, pp. 587–608. DOI: 10.1046/j.1365-8711.2002.05162.x. eprint: astro-ph/0109323.

Monaco, Pierluigi, Enea Di Dio, and Emiliano Sefusatti (2019). "A blind method to recover the mask of a deep galaxy survey". In: J. Cosmology Astropart. Phys. 4, p. 023. DOI: 10.1088/1475-7516/2019/04/023. arXiv: 1812.02104 [astro-ph.CO].

Monaco, Pierluigi et al. (2002). "Predicting the Number, Spatial Distribution, and Merging History of Dark Matter Halos". In: ApJ 564, pp. 8–14. DOI: 10.1086/324182. arXiv: astro-ph/0109322 [astro-ph].

Mortonson, Michael J. and Uroš Seljak (2014). "A joint analysis of Planck and BICEP2 B modes including dust polarization uncertainty". In: J. Cosmology Astropart. Phys. 10, p. 035. DOI: 10.1088/1475-7516/2014/10/035. arXiv: 1405.5857 [astro-ph.CO].

Moutarde, F. et al. (1991). "Precollapse scale invariance in gravitational instability". In: ApJ 382, pp. 377–381. DOI: 10.1086/170728.

Muciaccia, P. F., P. Natoli, and N. Vittorio (1997). "Fast Spherical Harmonic Analysis: A Quick Algorithm for Generating and/or Inverting Full-Sky, High-Resolution Cosmic Microwave Background Anisotropy Maps". In: ApJ 488.2, pp. L63–L66. DOI: 10.1086/310921. arXiv: astro-ph/9703084 [astro-ph].

Mukhanov, Viatcheslav (2005). *Physical foundations of cosmology*. Cambridge University Press.

Mukherjee, S. and B. D. Wandelt (2018). "Beyond the classical distance-redshift test: cross-correlating redshift-free standard candles and sirens with redshift surveys". In: *ArXiv e-prints*. arXiv: 1808.06615.

Münchmeyer, Moritz and Kendrick M. Smith (2019). "Fast Wiener filtering of CMB maps with Neural Networks". In: *arXiv e-prints*. arXiv: 1905.05846 [astro-ph.CO].

Mustafa, Mustafa et al. (2019). "CosmoGAN: creating high-fidelity weak lensing convergence maps using Generative Adversarial Networks". In: *Computational Astrophysics and Cosmology* 6.1, p. 1. DOI: 10.1186/s40668-019-0029-9. arXiv: 1706.02390 [astro-ph.IM].

Naess, Sigurd et al. (2014). "The Atacama Cosmology Telescope: CMB polarization at 200 < l < 9000". In: J. Cosmology Astropart. Phys. 10, p. 007. DOI: 10.1088/1475-7516/2014/10/007. arXiv: 1405.5524 [astro-ph.CO].

Nair, Vinod and Geoffrey E. Hinton (2010). "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *Proceedings of the 27th International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, pp. 807–814. ISBN: 978-1-60558-907-7. URL: http://dl.acm.org/citation.cfm?id=3104322.3104425.

Nash, John (1951). "Non-Cooperative Games". In: *The Annals of Mathematics* 54.2, p. 286. ISSN: 0003486X. DOI: 10.2307/1969529.

Neal, Radford (2000). "Slice Sampling". In: *Annals of Statistics* 31, pp. 705–767.

Neal, Radford M (1993). "Probabilistic inference using Markov chain Monte Carlo methods". In: *Technical Report CRG-TR-93-1*.

— (2003). "Slice sampling". In: *Annals of statistics*, pp. 705–741.

Neyrinck, M. C. et al. (2014). "A halo bias function measured deeply into voids without stochasticity". In: MNRAS 441, pp. 646–655. DOI: 10.1093/mnras/stu589. arXiv: 1309.6641.

Nielsen, Michael A. (2015). *Neural networks and deep learning*. Determination Press.

