Ludwig-Maximilians-Universität

Sigillum Universitatis Ludovici Maximiliani

Bayesian Methods for Analyzing the Large Scale Structure of the Universe

Dissertation an der Fakultät für Physik

der Ludwig-Maximilians-Universität München

für den Grad des

Doctor rerum naturalium

vorgelegt von Jens Jasche

aus Essen

München, den 12.02.2010

Jens Jasche:
Bayesian methods for analyzing the large scale structure of the Universe
Dissertation der Fakultät für Physik der Ludwig-Maximilians Universität München
ausgeführt am Max-Planck-Institut für Astrophysik

to María de los Ángeles

"We take off into the cosmos, ready for anything: for solitude, for hardship, for exhaustion, death. Modesty forbids us to say so, but there are times when we think pretty well of ourselves. And yet, if we examine it more closely, our enthusiasm turns out to be all a sham. We don't want to conquer the cosmos, we simply want to extend the boundaries of Earth to the frontiers of the cosmos. For us, such and such a planet is as arid as the Sahara, another as frozen as the North Pole, yet another as lush as the Amazon basin. We are humanitarian and chivalrous; we don't want to enslave other races, we simply want to bequeath them our values and take over their heritage in exchange. We think of ourselves as the Knights of the Holy Contact. This is another lie. We are only seeking Man. We have no need of other worlds. A single world, our own, suffices us; but we can't accept it for what it is. We are searching for an ideal image of our own world: we go in quest of a planet, a civilization superior to our own but developed on the basis of a prototype of our primeval past. At the same time, there is something inside us which we don't like to face up to, from which we try to protect ourselves, but which nevertheless remains, since we don't leave Earth in a state of primal innocence. We arrive here as we are in reality, and when the page is turned and that reality is revealed to us - that part of our reality which we would prefer to pass over in silence - then we don't like it anymore."

– Stanisław Lem (Solaris)

Contents

Contents

iv

1. Abstract

Bayesian methods for Large Scale Structure analysis

The cosmic large scale structure is of special relevance for testing current cosmological theories about the origin and evolution of the Universe. Throughout cosmic history, it evolved from tiny quantum fluctuations, generated during the early epoch of inflation, to the filamentary cosmic web presently observed by our telescopes. Observations and analyses of this large scale structure will hence test this picture, and will provide valuable information on the processes of cosmic structure formation as well as they will reveal the cosmological parameters governing the dynamics of the Universe.

Beside measurements of the cosmic microwave backround, galaxy observations are of particular interest to modern precision cosmology. They are complementary to many other sources of information, such as cosmic microwave background experiments, since they probe a different epoch. Galaxies report the cosmic evolution over an enormous period ranging from the end of the epoch of reionization, when luminous objects first appeared, till today. For this reason, galaxy surveys are excellent probes of the dynamics and evolution of the Universe. Especially the Sloan Digital Sky Survey is one of the most ambitious surveys in the history of astronomy. It provides measurements of 930,000 galaxy spectra as well as the according angular and redshift positions of galaxies over an area which covers more than a quarter of the sky. This enormous amount of precise data allows for an unprecedented access to the three dimensional cosmic matter distribution and its evolution. However, observables, such as positions and properties of galaxies, provide only an inaccurate picture of the cosmic large scale structure due to a variety of statistical and systematic observational uncertainties. In particular, the continuous cosmic density field is only traced by a set of discrete galaxies introducing statistical uncertainties in the form of Poisson distributed noise. Further, galaxy surveys are subject to a variety of complications such as instrumental limitations or the nature of the observation itself. The solution to the underlying problem of characterizing the large scale structure in the Universe therefore requires a statistical approach.

The main theme of this PhD-thesis is the development of new Bayesian data analysis methods which provide a complete statistical characterization and a detailed cosmographic description of the large scale structure in our Universe. The required epistemological concepts, the mathematical framework of Bayesian statistics as well as numerical considerations are thoroughly discussed. On this basis two Bayesian data analysis computer algorithms are developed. The first of which is called ARES (Algorithm for REconstruction and Sampling). It aims at the joint inference of the three dimensional density field and its power-spectrum from galaxy observations. The ARES algorithm accurately treats many observational systematics and statistical uncertainties, such as the survey geometry, galaxy selection effects, blurring effects and noise. Further, ARES provides a full statistical characterization of the three dimensional density field, the power-spectrum and their joint uncertainties by exploring the high dimensional space of their joint posterior via a very efficient Gibbs sampling scheme. The posterior is the probability of the model given the observations and all other available informations. As a result, ARES provides a sampled representation of the joint posterior, which conclusively characterizes many of the statistical properties of the large scale structure. This probability distribution allows for a variety of scientific applications, such as reporting any desired statistical summary or testing of cosmological models via Bayesian model comparison or Bayesian odds factors. The second computer algorithm, HADES (Hamiltonian Density Estimation and Sampling), is specifically designed to infer the fully evolved cosmic density field deep

into the non-linear regime. In particular, HADES accurately treats the non-linear relationship between the observed galaxy distribution and the underlying continuous density field by correctly accounting for the Poissonian nature of the observables. This allows for very precise recovery of the density field even in sparsely sampled regions. HADES also provides a complete statistical description of the non-linear cosmic density field in the form of a sampled representation of a cosmic density posterior. Beside the possibility of reporting any desired statistical summary of the density field or power-spectrum, such representations of the according posterior distributions also allow for simple non-linear and non-Gaussian error propagation to any quantity finally inferred from the analysis results. The application of HADES to the latest Sloan Digital Sky Survey data denotes the first fully Bayesian non-linear density inference conducted so far. The results obtained from this procedure represent the filamentary structure of our cosmic neighborhood in unprecedented accuracy.

Bayesische Methoden zur Analyse der großskaligen Struktur

Die großskalige Struktur der kosmischen Materieverteilung ist von besonderer Bedeutung für die Überprüfung moderner kosmologischer Modelle des Ursprungs und der weiteren Entwicklung des Universums. Im Verlauf der kosmischen Geschichte hat sich diese Struktur aus mikroskopischen Quantenfluktuation, welche zur Epoche der Inflation erzeugt wurden, zu dem heute mittels Teleskopen beobachteten faserförmigen kosmischen Netz entwickelt. Beobachtungen und Analysen dieser großskalige Struktur prüfen daher dieses Bild, und liefern wertvolle Information über die Prozesse der kosmologischen Strukturbildung als auch über die kosmologischen Parameter, welche die Dynamik des Universums bestimmen.

Neben Messungen des kosmischen Mikrowellenhintergrundes, sind insbesondere Galaxienbeobachtungen von großem Interesse für die moderne Präzisionskosmologie. Sie ergänzen viele andere Informationsquellen, wie den Mikrowellenhintergrund, da sie Aufschluss über verschiedene kosmische Epochen geben. Galaxien verfolgen die kosmische Entwicklung über enorme Zeiträume, ausgehend vom Ende der Reionisationsepoche, als die ersten leuchtenden Objekte entstanden, bis heute. Daher sind Galaxienbeobachtungen besonders geeignet die Dynamik und Entwicklung unseres Universums zu studieren. Insbesondere der Sloan Digital Sky Survey ist eines der ambitioniertesten Galaxienbeobachtungsprojekte in der Geschichte der Astronomie. In diesem Projekt wurden Spektren von 930.000 Galaxien, so wie deren Winkelpositionen und Rotverschiebungen, in einer Fläche erfasst, welche mehr als ein Viertel des Himmels abdeckt. Diese enorme Datenmenge ermöglicht einen noch nie dagewesenen Zugang zu der dreidimensionalen kosmischen Materieverteilung und deren Entwicklung. Jedoch liefern Beobachtungsgrössen, wie die Positionen und Eigenschaften der Galaxien, wegen einer Vielzahl von statistischen und systematischen Beobachtungsunsicherheiten, nur ein ungenaues Bild der großskaligen Struktur im Universum. Insbesondere wird das kontinuierliche kosmische Dichtefeld nur durch eine Menge von diskreten Galaxien nachgezeichnet, was statistische Unsicherheiten in der Form von poisson-verteiltem Rauschen erzeugt. Des Weiteren unterliegen Galaxienbeobachtungen einer Menge von zusätzlichen Komplikationen, wie instrumentelle Limitierungen oder der intrinsischen Natur der Beobachtung selbst. Die Lösung des unterliegenden Problems, der Charakterisierung der großskaligen Struktur im Universum, bedarf daher im Allgemeinen eines statistischen Ansatzes.

Das zentrale Thema dieser Dissertation ist die Entwicklung neuartiger bayesischer Datenanalysemethoden, welche eine komplette statistische Charakterisierung, so wie eine detaillierte kosmographische Beschreibung der großskaligen Struktur unseres Universum, ermöglichen. Die notwendigen erkenntnistheoretischen Konzepte, der mathematische Rahmen der bayesischen Statistik, als auch numerische Überlegungen werden ausgiebig diskutiert. Auf dieser Basis wird die Entwicklung zweier bayesischer Computeralgorithmen zur Datenanalyse präsentiert. Der erste dieser beiden Algorithmen trägt den Namen ARES (Algorithm for REconstruction and Sampling). Sein Ziel ist die gemeinsame Deduktion des drei-