Nowlan, Steven J and Geoffrey E Hinton (1992). "Simplifying neural networks by soft weight-sharing". In: *Neural computation* 4.4, pp. 473–493.

Ntampaka, Michelle et al. (2019). "The Role of Machine Learning in the Next Decade of Cosmology". In: *arXiv e-prints*, arXiv:1902.10159. arXiv: 1902.10159 [astro-ph.IM].

Nusser, A. (2005). "The Alcock-Paczyński test in redshifted 21-cm maps". In: MNRAS 364, pp. 743–750. DOI: 10.1111/j.1365-2966.2005.09603.x. eprint: astro-ph/0410420.

Nusser, A. and A. Dekel (1992). "Tracing large-scale fluctuations back in time". In: ApJ 391, pp. 443–452. DOI: 10.1086/171360.

O'Dwyer, I. J. et al. (2004). "Bayesian Power Spectrum Analysis of the First-Year Wilkinson Microwave Anisotropy Probe Data". In: ApJ 617, pp. L99–L102. DOI: 10.1086/427386. eprint: astro-ph/0407027.

Oh, S. P., D. N. Spergel, and G. Hinshaw (1999). "An Efficient Technique to Determine the Power Spectrum from Cosmic Microwave Background Sky Maps". In: ApJ 510, pp. 551–563. DOI: 10.1086/306629. eprint: astro-ph/9805339.

Padmanabhan, N. and M. White (2008). "Constraining anisotropic baryon oscillations". In: Phys. Rev. D 77.12, p. 123540. DOI: 10.1103/PhysRevD.77.123540. arXiv: 0804.0799.

Padmanabhan, Thanu (1993). *Structure formation in the universe*. Cambridge university press.

Papež, J., L. Grigori, and R. Stompor (2018). "Solving linear equations with messenger-field and conjugate gradient techniques: An application to CMB data analysis". In: A&A 620, A59. DOI: 10.1051/0004-6361/201832987. arXiv: 1803.03462 [astro-ph.CO].

Peacock, John A (1999). *Cosmological physics*. Cambridge university press.

Peebles, P. J. E. (1980). *The large-scale structure of the universe*. Princeton University Press.

Peebles, P. J. E. and J. T. Yu (1970). "Primeval Adiabatic Perturbation in an Expanding Universe". In: ApJ 162, p. 815. DOI: 10.1086/150713.

Peirani, Sébastien et al. (2014). "LyMAS: Predicting Large-scale Lyα Forest Statistics from the Dark Matter Density Field". In: ApJ 784, p. 11. DOI: 10.1088/0004-637X/784/1/11. arXiv: 1306.1533 [astro-ph.CO].

Penzias, A. A. and R. W. Wilson (1965). "A Measurement of Excess Antenna Temperature at 4080 Mc/s." In: ApJ 142, pp. 419–421. DOI: 10.1086/148307.

Perez, Luis and Jason Wang (2017). "The Effectiveness of Data Augmentation in Image Classification using Deep Learning". In: *arXiv e-prints*. arXiv: 1712.04621 [cs.CV].

Perraudin, N. et al. (2019). "DeepSphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications". In: *Astronomy and Computing* 27, p. 130. DOI: 10.1016/j.ascom.2019.03.004. arXiv: 1810.12186 [astro-ph.CO].

Phillipps, S. (1994). "A Possible Geometric Measurement of the Cosmological Constant". In: MNRAS 269, p. 1077. DOI: 10.1093/mnras/269.4.1077.

Planck Collaboration et al. (2014a). "Planck 2013 results. XXIII. Isotropy and statistics of the CMB". In: A&A 571, A23. DOI: 10.1051/0004-6361/201321534. arXiv: 1303.5083.

— (2014b). "Planck 2013 results. XXIV. Constraints on primordial non-Gaussianity". In: A&A 571, A24. DOI: 10.1051/0004-6361/201321554. arXiv: 1303.5084.

Planck Collaboration et al. (2016a). "Planck 2015 results. IX. Diffuse component separation: CMB maps". In: A&A 594, A9. DOI: 10.1051/0004-6361/201525936. arXiv: 1502.05956 [astro-ph.CO].

Planck Collaboration et al. (2016b). "Planck 2015 results. XIII. Cosmological parameters". In: A&A 594, A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589.

Planck Collaboration et al. (2016c). "Planck 2015 results. XVII. Constraints on primordial non-Gaussianity". In: A&A 594, A17. DOI: 10.1051/0004-6361/201525836. arXiv: 1502.01592 [astro-ph.CO].

Planck Collaboration et al. (2016d). "Planck 2015 results. XVII. Constraints on primordial non-Gaussianity". In: A&A 594, A17. DOI: 10.1051/0004-6361/201525836. arXiv: 1502.01592.

Planck Collaboration et al. (2016e). "Planck intermediate results. XXX. The angular power spectrum of polarized dust emission at intermediate and high Galactic latitudes". In: A&A 586, A133. DOI: 10.1051/0004-6361/201425034. arXiv: 1409.5738 [astro-ph.CO].

Planck Collaboration et al. (2018a). "Planck 2018 results. I. Overview and the cosmological legacy of Planck". In: *ArXiv e-prints*. arXiv: 1807.06205.

Planck Collaboration et al. (2018b). "Planck 2018 results. VI. Cosmological parameters". In: *arXiv e-prints*. arXiv: 1807.06209 [astro-ph.CO].

Planck Collaboration et al. (2019). "Planck 2018 results. IX. Constraints on primordial non-Gaussianity". In: *arXiv e-prints*, arXiv:1905.05697. arXiv: 1905.05697 [astro-ph.CO].

Pollina, G. et al. (2018). "On the relative bias of void tracers in the Dark Energy Survey". In: *ArXiv e-prints*. arXiv: 1806.06860.

Popowski, P. A. et al. (1998). "Quasar Clustering and Spacetime Geometry". In: ApJ 498, pp. 11–25. DOI: 10.1086/305528. eprint: astro-ph/9707175.

Porqueres, Natalia et al. (2019). "Explicit Bayesian treatment of unknown foreground contaminations in galaxy surveys". In: A&A 624, A115. DOI: 10.1051/0004-6361/201834844. arXiv: 1812.05113 [astro-ph.CO].

Puglisi, G. et al. (2018). "Iterative map-making with two-level preconditioning for polarized Cosmic Microwave Background data sets - A worked example for ground-based experiments". In: *A&A* 618, A62. DOI: 10.1051/0004-6361/201832710. arXiv: 1801.08937.

Racca, G. D. et al. (2016). "The Euclid mission design". In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 9904. Proc. SPIE, 99040O. DOI: 10.1117/12.2230762. arXiv: 1610.05508 [astro-ph.IM].

Radford, A., L. Metz, and S. Chintala (2015). "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". In: *ArXiv e-prints*. arXiv: 1511.06434 [cs.LG].

Reiman, D. M. and B. E. Göhre (2018). "Deblending galaxy superpositions with branched generative adversarial networks". In: *ArXiv e-prints*. arXiv: 1810.10098 [astro-ph.IM].

Ribli, Dezső, Bálint Ármin Pataki, and István Csabai (2019). "An improved cosmological parameter inference scheme motivated by deep learning". In: *Nature Astronomy* 3, pp. 93–98. DOI: 10.1038/s41550-018-0596-8. arXiv: 1806.05995 [astro-ph.CO].

Robertson, H. P. (1935). "Kinematics and World-Structure". In: ApJ 82, p. 284. DOI: 10.1086/143681.

— (1936a). "Kinematics and World-Structure II." In: ApJ 83, p. 187. DOI: 10.1086/143716.

— (1936b). "Kinematics and World-Structure III." In: ApJ 83, p. 257. DOI: 10.1086/143726.