dimensionalen Dichtefeldes und des zugehörigen Leistungsspektrums von Galaxienbeobachtungen. Der ARES-Algorithmus behandelt mit großer Genauigkeit viele systematische und statistische Unsicherheiten der Galaxienbeobachtung, wie z.B. die Geometrie der beobachteten Himmelsareale, Selektions- und Verschmierungseffekte als auch Rauschen. Des Weiteren liefert ARES eine vollständige statistische Charakterisierung des dreidimensionalen Dichtefeldes, des Leistungsspektrums sowie deren gemeinsamen Unsicherheiten mittels einer Durchmusterung der gemeinsamen Posteriorverteilung durch ein sehr effizientes Gibbs-Abtastungsverfahren. Der Posterior ist die Wahrscheinlichkeit des Modells gegeben die Beobachtungen und alle anderen vorliegenden Informationen. Als Ergebnis liefert ARES daher eine diskretisierte Darstellung der gemeinsamen Posteriorverteilung, welche die statistischen Eigenschaften der großskaligen Struktur überzeugend charakterisiert. Diese Wahrscheinlichkeitsverteilung ermöglicht eine Vielzahl von wissenschaftlichen Anwendungen, beispielsweise das Auswerten jeder beliebigen statistischen Zusammenfassung oder das Testen von kosmologischen Modellen mittels bayesischer Modellvergleiche oder bayesischer Chancenverhältnisse. Der zweite Computeralgorithmus, genannt HADES (Hamiltonian Density Estimation and Sampling), ist speziell auf die Deduktion des voll entwickelten kosmischen Dichtefeldes in besonders nichtlinearen Regimen ausgelegt. Im Besonderen behandelt HADES die stark nichtlineare Beziehung zwischen der beobachteten Galaxienverteilung und dem unterliegenden kontinuierlichem Dichtefeld mit großer Genauigkeit, durch die explizite Berücksichtigung der poissonischen Eigenschaften der Observablen. HADES liefert eine komplette statistische Beschreibung des nichtlinearen kosmischen Dichtefeldes in der Form einer abgetasteten Darstellung der kosmischen Dichteposteriorverteilung. Neben der schon erwähnten Möglichkeit sämtliche statistischen Zusammenfassungen auszuwerten, bietet eine derartige Darstellung der entsprechenden Posteriorverteilung einfache Methoden zur nichtlinearen und nichtgaussschen Fehlerfortpflanzung zu jeglicher Messgrösse, die schlussendlich aus dem Deduktionsergebnis ermittelt werden soll. Die Anwendung von HADES auf die jüngsten Sloan Digital Sky Survey Daten stellt die erste komplett bayesische Deduktion des nichtlinearen kosmischen Dichtefeldes dar. Die Ergebnisse diese Prozedur zeigen die filamentare Struktur unserer kosmischen Nachbarschaft in noch nie dagewesener Detailtreue auf.

2. Introduction and motivation

2.0.1. Motivation

The subject of cosmology has always been amongst the prime interests of humanity, since it tries to describe and explain the origin and the properties of the Universe and eventually the existence of man. While originally being defined by religious or philosophical ideology, over the past centuries natural sciences started to govern our modern view of the Universe. Particularly within the last century great advancements have been made, elevating the subject of cosmology from a mere philosophical to an accurate scientific discipline. Physics and Astrophysics hereby played the crucial role in shaping the understanding of the Universe by establishing a scientific theory about its origin and evolution. The foundations of modern cosmological models have been laid by the early discovery of E. Hubble and the development of general relativity by A. Einstein. Hubble observed that the recession velocities of galaxies are proportional to their distances, indicating that spacetime itself is not static, as expected from Newtonian theories, but expanding. This observation induced the development of relativistic cosmological models based on solutions of Friedmann's equations by A. Einstein and W. de Sitter. These solutions describe an expanding universe suggesting that it must have originated from a much hotter and denser state as can be observed today. The idea of a hot Big Bang as the origin of the Universe was further shaped by theoretical investigations of R. Alpher, H. Bethe and G. Gamov. They studied the thermal history of an expanding universe and found that at early times it was hot and dense enough to allow for thermonuclear synthesis of light elements. Support for their theory was given by measurements of the cosmic abundance of light elements, in particular of deuterium. Their theory further predicted the existence of a cosmic microwave background, which was subsequently detected by A. A. Penzias and R. W. Wilson.

In the course of the last and the beginning of this century, cosmologists have constrained the parameters governing the homogeneous dynamics of the Universe, as described by Friedmann's equations, to the few percent level and cosmology turned to investigating the cosmic inhomogeneities.

According to the current cosmological paradigm, all observed structures have their origin in microscopic primordial quantum fluctuations generated during the first split seconds of the Universe. I. Novikov and Y .B. Zel'dovich suggested that over the following 13.6 billion years, these tiny seed perturbations formed the presently observed matter distribution via gravitational amplification. J. Peebles then realized that baryonic models of structure formation are insufficient to explain observations and introduced a new not electromagnetically interacting matter component. Therefore, currently it is believed that structure formation is governed by the gravitational aggregation of such a dark matter fluid. As proposed by J. P. Ostriker, M. Rees and S. D. M. White, luminous objects like galaxies then form inside dark matter structures by condensation and cooling of baryons.

The process of structure formation hence involves very exciting physics ranging from quantum field theory, general relativity to the dynamics of collisionless dark matter and the behavior of the baryonic sector. Throughout cosmic history, these processes imprinted a wealth of information on the origin and evolution of the Universe to the large scale matter distribution as we observe it today. Thus, observations of the cosmic large scale structure have the potential to answer many open questions in physics and cosmology, such as:

- What does the Universe consist of?

- Is dark matter hot or cold?

- What is the dynamical behavior of our Universe?

- How is matter distributed?

- Does dark energy exist and how does it behave?

- Is our model of gravity correct or does it require modification?

- How do galaxies form, and how are they tracing dark matter?

- What were the conditions in the early Universe?

Harvesting this information from present and future probes of the large scale structure, such as galaxy surveys or cosmic microwave background experiments hence allows for testing current physical and cosmological theories. Especially, the large scale distribution of galaxies provides information about the clustering of dark matter on small scales and the transition from linear to nonlinear structure formation. Theories of cosmological structure formation are usually tested by determining the statistical properties of the dark matter density field or any tracer of it, like the galaxy distribution. In particular, statistical characterization in terms of power-spectra or n-point correlation functions has become a major method for large scale structure analyses over the past decades. Precise determination of the overall shape of the power-spectrum can for instance place important constraints on neutrino masses, help to identify the primordial power-spectrum and break degeneracies for cosmological parameter estimation from cosmic microwave background data by measuring the parameter combination Ω_m/h. In addition, several characteristic length scales have been imprinted to the matter distribution, which can serve as new standard rulers to measure the Universe. A prominent example of these length scales is the sound horizon, which yields oscillatory features in the power-spectrum, the so-called baryon acoustic oscillations. Such a new, precise standard ruler will enable us to measure the Universe through the distance redshift relation and to test various dark energy scenarios. The aim of all these observations is to determine the cosmological parameters governing the cosmological structure formation to a degree of accuracy comparable to those governing the homogeneous dynamics of the Universe. Beside direct observations, theories of structure formation are investigated with numerical computer simulations. These computer simulations solve the equations of motion for the matter content in the Universe and provide insight into dark matter dynamics in the nonlinear stages of structure formation and baryonic physics in the core of galaxies and clusters of galaxies.

However, contact between theory and observations cannot be made directly, since observational data is subject to a variety of systematic effects and statistical uncertainties. Careful data analysis methods are required to prescind the signal we seek to measure from all known systematic and statistical uncertainties. Also unique recovery of the underlying signal from observations afflicted with noise is generally not possible due to the stochastic nature of the noise contribution. Therefore, precision data analysis has to rely on statistical methods in order to provide the required uncertainty information and confidence limits and to allow for authentic testing of current cosmological models. The scientific progress of any natural science is thus crucially determined by the four scientific steps of theoretical modeling, simulating, observing and analyzing the data as outlined in Fig. 2.1. Theoretical models of the world are built in order to describe the behavior of physical systems. These models characterize our understanding of the Universe. Testable predictions of these theories are provided by simulations, which follow the behavior of the physical system to a state expected to be observable in our experiments. Experiments, conducted to test the physical system, yield a set of measurements. These measurements do not allow for direct comparison to the results of simulations due to the fact that there exists no ideal observation device in

6
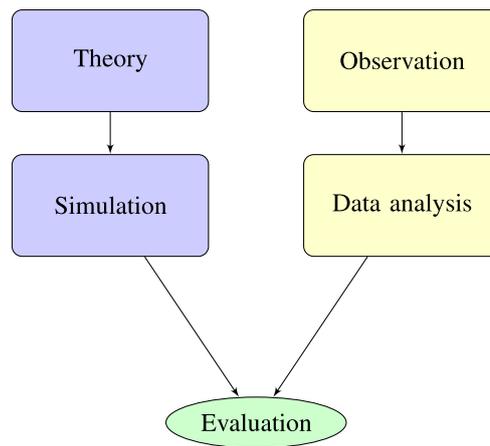
Figure 2.1.: The scientific progress in natural sciences depends on the four steps of theoretical modelling, simulations, observations and data analysis. The results of simulations and data analysis can be compared to judge and evaluate current models.

reality, and each measurement is subject to a variety of uncertainties. A following data analysis procedure must abstract the signal one seeks to recover from those uncertainties and must also provide a measure of confidence in the obtained result. The agreement of the theoretical prediction and the result of the data analysis can then be compared within the confidential limits, and only then a theory can be evaluated or judged to be plausible or not. This demonstrates the importance of data analysis procedures for any natural science including cosmology.

This thesis is dedicated to the development of new Bayesian methods to perform joint inferences of full three dimensional density fields and their power-spectra from cosmological data sets, especially from galaxy redshift surveys. In particular, the developed computer algorithms are designed to correct observational systematics, such as survey geometries and selection effects, and statistical uncertainties, such as noise and cosmic variance. Based on very efficient Markov Chain Monte Carlo algorithms, required to explore extremely high dimensional parameter spaces, these algorithms provide sampled representations of the according posterior distributions conditional on galaxy observations. These posterior distributions contain the full joint information on the signal one seeks to recover and its uncertainties. Any desired statistical summary can easily reported and theoretical models can be estimated via Bayesian model comparison. The data analysis methods as presented in this thesis provide a full statistical characterization of the observed large scale structure and its uncertainties as well as a sound cosmographic description in the form of three dimensional maps of the cosmic density field. Results obtained from the applications of these methods to the latest Sloan Digital Sky Survey data, therefore allow for precise studies of the cosmological structure formation, the correlation of galaxy properties and the large scale structure as well as for the analysis of the cosmological parameters governing the dynamics of the Universe.

2.0.2. Structure

In chapter 3 an introduction to the key concepts of modern cosmology and the theory of cosmic structure formation, as relevant for this thesis, is presented. A short description of the Sloan Digital Sky Survey, its instrumentation and scientific achievements is given in Chapter 4. The Sloan Digital Sky Survey is one of the most ambitious surveys in the history of astronomy. It covers an area of more than a quarter of the sky and provides angular and precise spectroscopic redshift positions of about $930,000$ galaxies. This enormous amount of data provides the opportunity to analyze the three dimensional large scale structure

to unprecedented details.

Chapter 5 is more technical in nature. Here the basic concepts of data acquisition and processing with digital computers are presented. In particular, the implications of Shannon's sampling theorem and the application of Fast Fourier Transform techniques for information recovery from sampled data systems are discussed to great detail. Further, a super sampling method is presented which can greatly alleviate sampling artifacts, such as aliasing, in power-spectrum estimations. It is particularly designed for iterative data analysis methods, such as Wiener filtering, in which artificial mode coupling, as introduced by systematic observational effects, is to be corrected correctly.

In Chapter 6, basic epistemological concepts and the mathematical framework of Bayesian statistics and statistical data analysis are presented. A brief overview of frequently encountered statistical and systematic uncertainties in galaxy redshift surveys and a motivation for the requirement of statistical data analysis is provided. Further, Bayesian and conventional statistics are compared and differences, particularly concerning the notion of probability, are highlighted. Then key concepts of Markov Chain Monte Carlo techniques will be presented and discussed to a degree required for this thesis.

Chapter 7 describes the development and implementation of the Bayesian data analysis computer algorithm ARES (Algorithm for REconstruction and sampling). This algorithm aims at the joint inference of three dimensional density fields and power-spectra from a given cosmological data set by exploring the joint posterior distribution via an efficient Gibbs sampling scheme. In this fashion, ARES provides a sampled representation of the joint posterior distribution from which any desired statistical summary, such as mean, mode or variance, can be easily reported. Mock tests demonstrate the ability of ARES to recover the power-spectra from observations with highly structured survey geometries, selection effects and noise to great accuracy. In particular, the problem of artificial mode coupling can be accounted for correctly yielding a high detectability of the baryon acoustic oscillations in the inferred power-spectrum.

In Chapter 8, a new method for non-linear three dimensional density field inference is presented. The resultant computer algorithm HADES (HAmiltonian Density Estimation and Sampling), provides samples from a lognormal Poissonian posterior via an efficient hybrid Hamiltonian Monte Carlo sampler. The method is tested in various mock scenarios, which demonstrate the ability of HADES to recover the fully evolved non-linear density field to great accuracy.

Chapter 9 presents the application of the non-linear sampler HADES to the latest Sloan Digital Sky Survey data. This procedure yields a sampled representation of the three dimensional density posterior conditional on galaxy observations. The estimated ensemble mean density field reveals the filamentary cosmic-web as predicted by numerical simulations. Further, a dynamical cosmic web type analysis is conducted to identify the four different structure types halos, filaments, sheets and voids. A preliminary analysis of correlations between galaxy properties and large scale environment is also provided.

In Chapter 10 the core results of this thesis are summarized. In addition, future applications and possible extensions of the developed methods are pointed out.

3. Cosmology and cosmic structure formation

Abstract

This chapter provides an overview of the current cosmological theory and the process of cosmic structure formation as relevant for this thesis. In Sect. 3.1 key concepts of general relativistic Friedmann-Lemaître cosmological models will be discussed, followed by the presentation of the theory of cosmological structure formation in Sect. 3.2 and the description of the statistical properties of the large-scale structure is presented in Sect. 3.3.

3.1. Friedmann-Lemaître cosmological models

The overall evolution of the Universe from the beginning to its present state is generally governed by gravitational interactions. For this reason, modern cosmological world models are based on the theory of general relativity, as presented by A. Einstein in 1915 (Einstein 1915). Einstein's theory describes gravity as an emergent geometric property of space and time. In the following, a brief overview over the foundations of modern cosmological world models is given.

3.1.1. A general relativistic universe

In general relativity spacetime is a four-dimensional manifold described by a metric tensor $g_{\mu\nu}$, which defines the geometric relation between two spacetime events. Such spacetime events are given by their four dimensional world coordinates, described by four-vectors x^μ containing the time coordinates and three spatial coordinates. Any infinitesimal separation ds between these events, separated by dx^μ, can then be calculated with the metric tensor by $ds = g_{\mu\nu}dx^\mu dx^\nu$. Here we used Einstein summation convention, which means to sum over repeated indices. According to general relativity the components of the metric tensor $g_{\mu\nu}$ are the dynamical variables which generate the gravitational interaction.

A cosmological world model hence requires to describe the dynamical behavior of the metric tensor along with that of the material and energetic contents of the Universe. The dynamics of such a model can be derived from a generalized Einstein-Hilbert action (see e.g. Gibbons & Hawking 1977, Carroll 2004):

$$S = \int d^4x \sqrt{-g}\, \mathcal{L} = \int d^4x \sqrt{-g}\, \left(\mathcal{L}_g + \mathcal{L}_m + \mathcal{L}_q \right), \tag{3.1}$$

where $\mathcal{L}_g$, $\mathcal{L}_m$ and $\mathcal{L}_q$ are the Lagrange densities for the metric tensor, the matter field and a possible dark energy component respectively. For models of Einstein gravity with a cosmological constant the sum of Lagrangians for gravity and dark energy can be written as:

$$\mathcal{L}_g + \mathcal{L}_q = \frac{c^4}{16\pi G}\, (R - 2\Lambda) \tag{3.2}$$

where $R = g^{\mu\nu} R^\delta_{\mu\delta\nu}$ is the Ricci scalar, with $R^\delta_{\mu\gamma\nu}$ being the Riemann tensor[1], and Λ denotes the cosmological constant. Variation of the Einstein-Hilbert action, given in eq. (3.1), with respect to the metric tensor $g_{\mu\nu}$ yields then Einstein's field equations (Landau & Lifshitz 1975, Carroll 2004):

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{R}{2}g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} + \Lambda g_{\mu\nu}, \tag{3.3}$$

---

[1]The Riemann tensor is defined as $R^\delta_{\mu\gamma\nu} \equiv \partial_\gamma \Gamma^\delta_{\mu\nu} - \partial_\nu \Gamma^\delta_{\mu\gamma} + \Gamma^\rho_{\mu\nu}\Gamma^\delta_{\rho\gamma} - \Gamma^\rho_{\mu\gamma}\Gamma^\delta_{\rho\nu}$, where the individual terms are the Christoffel symbols (for details see Landau & Lifshitz 1975).

where $G_{\mu\nu}$ is is the Einstein tensor, $R_{\mu\nu} = R^{\delta}_{\mu\delta\nu}$ is the Ricci tensor and $T_{\mu\nu}$ is the energy momentum tensor of the matter component obtained by variation of $\mathcal{L}_m$ with respect to the metric tensor (see e.g. Gibbons & Hawking 1977, Carroll 2004). Einstein's equation form a set of ten independent second-order differential equations for the metric tensor $g_{\mu\nu}$ (Carroll 2004). However, the Bianchi identity $\nabla^{\mu}G_{\mu\nu} = 0$, with ∇^{μ} being the covariant derivative, provides four constraints on the Ricci tensor which reduces the amount of equations to six truly independent equations, as is required for the metric to be a solution to Einstein's equations in any coordinate system. Another consequence of the contracted Bianchi identity is that the energy momentum tensor is a conserved current, i.e. $\nabla^{\mu}T_{\mu\nu} = 0$. The major difficulty for solving Einstein's equations arises from the high degree of non-linearity. Hence, two known solutions can generally not be superposed to find a new one. Solutions of the Einstein's equations therefore often require simplifying assumptions as will be discussed in the following. Also note, that the action given in eq. (3.1) is the most fundamental way to summarize our current understanding of the Universe.

3.1.2. Cosmological principles and the Robertson-Walker metric

Einstein's equations are fundamentally governing the dynamical evolution of the Universe. Hence, any conclusive cosmological model must be based on a solution to this set of non-linear equations. Since no general solution to Einstein's equations is known, one usually has to be content with an ansatz for the metric tensor. In order to further reduce the number of coupled equations to solve, one usually requires metric tensor to be highly symmetric. Such an ansatz for the metric in a cosmological model can be found by requiring the metric to fulfill the cosmological principles, as stated below.

3.1.2.1. The cosmological principles

In modern standard cosmological theories, Einstein's equations are commonly solved by assuming a spherically symmetric metric tensor $g_{\mu\nu}$, which describes the expansion dynamics of the Universe. This simplifying assumption for the cosmological model is based on two fundamental postulates.

- **Isotropy:**
  When averaged over sufficiently large scales, there exists a mean motion of matter and radiation in the Universe. For an observer co-moving with such mean motion, all averaged observables appear to be isotropic.

- **Homogeneity:**
  There exists no preferred position for any observer following this mean motion. Thus, all observers measure the same averaged observables and experience the same history of the Universe (also known as the Copernican principle (Peacock 1999)).

These postulates reflect our believe that the large scale universe is homogeneous and isotropic. Throughout most of the twentieth century, when the foundations for modern cosmology were developed, in particular by Friedmann and Lemaître, the cosmological principles were an educated guess to solve Einstein's equations. Today firm empirical data exists, confirming the large scale homogeneity and isotropy of the Universe (Mukhanov 2005). Support for the first postulate is given by the observation of perfect rotational invariance of the cosmic microwave background (CMB) temperature in the co-moving frame. By assuming the isotropy of the spatial hyper-surface around any such co-moving observer and applying the cosmological principle in space yields isotropy around any point on the spatial hyper-surface. Given the space-time metric is an analytic function of the coordinates, this immediately implies homogeneity. The observable region of the Universe is on the order of 3 Gpc, and modern galaxy redshift surveys suggest that the Universe is only homogeneous and isotropic when averaged on scales $\gtrsim 100$ Mpc. On smaller scales the Universe is highly inhomogeneous, filled with aggregations of matter, such as galaxies, clusters and superclusters (Mukhanov 2005). Therefore, the cosmological principle is only applicable within a limited range of scales.