Rodríguez, Andres C. et al. (2018). "Fast cosmic web simulations with generative adversarial networks". In: *Computational Astrophysics and Cosmology* 5.1, p. 4. DOI: 10.1186/s40668-018-0026-4. arXiv: 1801.09070 [astro-ph.CO].

Rogers, K. K. et al. (2016). "Spin-SILC: CMB polarization component separation with spin wavelets". In: MNRAS 463, pp. 2310–2322. DOI: 10.1093/mnras/stw2128. arXiv: 1605.01417.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.

Rosasco, Lorenzo et al. (2004). "Are loss functions all the same?" In: *Neural Computation* 16.5, pp. 1063–1076.

Ross, A. J. et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: observational systematics and baryon acoustic oscillations in the correlation function". In: MNRAS 464, pp. 1168–1191. DOI: 10.1093/mnras/stw2372. arXiv: 1607.03145.

Ross, Ashley J. et al. (2011). "Ameliorating systematic uncertainties in the angular clustering of galaxies: a study using the SDSS-III". In: MNRAS 417, pp. 1350–1373. DOI: 10.1111/j.1365-2966.2011.19351.x. arXiv: 1105.2320 [astro-ph.CO].

Rumelhart, David E, Geoffrey E Hinton, Ronald J Williams, et al. (1988). "Learning representations by back-propagating errors". In: *Cognitive modeling* 5.3, p. 1.

Ryden, B. S. (1995). "Measuring Q 0 from the Distortion of Voids in Redshift Space". In: ApJ 452, p. 25. DOI: 10.1086/176277. eprint: astro-ph/9506028.

Saad, Yousef (2003). *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics.

Saadeh, D. et al. (2016). "How Isotropic is the Universe?" In: *Physical Review Letters* 117.13, p. 131302. DOI: 10.1103/PhysRevLett.117.131302. arXiv: 1605.07178.

Sarkar, Prakash et al. (2009). "The scale of homogeneity of the galaxy distribution in SDSS DR6". In: MNRAS 399.1, pp. L128–L131. DOI: 10.1111/j.1745-3933.2009.00738.x. arXiv: 0906.3431 [astro-ph.CO].

Saunders, W. and W. E. Ballinger (2000). "Interpolation of Discretely-Sampled Density Fields". In: *Mapping the Hidden Universe: The Universe behind the Mily Way - The Universe in HI*. Ed. by R. C. Kraan-Korteweg, P. A. Henning, and H. Andernach. Vol. 218. Astronomical Society of the Pacific Conference Series, p. 181. eprint: astro-ph/0005606.

Schawinski, K. et al. (2017). "Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit". In: MNRAS 467, pp. L110–L114. DOI: 10.1093/mnrasl/slx008. arXiv: 1702.00403 [astro-ph.IM].

Schlegel, D. J., D. P. Finkbeiner, and M. Davis (1998). "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds". In: ApJ 500, pp. 525–553. DOI: 10.1086/305772. eprint: astro-ph/9710327.

Schmelzle, J. et al. (2017). "Cosmological model discrimination with Deep Learning". In: *ArXiv e-prints*. arXiv: 1707.05167.

Schmidt, F. et al. (2018). "A rigorous EFT-based forward model for large-scale structure". In: *ArXiv e-prints*. arXiv: 1808.02002.

Schmoldt, I. M. et al. (1999). "On Density and Velocity Fields and beta from the IRAS PSCZ Survey". In: AJ 118, pp. 1146–1160. DOI: 10.1086/301001. eprint: astro-ph/9906035.

Scoccimarro, R. (2000). "The Bispectrum: From Theory to Observations". In: ApJ 544, pp. 597–615. DOI: 10.1086/317248.

Scoccimarro, R. and R. K. Sheth (2002). "PTHALOS: a fast method for generating mock galaxy distributions". In: MNRAS 329, pp. 629–640. DOI: 10.1046/j.1365-8711.2002.04999.x.

Scoccimarro, Roman (1998). "Transients from initial conditions: a perturbative analysis". In: MNRAS 299, pp. 1097–1118. DOI: 10.1046/j.1365-8711.1998.01845.x. arXiv: astro-ph/9711187 [astro-ph].

Scranton, Ryan et al. (2002). "Analysis of Systematic Effects and Statistical Uncertainties in Angular Clustering of Galaxies from Early Sloan Digital Sky Survey Data". In: ApJ 579, pp. 48–75. DOI: 10.1086/342786. arXiv: astro-ph/0107416 [astro-ph].

Scrimgeour, Morag I. et al. (2012). "The WiggleZ Dark Energy Survey: the transition to large-scale cosmic homogeneity". In: MNRAS 425.1, pp. 116–134. DOI: 10.1111/j.1365-2966.2012.21402.x. arXiv: 1205.6812 [astro-ph.CO].

Seljak, U. and C. M. Hirata (2004). "Gravitational lensing as a contaminant of the gravity wave signal in the CMB". In: Phys. Rev. D 69.4, p. 043005. DOI: 10.1103/PhysRevD.69.043005. eprint: astro-ph/0310163.

Seljak, U. and M. Zaldarriaga (1997). "Signature of Gravity Waves in the Polarization of the Microwave Background". In: *Physical Review Letters* 78, pp. 2054–2057. DOI: 10.1103/PhysRevLett.78.2054. eprint: astro-ph/9609169.

Seljebotn, D. S. et al. (2014). "A Multi-level Solver for Gaussian Constrained Cosmic Microwave Background Realizations". In: ApJS 210, p. 24. DOI: 10.1088/0067-0049/210/2/24. arXiv: 1308.5299 [astro-ph.IM].

Seljebotn, D. S. et al. (2019). "Multi-resolution Bayesian CMB component separation through Wiener filtering with a pseudo-inverse preconditioner". In: A&A 627, A98. DOI: 10.1051/0004-6361/201732037.

Seo, H.-J. and D. J. Eisenstein (2003). "Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys". In: ApJ 598, pp. 720–740. DOI: 10.1086/379122. eprint: astro-ph/0307460.

Sheth, R. K. (1995). "Constrained realizations and minimum variance reconstruction of non-Gaussian random fields". In: MNRAS 277, pp. 933–944. DOI: 10.1093/mnras/277.3.933. eprint: astro-ph/9511096.

Shewchuk, Jonathan Richard (1994). *An introduction to the conjugate gradient method without the agonizing pain*. Carnegie-Mellon University. Department of Computer Science.

Shirasaki, Masato, Naoki Yoshida, and Shiro Ikeda (2018). "Denoising Weak Lensing Mass Maps with Deep Learning". In: *arXiv e-prints*. arXiv: 1812.05781 [astro-ph.CO].

Shoji, M., D. Jeong, and E. Komatsu (2009). "Extracting Angular Diameter Distance and Expansion Rate of the Universe From Two-Dimensional Galaxy Power Spectrum at High Redshifts: Baryon Acoustic Oscillation Fitting Versus Full Modeling". In: ApJ 693, pp. 1404–1416. DOI: 10.1088/0004-637X/693/2/1404. arXiv: 0805.4238.

Smith, Adrian FM and Gareth O Roberts (1993). "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 55.1, pp. 3–23.

Smith, K. M. (2006a). "Pseudo-C_ℓ estimators which do not mix E and B modes". In: Phys. Rev. D 74.8, p. 083002. DOI: 10.1103/PhysRevD.74.083002. eprint: astro-ph/0511629.

— (2006b). "Pure pseudo- C_ℓ estimators for CMB B-modes". In: New Astron. Rev. 50, pp. 1025–1029. DOI: 10.1016/j.newar.2006.09.015. eprint: astro-ph/0608662.

Smith, K. M., O. Zahn, and O. Doré (2007). "Detection of gravitational lensing in the cosmic microwave background". In: Phys. Rev. D 76.4, p. 043510. DOI: 10.1103/PhysRevD.76.043510. arXiv: 0705.3980.