3.1.2.2. The Robertson-Walker metric

The assumption of the cosmological principles singles out a preferred class of fundamental, or co-moving, observers for whom the universe appears to be homogeneous and isotropic. The properties of the metric tensor can then be best

deduced from a coordinate system x^μ in the reference frame of such a fundamental observer. Then the spacetime interval $\mathrm{d}s^2$ can be separated out (Padmanabhan 1993):

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = g_{00}\,c^2\,\mathrm{d}t^2 + 2\,g_{0i}\,c\,\mathrm{d}t\,\mathrm{d}x^i + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\,, \tag{3.4}$$

where Latin indices denote summations only over the spatial components. In the reference frame of a fundamental observer one may use the proper time of such an observer to label the spatial hyper-surfaces, which implies $g_{00} = 1$. Further, if the metric tensor is to fulfill the requirement of isotropy the components g_{i0} must vanish, since otherwise a particular direction in space is preferred. Thus, a metric which satisfies the cosmological principles yields a line element of the form:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 + g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\,, \tag{3.5}$$

Since isotropy has to be conserved, the spatial part of the metric tensor may only scale with a function $a(t)$ depending solely on the proper cosmic time t, yielding the line element:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\mathrm{d}q^2, \tag{3.6}$$

where $\mathrm{d}q$ is the line element on the spatial hyper-surfaces. The Robertson-Walker line element for a homogeneous and isotropic universe is then obtained by introducing spherical coordinates $q = (w, \theta, \phi)$ (Misner et al. 1973, d'Inverno 1992):

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\left[\mathrm{d}w^2 + f_K^2(w)\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right]. \tag{3.7}$$

Due to the requirement of homogeneity the function $f_K(w)$ has to be either trigonometric for positive values of the curvature K, linear for vanishing K or hyperbolic for negative K:

$$f_K(w) = \begin{cases} \frac{1}{\sqrt{K}}\sin\left(\sqrt{K}w\right) & , (K > 0),\ \text{spherical}, \\ w & , (K = 0),\ \text{flat}, \\ \frac{1}{\sqrt{|K|}}\sinh\left(\sqrt{|K|}w\right), & (K < 0),\ \text{hyperbolic}. \end{cases} \tag{3.8}$$

As one can see the highly symmetric Robertson-Walker metric is completely determined by two parameters, the curvature K and the cosmic scale factor $a(t)$. The curvature parameter K distinguishes between different geometries for the spatial hyper-surfaces. An open universe is determined by $K < 0$, a flat by $K = 0$ and a closed universe by $K > 0$. Today it is commonly assumed that the Universe is flat or at least close to flat. The cosmological scale factor on the other hand, describes the conformal mapping between hyper-surfaces separated by time-like vectors. From a dynamical point of view this parameter describes the dynamical behavior of the Universe. Hence, the cosmic scale factor and the dependence of K on the matter content of the Universe uniquely determine the spacetime of a homogeneous and isotropic universe.

3.1.2.3. The cosmological redshift

The expansion of the Universe, due to the cosmological scale factor, has consequences for observing distant light sources. In an expanding universe, photons of wavelength λ_s emitted by a co-moving light source at time t_s will generally be observed with the redshifted wavelength λ_o at t_o. Since photons are massless particles, they travel on null geodesics of zero proper time, implying the relation $c^2\mathrm{d}t^2 = a^2(t)\mathrm{d}q^2$. The co-moving distance d between emitter and observer can then be calculated according to:

$$d = \int_0^d \mathrm{d}q = \int_{t_s}^{t_o} \mathrm{d}t\,\frac{c}{a(t)}\,. \tag{3.9}$$

Already by construction, the co-moving distance between emitter and observer is constant. For this reason, subsequently emitted photons at the time $t_s + \lambda_s/c$ which are observed at the time $t_o + \lambda_o/c$ will have traveled the same co-moving distance d. One can therefore write:

$$\int_{t_s}^{t_o} \mathrm{d}t\,\frac{c}{a(t)} = \int_{t_s+\lambda_s/c}^{t_o+\lambda_o/c} \mathrm{d}t\,\frac{c}{a(t)}\,. \tag{3.10}$$

Parameter	symbol	value
Hubble parameter	h	0.705 ± 0.013 [1]
Total matter density	Ω_m	0.27 ± 0.04 [2]
Baryon density	Ω_b	0.0456 ± 0.0015 [1]
Cosmological constant	Ω_Λ	0.726 ± 0.015 [1]
Curvature parameter	Ω_c	0.02 ± 0.02 [2]

[1] Komatsu et al. (2009)
[2] Spergel et al. (2003)

Table 3.1.: Cosmological parameters.

If the cosmic scale factor $a(t)$ is approximately constant over the small time intervals λ_s/c and λ_o/c then eq. (3.10) yields the relation:

$$\frac{\lambda_o}{\lambda_s} = \frac{a(t_o)}{a(t_s)} \equiv 1 + z \,, \tag{3.11}$$

where $1 + z$ is the relative change in wavelength, and z is referred to as the cosmological redshift. Note that this redshift is not the same as the conventional Doppler effect in special relativity, since it is due to the expansion of space itself, not to the relative velocities of the observer and emitter (Carroll 2004).

3.1.3. Friedmann world model

To study the average dynamical behavior of the Universe, Friedmann proposed a solution to Einstein's equations, which describe an expanding homogeneous and isotropic universe (Friedmann 1922, 1924). This model can be derived by assuming the contents of the Universe to behave as a homogeneous perfect fluid. The energy momentum tensor of such a fluid is completely determined by the energy density ρ, the pressure p and its four velocity u_μ:

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u_\mu \, u_\nu - p \, g_{\mu\nu} \tag{3.12}$$

Together with the Robertson-Walker ansatz for the metric tensor $g_{\mu\nu}$, this assumption allows for reducing the set of Einstein's equations given in eq. (3.3) to two ordinary differential equations:

$$\frac{\dot{a}}{a} = \sqrt{\frac{8\pi G}{3}\rho - K\frac{c^2}{a^2} + \frac{\Lambda c^2}{3}} \quad \text{and} \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \tag{3.13}$$

which describe the time evolution of the cosmic scale factor $a(t)$ depending on the energy density $\rho(t)$ and the pressure $p(t)$ of a cosmological fluid. By virtue of the Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$, and consequently $\nabla^\mu T_{\mu\nu} = 0$, one can further derive the adiabatic equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3(t)\, c^2\, \rho(t)\right] + p(t)\frac{\mathrm{d}}{\mathrm{d}t}a^3(t) = 0, \tag{3.14}$$

which describes the time evolution of the energy contained in a fixed co-moving volume expanding with the Hubble flow. Further, any change in internal energy $\mathrm{d}\left(a^3(t)\, c^2\, \rho(t)\right)$ inside a volume is equal to the work $p(t)\mathrm{d}\left(a^3(t)\right)$ done to change the proper volume. The adiabatic equation therefore corresponds to the first law of thermodynamics applied to the cosmological expansion. Alternatively this adiabatic equation can be derived by combining the two Friedmann equations (3.13). The expansion dynamics of the Universe is conveniently characterized by the Hubble function $H(t)$, which is defined as the logarithmic derivative of the cosmic scale factor $a(t)$:

$$H(t) \equiv \frac{\mathrm{d}}{\mathrm{d}t}\ln(a) = \frac{\dot{a}}{a} \quad \Longrightarrow \quad H^2(t) = H_0^2\left[\frac{8\pi G}{3}\rho - K\frac{c^2}{a^2} + \frac{\Lambda c^2}{3}\right]. \tag{3.15}$$

Equation (3.15) can be more conveniently written when expressing energy densities in units of the critical density

ρ_{crit}:

$$\rho_{\text{crit}} \equiv \frac{3H_0^2}{8\pi G}, \tag{3.16}$$

which defines the critical point between an expanding and a contracting universe. This means, if the sum of all cosmological fluid densities gives ρ_{crit}, the curvature parameter K vanishes and all spatial hyper-surfaces are flat. Expressing the energy densities of all cosmological fluids (matter ρ, curvature K, the cosmological constant Λ and radiation $3p/c^2$) in units of ρ_{crit} :

$$\Omega_m = \frac{\rho}{\rho_{\text{crit}}}, \quad \Omega_r = \frac{8\pi G p}{c^2 H_0^2}, \quad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \quad \Omega_c = \frac{Kc^2}{H_0^2}, \tag{3.17}$$

allows for rewriting the Hubble function in its usual form as:

$$H^2(t) = H_0^2 \left[\frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} - \frac{\Omega_c}{a^2} + \Omega_\Lambda \right]. \tag{3.18}$$

Today, the cosmological parameters governing the homogeneous evolution of the Universe have been measured to a few percent accuracy. Particularly measurements of the CMB temperature fluctuations, as carried out by the WMAP satellite, in combination with Type Ia supernovae and baryon accoustic oscillations data, provide the most precise estimates for the set of cosmological parameters (Spergel et al. 2003, Komatsu et al. 2009). Currently, the most uncertain parameter is the value of the Hubble function at the present time $H(t_0) = H_0 = 100\,h\,\text{kms}^{-1}\,\text{Mpc}^{-1}$, with h expressing its uncertainty. Nevertheless, measurements of CMB anisotropies and from Cepheid variable stars in distant galaxies now converge on a value of $h \simeq 0.7$ (Freedman et al. 2001, Spergel et al. 2003, Komatsu et al. 2009). Table 3.1 shows the values for the cosmological parameters matter density Ω_m, baryonic density Ω_b, curvature $\Omega_c =$ and cosmological constant Ω_Λ. Since the energy density of radiation decreases rapidly Ω_r does not play a crucial role for the cosmological dynamics at the present epoch. One of the most important and interesting questions about the homogeneous evolution of the Universe concerns the origin and the nature of the cosmological constant Λ. Theoretical models propose that the currently observed accelerated expansion of the Universe can be explained by a new homogeneous dynamical component of the cosmological fluid the so-called dark energy, rather than by a cosmological constant. These theories suggest that the accelerated expansion of the Universe is driven by a new quantum scalar field ϕ_q, which exhibits negative pressure (Wetterich 1995, Doran & Wetterich 2003). At present this dark energy component constitutes about 73% of the total energy density in the Universe.

3.1.4. Distance measures in cosmology

General relativity describes gravity as a geometric property of the spacetime manifold. Since the metric tensor $g_{\mu\nu}$ itself is a dynamical field, the notion of "distance" has no longer a unique meaning in an arbitrary curved and non-stationary spacetime. For this reason, in cosmology there exist many ways to specify "distances" between two points on the manifold. A unifying aspect of all these possible measures of "distance" is that they all estimate somehow the separation between two spacetime events on radial null-geodesics $ds^2 = c^2 dt^2 - a^2(t)dw^2 = 0$ (Hogg 1999). In general relativity "distances" are therefore measured by traveling times of light signals emitted by a source at t_s and observed at t_o (e.g. Bartelmann & Schneider 2001, Hogg 1999). Below the co-moving and proper distance as required in the context of this thesis will be described.

Co-moving distance

The co-moving distance $d_{\text{com}}(z_o, z_s)$ measures the radial distance between the world lines of two fundamental observers on the same spatial hyper-surface labeled by the cosmic time $t = t_0$. The distance between two spacetime events, the emission of light at the source at time t_s and subsequent observation at time t_o, can be measured by following photons along null-geodesics $ds = 0$. In analogy to the calculation of the cosmological redshift (see section 3.1.2.3) one yields the co-moving distance between source and emitter as:

$$d_{\text{com}}(z_o, z_s) \equiv \int_{w(z_s)}^{w(z_o)} dw = \int_{t(z_s)}^{t(z_o)} dt \frac{c}{a(t)} = \int_{a(z_s)}^{a(z_o)} da \frac{c}{a^2 H(a)} \tag{3.19}$$
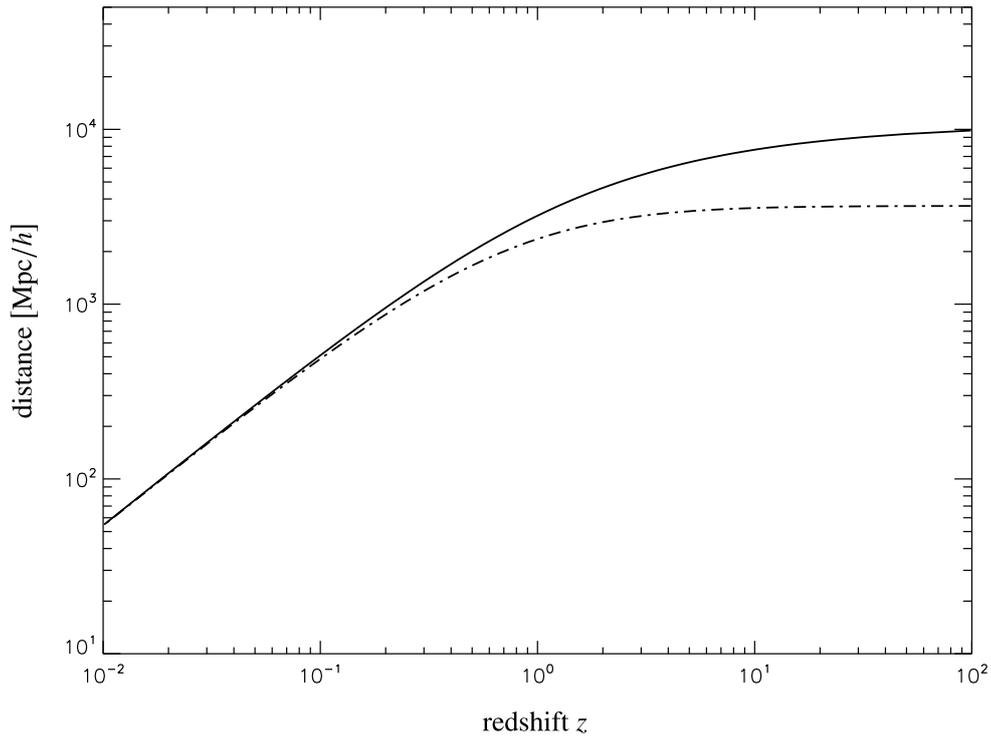
Figure 3.1.: Two different distance measures in cosmology: The co-moving distance $d_{com}(0, z)$ (solid line) and the proper distance $d_{prop}(0, z)$ (dash-dotted line).

where the Hubble function is given by eq. (3.15). Here we used the fact that the Hubble function is the logarithmic derivative of the cosmic scale factor with respect to time. Also note that due to eq. (3.19) comoving distances are additive.

Proper distance and look back time

The proper distance measures the elapsed coordinate time a photon needs to travel from the source to the observer. It can be obtained in a similar way as the co-moving distance by integration along null-geodesics $ds = 0$:

$$d_{prop}(z_o, z_s) = \int_{a(z_s)}^{a(z_o)} da \frac{c}{aH(a)} \,. \tag{3.20}$$

Dividing the proper distance by the speed of light c yields the look back time. The look back time and the proper distance are also additive. In Fig. 3.1 co-moving and proper distance are compared.

3.2. Structure formation

The outstanding problem in today's cosmology is the origin and evolution of large scale structure. The solution of this problem will give us better understanding of the epoch of galaxy formation, the clustering in the galaxy distribution as well as sheds light on the amplitudes and form of anisotropies in the microwave backround. At present it is commonly accepted that all observed structure in the Universe formed via gravitational instability from primordial microscopic quantum fluctuations, generated at the Planck epoch. The origin and properties of this initial density perturbations are yet to be understood. Amongst other theories, inflationary models provide promising explanations for the origin and statistical properties of this seed perturbations while also resolving some other cosmological problems, such as the flatness and horizon problem (Guth 1981, Albrecht & Steinhardt 1982). In addition, one

needs to understand the processes governing the growth and time evolution of density fluctuations in the Universe. This study of cosmological structure formation generally requires analytic understanding by performing perturbation theoretical analyses of the governing equations as well as numerical simulations to follow the dynamics of a gravitating system in an expanding universe.

3.2.1. Growth of density perturbations in cold dark matter models

3.2.1.1. Dark matter

Today, it is widely known that ordinary baryonic matter alone cannot explain the formation of the currently observed structures in the Universe via gravitational interaction. In the process of gravitational clustering, matter flows away from regions where the density is below average and aggregates in places with higher densities. In general, this gravitational clustering is a slow process. Thus, in order to explain the presently observed cosmic structures by a purely baryonic matter component, the fluctuations in the cosmic microwave backround would be expected to be at least two orders of magnitude larger in amplitude than actually measured by modern cosmic microwave background experiments (Einasto 2009). Cosmologists resolved this mass discrepancy by introducing a hypothetical material component of the Universe, which does not interact via electromagnetic or strong forces, but possibly by weak nuclear interactions. Observational indications for such a "dark matter" component have been found already very early. Zwicky (1933) studied the radial velocities of galaxies in the Coma cluster, and found that in order to explain the observed orbital velocities the cluster must contain huge amounts of invisible matter. Further dynamical evidence for dark matter was obtained in the 1970's by measuring rotation curves of galaxies (Rubin & Ford 1970). A peculiarity in this measurements was that rotation curves do not exhibit the expected Keplerian fall but tend to be flat at radii larger than a few kpc from the center. These observations, are consistent with extended massive haloes, containing about ten times more mass than the galactic mass observed optically (Del Popolo 2007). However, as noted in the beginning most stringent evidence for dark matter comes from measurements of the cosmic microwave background temperature fluctuations, such as carried out by the COBE or the WMAP satellites. Also weak lensing or discrepancies in the estimated mass of clusters of galaxies via the virial theorem give strong indications for the existence of a dark matter component. A very impressive evidence for dark matter was provided by the Chandra X-ray observatory which observed the bullet cluster of galaxies and a joint analysis of its lensing properties. This analysis revealed that the gravitational centers do not coincide with the visible luminous matter.

Although much evidence in favor of dark matter has been provided, so far the properties of this hypothetical form of matter are unknown. It is believed to be composed of yet undiscovered gravitationally interacting elementary particles, which carry neither an electromagnetic nor a strong charge. However, they possibly interact by the weak nuclear force. These particles are expected to be stable and only weakly self-interacting in order to account for a significant contribution to the critical density. Further constraints for a small self-interaction cross section can be deduced from the impact of dark matter on the central cores of dark matter halos (Yoshida et al. 2000). The detection and identification of the dark matter particle is a major scientific task in modern cosmology and particle physics. Currently there exist a variety of experiments which try to detect the dark matter particle either directly or indirectly. There exist a huge variety of direct detection experiments, which search for scattering events of weakly interacting massive particles with atomic nuclei. These experiments are usually carried out in very deep underground laboratories to eliminate the background generated by cosmic rays (CRESST [2], CDMS[3], EURECA [4], XENON[5], DAMA[6], LUX[7], EDELWEISS[8]). Also the Large Hadron Collider might be able to provide a direct measurement of the dark matter particle (LHC [9]). Dark matter particles can also be indirectly detected through their annihilation radiation or other products of dark matter interactions. In particular, the Fermi Gamma-ray Space telescope searches for dark matter annihilation signals (Atwood et al. 2009).

---

[2]http://www.cresst.de/

[3]http://cdms.berkeley.edu/

[4]http://www.eureca.ox.ac.uk/

[5]http://xenon.astro.columbia.edu/

[6]http://www.lngs.infn.it/lngs/htexts/dama/welcome.html

[7]http://lux.brown.edu/

[8]http://edelweiss.in2p3.fr/indexedwe.html

[9]http://lhc.web.cern.ch/lhc/

3.2.1.2. Vlasov equation

As described above, the detailed nature of dark matter is still unknown. However, all dark matter candidates are extremely light compared to the mass scale of typical galaxies. The number densities of these dark matter particles are therefore expected to be very high, at least on the order of 10^{50} particles per Mpc^3 (Kolb & Turner 1990). As in this limit, $N \gg 1$, discreteness effects are negligible, dark matter is thought to obey the Vlasov equation for the phase space distribution function (Bernardeau et al. 2002). It is also known, that at scales much smaller than the Hubble radius $d \ll c/H_0$ the equations of motions can be essentially described by Newtonian dynamics. For a detailed discussion of the Newtonian limit derived from general relativity see e.g. Peebles (1980). If we define the number density in phase space as $f(r, p, t)$ then phase space conservation implies the Vlasov equation (Bernardeau et al. 2002). In the Newtonian limit the Vlasov equation for collision-less dark matter thus is given as:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \frac{p}{m}\frac{\partial f}{\partial r} - m\nabla\Phi\frac{\partial f}{\partial p} = 0\,, \tag{3.21}$$

with $p = m\,\mathrm{d}r/\mathrm{d}t$ being the momentum, m the particle mass, and Φ is the Newtonian potential given by the Poisson equation:

$$\nabla^2\Phi(r, t) = 4\pi G m \int \mathrm{d}^3p\, f(r, p, t)\,. \tag{3.22}$$

This is a highly non-linear partial differential equation, involving seven dimensions. Here, the non-linearities arise from the Newtonian potential Φ, which depends through the Poisson equation on the integral of the distribution over momentum. There exist no solution to the Vlasov equation for collision-less dark matter. Therefore, in order to study the behavior of dark-matter in the Universe, two main approaches are persued. Numerical techniques, such as N-body simulations, sample the phase space distribution by a large number of discrete particles and follow their trajectories in phase space. In this fashion one obtains a sampled representation of the phase space density $f(r, p, t)$. The analytic approach, on the other hand, relies on studying the time evolution of the zeroth and first momentum moments of the distribution. This yields a fluid dynamics approach for the motion of collision-less dark matter.