Smith, K. M. and M. Zaldarriaga (2007). "General solution to the E-B mixing problem". In: Phys. Rev. D 76.4, p. 043001. DOI: 10.1103/PhysRevD.76.043001. eprint: astro-ph/0610059.

Smith, Michael J. and James E. Geach (2019). "Generative deep fields: arbitrarily sized, random synthetic astronomical images through deep learning". In: *arXiv e-prints*. arXiv: 1904.10286 [astro-ph.IM].

Smoot, G. F. et al. (1992). "Structure in the COBE differential microwave radiometer first-year maps". In: ApJ 396, pp. L1–L5. DOI: 10.1086/186504.

Sousbie, T. et al. (2008). "MoLUSC: A Mock Local Universe Survey Constructor". In: ApJ 678, pp. 569–577. DOI: 10.1086/526507. arXiv: astro-ph/0612166 [astro-ph].

Springel, V. (2005). "The cosmological simulation code GADGET-2". In: MNRAS 364, pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x. eprint: astro-ph/0505010.

Srivastava, Nitish et al. (2014). "Dropout: a simple way to prevent neural networks from overfitting". In: *The Journal of Machine Learning Research* 15.1, pp. 1929–1958.

Starck, J. L., M. J. Fadili, and A. Rassat (2013). "Low-ℓ CMB analysis and inpainting". In: A&A 550, A15. DOI: 10.1051/0004-6361/201220332. arXiv: 1210.6587 [astro-ph.CO].

Stark, D. et al. (2018). "PSFGAN: a generative adversarial network system for separating quasar point sources and host galaxy light". In: MNRAS 477, pp. 2513–2527. DOI: 10.1093/mnras/sty764. arXiv: 1803.08925.

Sudevan, V. et al. (2017). "Improved Diffuse Foreground Subtraction with the ILC Method: CMB Map and Angular Power Spectrum Using Planck and WMAP Observations". In: ApJ 842, p. 62. DOI: 10.3847/1538-4357/aa7334. arXiv: 1612.03401.

Sutter, P. M. et al. (2012). "A First Application of the Alcock-Paczynski Test to Stacked Cosmic Voids". In: ApJ 761, p. 187. DOI: 10.1088/0004-637X/761/2/187. arXiv: 1208.1058.

Sutter, P. M. et al. (2014). "A measurement of the Alcock-Paczyński effect using cosmic voids in the SDSS". In: MNRAS 443, pp. 2983–2990. DOI: 10.1093/mnras/stu1392. arXiv: 1404.5618.

Suzuki, A. et al. (2016). "The Polarbear-2 and the Simons Array Experiments". In: *Journal of Low Temperature Physics* 184, pp. 805–810. DOI: 10.1007/s10909-015-1425-4. arXiv: 1512.07299 [astro-ph.IM].

Sylos Labini, F. et al. (2009). "Absence of self-averaging and of homogeneity in the large-scale galaxy distribution". In: *EPL (Europhysics Letters)* 86.4, p. 49001. DOI: 10.1209/0295-5075/86/49001. arXiv: 0805.1132 [astro-ph].

Szegedy, Christian et al. (2015). "Going Deeper with Convolutions". In: *Computer Vision and Pattern Recognition (CVPR)*. URL: http://arxiv.org/abs/1409.4842.

Szegedy, Christian et al. (2016). "Rethinking the Inception Architecture for Computer Vision". In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, URL: http://arxiv.org/abs/1512.00567.

Szegedy, Christian et al. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning." In: *AAAI Conference on Artificial Intelligence*. Vol. 4, p. 12.

Tassev, Svetlin and Matias Zaldarriaga (2012). "The mildly non-linear regime of structure formation". In: J. Cosmology Astropart. Phys. P. 013. DOI: 10.1088/1475-7516/2012/04/013. arXiv: 1109.4939 [astro-ph.CO].