3.2.1.3. Fluid approach

The complicated non-linear structure of the Vlasov equation forbids simple analytic analysis of the dark matter dynamics. For this reason one relies on approximations to Vlasov's equation, by taking the zeroth and first momentum moments of the phase space distribution. The zeroth order moment then relates the phase space density to the local mass density field:

$$\int \mathrm{d}^3p\, m\, f(r, p, t) =: \rho(r, t) \tag{3.23}$$

and the first momentum moment yields the peculiar velocity flow $v(r, t)$ as:

$$\int \mathrm{d}^3p\, p\, f(r, p, t) =: \rho(r, t)\, v(r, t)\,. \tag{3.24}$$

The next order moment yields the stress tensor, which characterizes the deviation of particle motions from a single coherent flow (Bernardeau et al. 2002). Neglecting the stress tensor, will therefore only be a good approximation for the early stages of structure formation before velocity dispersions due to multiple streams are generated. Hence, this approximation will eventually breakdown in regions where non-linear structure formation takes place. It is possible to incorporate the stress tensor in the analysis for the expense of more analytic complexity (see e.g. Buchert 2000, Bernardeau et al. 2002, Buchert & Domínguez 2005, Pueblas & Scoccimarro 2009). In the following we will persue the common approach of ignoring the stress tensor to describe the early and mildly non-linear stages of structure formation.

For a collisionless self gravitating cold dark matter fluid the continuity and Euler equations can then be derived by taking moments of the Vlasov equation (Bernardeau et al. 2002). The continuity and Euler equations are then given as:

$$\frac{\partial\rho(r, t)}{\partial t} + \nabla\left[\rho(r, t)v(r, t)\right] = 0\,, \tag{3.25}$$

and:

$$\frac{\partial v(r,t)}{\partial t} + (v(r,t)\nabla)v(r,t) + \nabla\Phi(r,t) = 0 \,. \tag{3.26}$$

The Newtonian potential $\Phi(r,t)$ is then related to the mass density $\rho(r,t)$ via the Poisson equation:

$$\nabla^2\Phi(r,t) = 4\pi G\rho(r,t) \,. \tag{3.27}$$

Unfortunately, there exists no general analytic solution to the fluid dynamics of collisionless self gravitating cold dark matter. Literature, however, provides a plenitude of different analytic perturbative techniques to yield approximate solutions for the dark matter dynamics (see e.g. Zel'Dovich 1970, Buchert 2000, Bernardeau et al. 2002, Short & Coles 2006, Crocce & Scoccimarro 2006). Below we will provide some common approximations to the fluid dynamics approach.

3.2.1.4. Linear growth

To study the linear growth of structure in the Universe the above hydrodynamic equations can be approximated to leading order by following the evolution of small perturbations against the expanding background. The density and velocity fields can then be expanded as:

$$\rho(r,t) = \rho_0(t) + \delta\rho(r,t) \,, \tag{3.28}$$

$$v(r,t) = v_0(t) + \delta v(r,t) \,, \tag{3.29}$$

and

$$\Phi(r,t) = \Phi_0(t) + \delta\Phi(r,t) \,, \tag{3.30}$$

where $\rho_0(t)$ is the homogeneous background density, $v_0(t)$ is the Hubble expansion, and $\Phi_0(t)$ is the background gravitational potential. Perturbations in the dark matter density field can be more conveniently be described by the density contrast $\delta(r,t)$:

$$\delta(r,t) = \frac{\delta\rho(r,t)}{\rho_0(t)} \,, \tag{3.31}$$

with the average cosmic density $\rho_0(t) = \Omega_m\rho_{\mathrm{crit}}a(t)^{-3}$. When co-moving length units are used the equations become particularly simple, since then the background density $\rho_0(t)$ is independent of time. For this reason we introduce co-moving spatial coordinates and a similar transformation for the peculiar velocities:

$$r(t) = a(t)x(t) \,, \tag{3.32}$$

and

$$\delta v(t) = a(t)u(t) \,. \tag{3.33}$$

The linearized equations are then given as:

$$\frac{\mathrm{d}\delta(x,t)}{\mathrm{d}t} = -\nabla_x u(x,t) \tag{3.34}$$

and

$$\frac{\mathrm{d}u(x,t)}{\mathrm{d}t} + 2\frac{\dot{a}}{a}u(x,t) = 4\pi G\rho_0\,\delta(x,t) \tag{3.35}$$

A second order differential equation for the density contrast $\delta(x,t)$ can be obtained by eliminating u:

$$\frac{\mathrm{d}^2\delta(x,t)}{\mathrm{d}t^2} + 2\frac{\dot{a}}{a}\frac{\mathrm{d}\delta(x,t)}{\mathrm{d}t} = 4\pi G\rho_0\,\delta(x,t) \,. \tag{3.36}$$

This equation governs the gravitational amplification of linear density perturbations. A thorough relativistic perturbative approach, shows that in the linear regime $|\delta| \ll 1$ perturbations grow differently with a, depending on which fluid dominates the cosmological dynamics, as long as the Einstein-de Sitter limit is fulfilled, i.e. $\Omega_m(a) \simeq 1$:

$$\delta(a) \propto a^{3w+1} \,, \quad \text{with } w = \begin{cases} \frac{1}{3} & \text{for } a < a_{\mathrm{eq}}, \text{ radiation dominated era,} \\ 0 & \text{for } a > a_{\mathrm{eq}}, \text{ matter dominated era.} \end{cases} \tag{3.37}$$
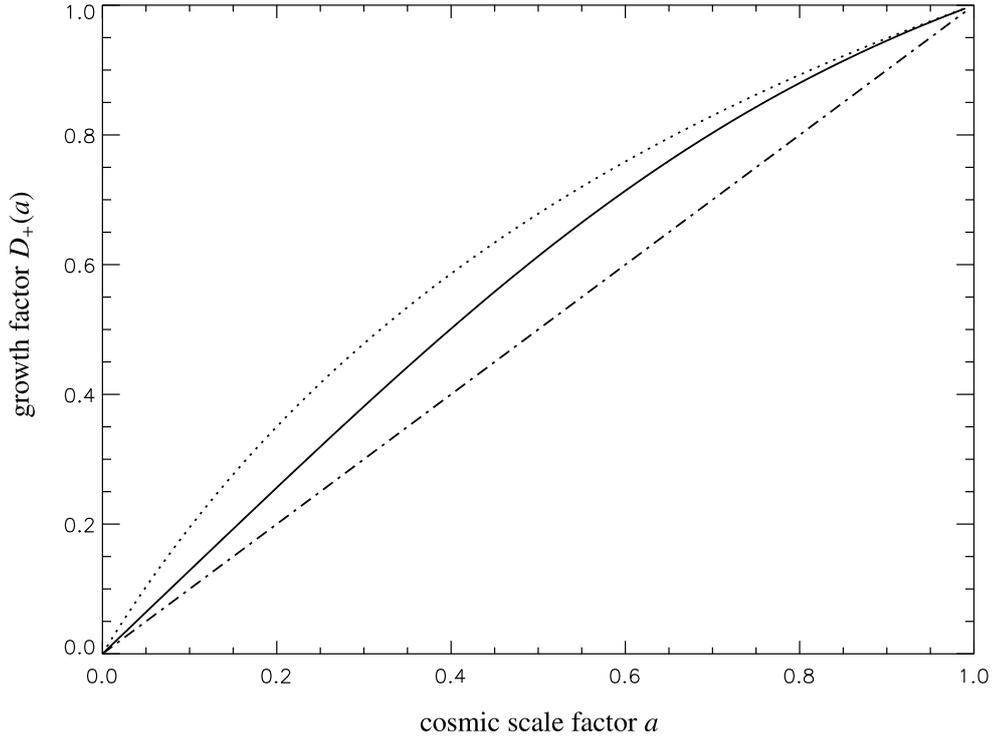
Figure 3.2.: The plot shows the growth function $D_+(a)$ for three different cosmologies: A low-density model with $\Omega_m = 0.3$ and vanishing cosmological constant $\Omega_\Lambda = 0.0$ (dotted line), a ΛCDM model (solid line) and a SCDM model (dash-dotted line).

Once either the matter density Ω_m has decreased sufficiently or the cosmological Ω_Λ has started dominating the Hubble expansion, as is the case for later times, the linear growth depends on the cosmic scale factor a according to:

$$\frac{\delta(\boldsymbol{x}, a)}{\delta(\boldsymbol{x}, 1)} = a \frac{g'(a)}{g'(1)} \equiv D_+(a), \tag{3.38}$$

where $D_+(a)$ is the solution to the growing mode of homogeneous structure formation. In general, the solution of equation (3.36) requires numerical integration. However, a good approximation to $g'(a)$ for the Ω_m-dominated phase of structure growth is provided by Carroll et al. (1992):

$$g'(a) = \frac{5}{2}\Omega_m(a)\left[\Omega_m^{4/7}(a) - \Omega_\Lambda(a) + \left(1 + \frac{1}{2}\Omega_m(a)\right)\left(1 + \frac{1}{70}\Omega_\Lambda(a)\right)\right]^{-1}. \tag{3.39}$$

The growth function $D_+(a)$ as a function of scale factor a of the ΛCDM model, the SCDM model and a low density model without cosmological constant Λ is depicted in Fig. 3.2.

3.2.1.5. Velocities in the large-scale structure

In the previous section we focused on the evolution of the density contrast. However, for the analysis of large scale structure also the peculiar velocity field is of great importance. The linear velocity field can be easily calculated from equation (3.35) in Fourier space. For a harmonic perturbation with wave vector $\boldsymbol{k}$, the peculiar velocity $\boldsymbol{u}$ is parallel to $\boldsymbol{k}$:

$$\boldsymbol{u}(\boldsymbol{k}, t) = -\mathrm{i}\frac{\boldsymbol{k}}{k^2}\frac{\mathrm{d}\delta(\boldsymbol{k}, t)}{\mathrm{d}t} \tag{3.40}$$

As described above, in the linear regime the time evolution of the density field δ is homogeneous and can therefore be expressed as $\delta(\boldsymbol{k}, a) = D_+(a)\delta(\boldsymbol{k}, 1)$. With the definition of the Hubble function $\dot{a} = aH(a)$ and the normalization of the growth function we hence yield:

$$\boldsymbol{u}(\boldsymbol{k}, a) = -\mathrm{i}H(a)f(\Omega)\frac{\boldsymbol{k}}{k^2}\delta(\boldsymbol{k}, a)\,, \tag{3.41}$$

where f describes the dependence of the equation of continuity on cosmic time and mainly depends on the mass density Ω_m (Peebles 1980, Lahav et al. 1991):

$$f(\Omega) = \frac{\mathrm{d}\ln\delta}{\mathrm{d}\ln a} = \frac{\mathrm{d}\ln D(a)}{\mathrm{d}\ln a} \simeq \Omega_m(a)^{0.6} \tag{3.42}$$

The proper physical peculiar velocity δv is then obtained by multiplication with the cosmic scalefactor a:

$$\delta\boldsymbol{v}(\boldsymbol{k}, a) = -\mathrm{i}\,a\,H(a)f(\Omega)\frac{\boldsymbol{k}}{k^2}\delta(\boldsymbol{k}, a)\,. \tag{3.43}$$

Investigation of peculiar velocity fields in our cosmic neighborhood is an interesting topic, since they can be used to test whether cosmic flows are irrotational and also to provide dynamical estimates of Ω_m and σ_8. For this reason, reconstructions of the cosmic velocity field have been carried out e.g. with the POTENT algorithm proposed by Bertschinger & Dekel (1989, 1991).