Taylor, J. F., M. A. J. Ashdown, and M. P. Hobson (2008). "Fast optimal CMB power spectrum estimation with Hamiltonian sampling". In: MNRAS 389, pp. 1284–1292. DOI: 10.1111/j.1365-2966.2008.13630.x. arXiv: 0708.2989.

Tegmark, M. (1997a). "How to Make Maps from Cosmic Microwave Background Data without Losing Information". In: ApJ 480, pp. L87–L90. DOI: 10.1086/310631. eprint: astro-ph/9611130.

— (1997b). "How to measure CMB power spectra without losing information". In: Phys. Rev. D 55, pp. 5895–5907. DOI: 10.1103/PhysRevD.55.5895. eprint: astro-ph/9611174.

Tegmark, M. and G. Efstathiou (1996). "A method for subtracting foregrounds from multifrequency CMB sky maps**". In: MNRAS 281, pp. 1297–1314. DOI: 10.1093/mnras/281.4.1297. eprint: astro-ph/9507009.

Tegmark, M. et al. (1998). "Measuring the Galaxy Power Spectrum with Future Redshift Surveys". In: ApJ 499, pp. 555–576. DOI: 10.1086/305663. eprint: astro-ph/9708020.

Tegmark, M. et al. (2004). "The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey". In: ApJ 606, pp. 702–740. DOI: 10.1086/382125. eprint: astro-ph/0310725.

The Dark Energy Survey Collaboration (2005). "The Dark Energy Survey". In: *arXiv e-prints*, astro–ph/0510346. arXiv: astro-ph/0510346 [astro-ph].

The POLARBEAR Collaboration et al. (2017). "A Measurement of the Cosmic Microwave Background *B*-Mode Polarization Power Spectrum at Sub-Degree Scales from 2 years of POLARBEAR Data". In: *ArXiv e-prints*. arXiv: 1705.02907.

Tröster, Tilman et al. (2019). "Painting with baryons: augmenting N-body simulations with gas using deep generative models". In: MNRAS 487.1, pp. L24–L29. DOI: 10.1093/mnrasl/slz075. arXiv: 1903.12173 [astro-ph.CO].

Trotta, Roberto (2008). "Bayes in the sky: Bayesian inference and model selection in cosmology". In: *Contemporary Physics* 49, pp. 71–104. DOI: 10.1080/00107510802066753. arXiv: 0803.4089 [astro-ph].

— (2017). "Bayesian Methods in Cosmology". In: *arXiv e-prints*. arXiv: 1701.01467 [astro-ph.CO].

Turner, Michael S. and Martin White (1997). "CDM models with a smooth component". In: Phys. Rev. D 56, R4439–R4443. DOI: 10.1103/PhysRevD.56.R4439. arXiv: astro-ph/9701138 [astro-ph].

Vansyngel, F. et al. (2016). "Semi-blind Bayesian inference of CMB map and power spectrum". In: A&A 588, A113. DOI: 10.1051/0004-6361/201424890. arXiv: 1409.0858.

Verde, L. (2010). "Statistical Methods in Cosmology". In: *Lectures on Cosmology Accelerated Expansion of the Universe by Georg Wolschin, Lecture Notes in Physics vol. 800, Springer Berlin / Heidelberg, ISSN 1616-6361; ISBN 978-3-642-10597-5, pp.147-177*. Ed. by Georg Wolschin, pp. 147–177. DOI: 10.1007/978-3-642-10598-2_4.

Villani, Cédric (2008). *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media.

Walker, A. G. (1937). "On Milne's theory of world-structure". In: *Proceedings of the London Mathematical Society* 42, p. 90. DOI: 10.1112/plms/s2-42.1.90.

Wandelt, B. D., D. L. Larson, and A. Lakshminarayanan (2004). "Global, exact cosmic microwave background data analysis using Gibbs sampling". In: Phys. Rev. D 70.8, p. 083511. DOI: 10.1103/PhysRevD.70.083511. eprint: astro-ph/0310080.