3.2.2. Zel'dovich approximation

In the sections above, the fluid motions were described in an Eulerian coordinate system. However, it is possible to develop non-linear perturbation theory in the so-called Lagrangian framework. In this approach one follows the trajectories of individual particles or fluid elements rather than studying the dynamics of density and velocity fields (for a review of Lagrangian perturbation theory see Buchert 1996, Bernardeau et al. 2002). To linear order this perturbation approach is equivalent to the Zel'dovich approximation, as proposed by Zel'Dovich (1970). Here, the objects of interest are the particle trajectories, which link the initial particle positions $\boldsymbol{q}$ to their positions $\boldsymbol{x}(\boldsymbol{q}, a)$ at a later time:

$$\boldsymbol{x}(\boldsymbol{q}, a) = \boldsymbol{q} - D_+(a)\nabla\Psi(\boldsymbol{q})\,, \tag{3.44}$$

where $\Psi(\boldsymbol{q})$ is the displacement potential given by the initial density contrast through a Poisson equation:

$$\nabla^2\Psi(\boldsymbol{q}) = \delta(\boldsymbol{q})\,. \tag{3.45}$$

Equation 3.44, therefore, describes the particle dynamics as straight inertial motion in the direction of its initial velocity vector. Strictly speaking, extrapolation of peculiar velocities is only exact in one dimension. One dimensional gravitational dynamics corresponds to following parallel sheets of matter, where the gravitational acceleration towards a sheet is independent of distance. Hence, the full equation of motion can be extrapolated from the particle's initial peculiar velocity, as long as sheets do not cross (Peacock 1999). The dynamics described by equation (3.44) do not take into account gravitational interactions between individual particles. Thus, Zel'dovich's approximation fails at sufficiently non-linear stages when particles are to be forming gravitationally bound objects instead of following straight lines. Nevertheless, Zel'dovich's approximation provides interesting insight into the formation of the large scale cosmic web. Figure 3.3 displays solutions of equation 3.44 for a set of 256^3 particles. The initial density contrast $\delta(\boldsymbol{q})$ was generated from a multivariate normal distribution on a 256^3 equidistant grid and according to a ΛCDM power-spectrum calculated according to the prescription presented in Eisenstein & Hu (1998) and Eisenstein & Hu (1999). Further, the set of cosmological parameters ($\Omega_m = 0.24$, $\Omega_\Lambda = 0.76$, $\Omega_b = 0.04$, $h = 0.73$, $\sigma_8 = 0.74$, $n_s = 1$) was adopted. The plots demonstrate the filamentary cosmic web at four different resolutions.

To obtain the Eulerian density field at any epoch one exploits mass conservation $[1 + \delta(\boldsymbol{q})]\mathrm{d}q^3 = [1 + \delta(\boldsymbol{x})]\mathrm{d}x^3$ and uses the Jacobian of the transformation between $\boldsymbol{q}$ and $\boldsymbol{x}$:

$$(1 + \delta(\boldsymbol{x})) = (1 + \delta(\boldsymbol{q}))\left(\det\left[\delta_{ij}^K - D_+\partial_i\partial_j\Psi\right]\right)^{-1}\,. \tag{3.46}$$

This result demonstrates that the tidal forces $\partial_i \partial_j \Psi$ cause compressions of the cosmic density field (Schäfer 2009). When using the three eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ of the Jacobian in equation (3.46) one can write:

$$(1 + \delta(\boldsymbol{x})) = \frac{(1 + \delta(\boldsymbol{q}))}{(1 - D_+\lambda_1)(1 - D_+\lambda_2)(1 - D_+\lambda_3)} \, . \tag{3.47}$$

The details of structure growth in a particular region depends on the values $(\lambda_1, \lambda_2, \lambda_3)$ with $(\lambda_1 \geq \lambda_2 \geq \lambda_3)$. One can see that if $\lambda_1 >> \lambda_2, \lambda_3$ then the gravitational collapse proceeds along one axis and forms a two dimensional pancake. In regions where $\lambda_1 \sim \lambda_2 >> \lambda_3$ the gravitational collapse is expected to be two dimensional and a one dimensional filament will form. For $\lambda_1 \sim \lambda_2 \sim \lambda_3$ the gravitational collapse proceeds along all three axis and will form a nearly spherical overdensity or halo. This behavior of gravitational structure formation lends itself to a cosmic web classification algorithm, which classifies different regions either as voids, sheets, filaments or halos. Such a dynamic web classification algorithm was proposed by Hahn et al. (2007) and refined by Forero-Romero et al. (2009). In chapter 9 an application of such a web classification method to observed data will be presented.

3.2.3. Numerical simulations of cosmic structure formation

The results presented above are only valid in the linear regime ($|\delta| \ll 1$). However, in the course of structure formation high density objects with overdensities $|\delta| \gg 1$ are formed, e.g. galaxies ($\delta \simeq 10^6$), clusters of galaxies ($\delta \simeq 100$) and superclusters ($\delta \simeq 10$). In these regimes, perturbation theory provides no valid description for the dynamics of these objects. In addition, non-linear structure formation generally proceeds heterogeneously, such, that the relation $\delta(\boldsymbol{x}, a) = D_+(a)\delta(\boldsymbol{x})$ is violated. The non-linear dynamics yield strong coupling between different modes $\delta(\boldsymbol{k})$ in Fourier space, which does not permit a simple analytic approach. And thirdly, non-linear structure formation produces non-Gaussian features, due to phase correlations. This is due to the equation of continuity which prevents the density contrast to assume values $\delta < -1$. The density fluctuation field therefore is bounded to small values $\delta > -1$, but an upper bound does not exist. Thus, in the course of non-linear structure formation the distribution of δ necessarily develops a non-vanishing skewness. For this reason, one cannot uniquely describe the statistical properties of the non-linear density field in terms of 2-point correlation functions or power spectra $P_\delta(k)$.

   In order to study the non-linear stages of cosmic structure formations, one usually relies on numerical simulation. These computer codes, the most notable of which is GADGET (Springel et al. 2001, Springel & Hernquist 2002), approximately solve the Vlasov equation (3.21), by following the trajectories of a set of discrete particles through phase space. In this fashion, one obtains a sampled representation of the phase space distribution $f(\boldsymbol{r}, \boldsymbol{p}, t)$. Extensions to GADGET include baryonic dynamics, magnetic fields and cosmic rays. In this work some artificial galaxy catalogs based on simulations, carried out with GADGET, are used to test the presented large scale structure analysis methods.

3.3. Large scale structure

3.3.1. Distribution of matter in the Universe

In the sections above we described the governing equations for the dynamics of the large scale structure. However, all these models are incomplete in a crucial aspect, since they do not explain the origin of the initial seed perturbations. These microscopic primordial density fluctuations yield the structure as observed in the present universe by gravitational amplification. A possible origin for these seed fluctuations may be topological defects that remain as the relics of high energy phase transitions in the early universe (Kibble 1976). However, currently it is widely accepted that the origin of the primordial density fluctuations can be well explained by the theory of inflation (Guth & Pi 1982, Starobinsky 1982, Hawking 1982, Bardeen et al. 1983, Brandenberger et al. 1983, Brandenberger & Kahn 1984). According to the inflationary paradigm, the equation of state of the early universe is governed by a quantum field with negative pressure, the so-called inflaton field. This quantum field drives the exponential growth of the cosmic scale factor and causes quantum fluctuations to be magnified to macroscopic scales, until they leave the horizon. These fluctuations spend a long time outside the horizon, before re-entering at later stages during the era of radiation- and matter-domination, where they set the initial conditions for gravitational structure formation. The theory of inflation is generally favored over other theories for the origin of seed perturbations, since it also provides explanations for some other characteristics of the observed universe, e.g. flatness, homogeneity and isotropy
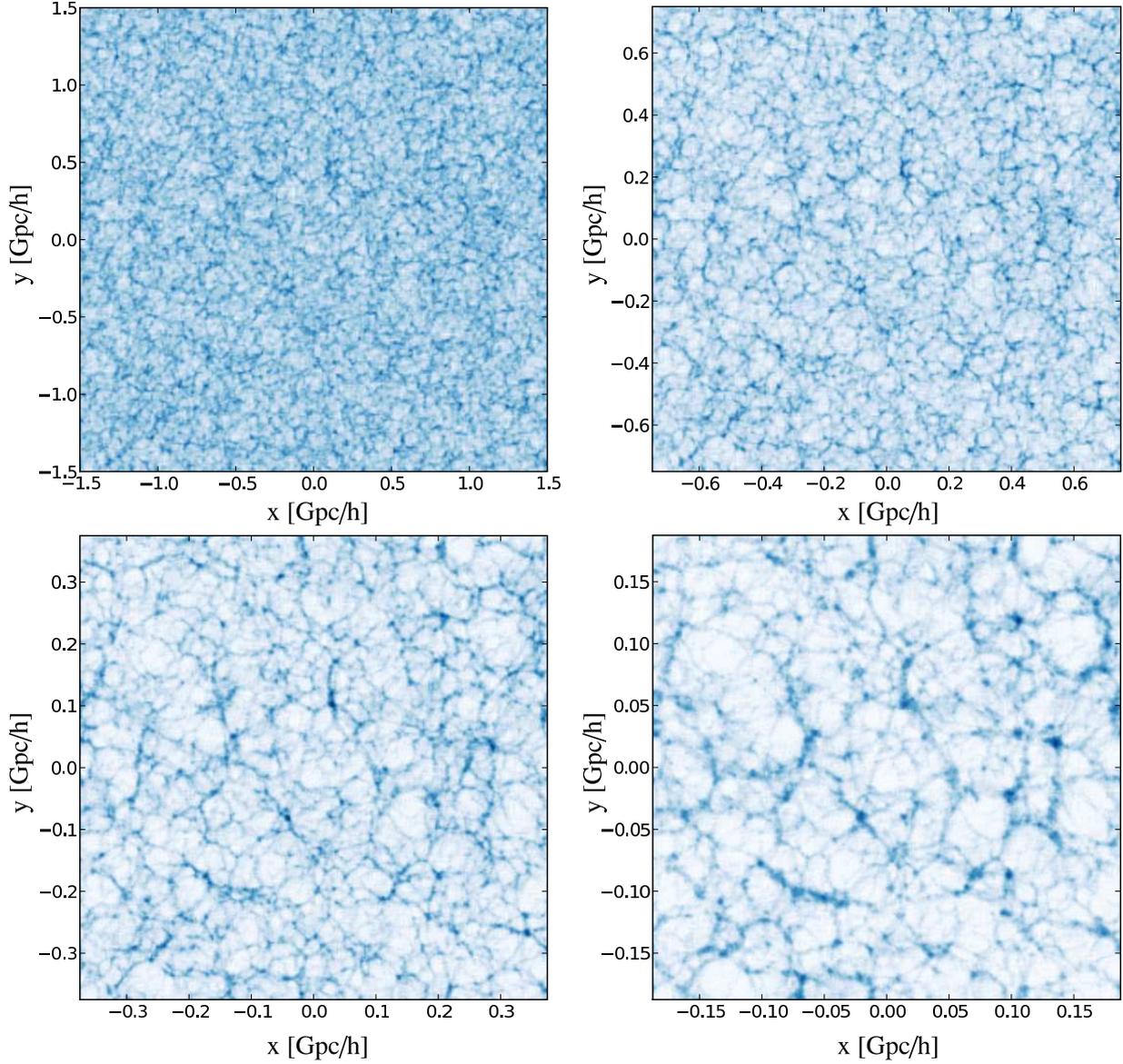
Figure 3.3.: Solution of the Zel'dovich approximation, given by equation (3.44) for a ΛCDM universe. The four panels correspond to different box sizes 1.5 Gpc (upper left panel), 0.75 Gpc (upper right panel), 0.375 Gpc (lower left panel) and 0.1875 Gpc (lower left panel). The filamentary structure of the cosmic web is clearly visible.