Wang, Limin and Paul J. Steinhardt (1998). "Cluster Abundance Constraints for Cosmological Models with a Time-varying, Spatially Inhomogeneous Energy Component with Negative Pressure". In: ApJ 508, pp. 483–490. DOI: 10.1086/306436. arXiv: astro-ph/9804015 [astro-ph].

Webster, M., O. Lahav, and K. Fisher (1997). "Wiener reconstruction of the IRAS 1.2-Jy galaxy redshift survey: cosmographical implications". In: MNRAS 287, pp. 425–444. DOI: 10.1093/mnras/287.2.425. eprint: astro-ph/9608021.

Weinberg, Steven and RH Dicke (1973). "Gravitation and cosmology: principles and applications of the general theory of relativity". In: *American Journal of Physics* 41.4, pp. 598–599.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*. Vol. 2. MIT press, Cambridge, MA.

Witten, Ian H. et al. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Yadav, Jaswant K., J. S. Bagla, and Nishikanta Khandai (2010). "Fractal dimension as a measure of the scale of homogeneity". In: MNRAS 405.3, pp. 2009–2015. DOI: 10.1111/j.1365-2966.2010.16612.x. arXiv: 1001.0617 [astro-ph.CO].

York, D. G. et al. (2000). "The Sloan Digital Sky Survey: Technical Summary". In: AJ 120, pp. 1579–1587. DOI: 10.1086/301513. eprint: astro-ph/0006396.

Zaldarriaga, M. (2001). "Nature of the E-B decomposition of CMB polarization". In: Phys. Rev. D 64.10, p. 103001. DOI: 10.1103/PhysRevD.64.103001. eprint: astro-ph/0106174.

Zaldarriaga, M. and U. Seljak (1997). "All-sky analysis of polarization in the microwave background". In: Phys. Rev. D 55, pp. 1830–1840. DOI: 10.1103/PhysRevD.55.1830. eprint: astro-ph/9609170.

— (1998). "Gravitational lensing effect on cosmic microwave background polarization". In: Phys. Rev. D 58.2, p. 023003. DOI: 10.1103/PhysRevD.58.023003. eprint: astro-ph/9803150.

Zamudio-Fernandez, Juan et al. (2019). "HIGAN: Cosmic Neutral Hydrogen with Generative Adversarial Networks". In: *arXiv e-prints*. arXiv: 1904.12846 [astro-ph.CO].

Zaroubi, S. (2002). "Unbiased reconstruction of the large-scale structure". In: MNRAS 331, pp. 901–908. DOI: 10.1046/j.1365-8711.2002.05229.x. eprint: astro-ph/0010561.

Zaroubi, S., Y. Hoffman, and A. Dekel (1999). "Wiener Reconstruction of Large-Scale Structure from Peculiar Velocities". In: ApJ 520, pp. 413–425. DOI: 10.1086/307473. eprint: astro-ph/9810279.

Zel'Dovich, Y. B. (1970). "Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations." In: A&A 500, pp. 13–18.

Zeldovich, Ya B (1972). "A hypothesis, unifying the structure and the entropy of the Universe". In: MNRAS 160.1, 1P–3P.

Zhang, Xinyue et al. (2019). "From Dark Matter to Galaxies with Convolutional Networks". In: *arXiv e-prints*. arXiv: 1902.05965 [astro-ph.CO].

Zhao, W. and D. Baskaran (2010). "Separating E and B types of polarization on an incomplete sky". In: Phys. Rev. D 82.2, p. 023001. DOI: 10.1103/PhysRevD.82.023001. arXiv: 1005.1201.

Zingales, T. and I. P. Waldmann (2018). "ExoGAN: Retrieving Exoplanetary Atmospheres Using Deep Convolutional Generative Adversarial Networks". In: *ArXiv e-prints*. arXiv: 1806.02906 [astro-ph.IM].