(Guth 1981, Linde 1982, Albrecht & Steinhardt 1982).

Many inflationary models predict the amplitudes of the primordial seed perturbations to be Gaussian distributed. Here, the Gaussianity is a consequence of the superposition of a very high number of uncorrelated quantum fluctuations, which according to the central limit theorem yield a Gaussian amplitude distribution (Kendall & Stuart 1969). Thus, the process of generating the seed perturbations is stochastic and cannot be derived from first principles. Therefore, a probability distribution function of the cosmological density fluctuations is the most fundamental characterization of the large scale structure in the Universe. The statistical properties of the large scale matter distribution can be inferred from observations of anisotropies in the cosmic microwave background and from probes of the large scale structure such as galaxy redshift surveys. Since the fluctuations are believed to arise from inflation, such measurements can also set constraints on parameters within inflationary theory.

3.3.2. Statistical description of the large scale structure: Gaussian random fields

As described above, inflationary models predict the initial density fluctuations to be Gaussian distributed. If this is the case, then the statistical behavior of the initial seed perturbations is completely determined by its two point statistics or accordingly the matter power-spectrum. Considering the density contrast $\delta_i \equiv \delta(x_i)$ defined at the comoving position x_i to be a stochastic variable we can define the joint probability distribution of the initial density field as a multivariate Gaussian:

$$\mathcal{P}(\{\delta_i\} : i = 1, ..., N) = \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2} \sum_{i,j=1}^{N} \delta_i S_{ij}^{-1} \delta_j}, \tag{3.48}$$

with N being an arbitrary positive integer and $S_{ij} = \langle \delta_i \delta_j \rangle$ is the according covariance matrix (see e.g. Lahav & Suto 2004). As can be seen from equation (3.48), the Gaussian distribution is completely specified by the two point correlation function $S_{ij} = \xi(x_i, x_j)$. However, although the two point statistics are of undoubtable importance, they do not provide any information on the phases of the initial density fluctuations. Hence, there exists no phase correlations in a Gaussian random field. Also note, that the Gaussian assumption has to break down at later epochs of structure formation since it predicts density amplitudes to be symmetrically distributed in the range $-\infty < \delta_i < \infty$ but weak and strong energy conditions require $\delta_i \geq -1$. The Gaussian assumption is therefore strictly speaking only valid in the limit of infinitesimal small density fluctuations $|\delta_i| << 1$. Equation (3.48) is most conveniently written in Fourier space, where for an homogeneous and isotropic universe the Fourier space covariance reduces to a diagonal form:

$$\langle \delta(k)\delta^*(k') \rangle = (2\pi)^3 \delta^D(k - k') P_\delta(k), \tag{3.49}$$

with $\delta(k) = \int d^3x\, \delta(x) \exp(-ikx)$ being the Fourier transform of the density contrast and $P_\delta(k)$ is the according density power-spectrum. Due to the diagonal structure of the Fourier space covariance matrix, in a homogeneous, isotropic universe different Fourier modes are uncorrelated and their probability distribution is given by:

$$\mathcal{P}(\delta(k)) = \frac{1}{\sqrt{2\pi P_\delta(k)}} e^{-\frac{1}{2} \frac{|\delta(k)|^2}{P_\delta(k)}}. \tag{3.50}$$

Therefore, if the initial density fluctuations in a homogeneous and isotropic universe were Gaussian distributed then the matter power-spectrum, completely characterizes their statistical properties. Also note, that the assumption of homogeneity and isotropy of the Universe can, in principle, be tested by measuring the off-diagonal elements of $\langle \delta(k)\delta^*(k') \rangle$. As already discussed above, the Gaussian assumption only applies to density amplitudes $|\delta_i| << 1$. However, gravitational amplification will dramatically increase the amplitudes of the density field and will also introduce mode coupling as well as phase correlations. Hence, the evolved density field in the non-linear regimes will not be isotropic and homogeneous, due to mode coupling, and will also not be Gaussian distributed, due to phase correlations. The statistical properties of the evolved density field will be discussed below.

3.3.2.1. The matter power-spectrum $P_\delta(k)$ and its shape

If the initial density fluctuations in a homogeneous and isotropic universe were Gaussian or close to Gaussian, then the detailed functional form of the matter power-spectrum $P_\delta(k)$ is of obvious relevance to cosmology. The primordial shape of $P_\delta(k)$ can be estimated with the aid of linear perturbation theory as described in 3.2.1.4. In linear perturbation theory gravity acts homogeneously on the density field causing each individual Fourier mode to

evolve independently:

$$\delta(\boldsymbol{x}, a) = D_+(a)\delta(\boldsymbol{x}) \longrightarrow \delta(\boldsymbol{k}, a) = D_+(a)\delta(\boldsymbol{k}), \tag{3.51}$$

as long as its wavelength is small compared to the horizon size $d_{\mathrm{H}} = c/(aH(a))$ where the Newtonian treatment is applicable. Here d_{H} is defined as the distance a photon could have traveled since the Big Bang. A Fourier mode $\delta(\boldsymbol{k})$ crosses the horizon at the epoch a_{start} if its wavelength $\lambda = 2\pi/k$ is equal to the $\lambda = d_{\mathrm{H}}(a_{\mathrm{start}})$ horizon size at that time.

However, at early times, during the era of radiation domination, structure growth on small scales is greatly suppressed due to the radiation driven expansion of the Universe. This can be seen by comparing the expansion time scale t_{Hubble} to the collapse time scale t_{dm} of dark matter:

$$t_{\mathrm{Hubble}} \propto \frac{1}{\sqrt{G\rho_r}} \leq \frac{1}{\sqrt{G\rho_m}} \propto t_{\mathrm{dm}}, \tag{3.52}$$

with $\rho_r > \rho_m$ at the epoch of radiation domination. This radiation driven expansion effectively suppresses structure growth from a_{start} until the epoch of matter radiation a_{eq}. Therefore, structures first start growing at a redshift $z \simeq 24500$ when $\rho_r(a_{\mathrm{eq}}) = \rho_m(a_{\mathrm{eq}})$. Then, according to equation (3.37), all fluctuations with $\lambda < d_{\mathrm{H}}(a_{\mathrm{eq}})$ are suppressed by $(a_{\mathrm{start}}/a_{\mathrm{eq}})^2$ and do not grow before matter radiation equality. The epochs a_{start} at which the different modes enter the horizon can now be expressed as a function of wavelength by $\lambda = d_{\mathrm{H}}(a_{\mathrm{start}})$. In the early Einstein-de Sitter phase, $d_{\mathrm{H}}(a)$ can be approximated as being proportional to the scale factor by a, yielding a suppression proportional to λ^2 on scales smaller than the horizon size at the epoch of matter-radiation equality. The transition from the mode suppression- to the mode growing region therefore happens at a wavelength corresponding to the horizon size at matter radiation equality, the numerical value of which is $0.025/(\Omega_m h)$ Hubble radii.

To estimate the shape of the primordial power-spectrum one commonly assumes the initial power-spectrum to be scale invariant on large scales, $P_\delta(k) \propto k^{n_s}$ with $n_s \simeq 1$ (Harrison 1970, Peebles & Yu 1970, Zeldovich 1972). With this assumption and the suppression of Fourier modes $\propto \lambda^2 = (2\pi/k)^2 \propto k^{-2}$, the asymptotic behavior of the power-spectrum on small scales can be estimated to be $P_\delta(k) \propto k^{n_s-4} = k^{-3}$. As already pointed out before, the transition point between the large scale and small scale regime is related to a horizon size of roughly $0.025/(\Omega_m h)$ Hubble radii. From this simple arguments it is clear that this point of turn over in the cosmic power-spectrum carries some information about the dark matter content of the Universe.

Nevertheless, much more accurate fitting formulas, which link these two asymptotic regimes in a smooth way have been obtained by large-volume n-body simulations. Especially the fitting function provided by Bardeen et al. (1986) has been proven to be very accurate:

$$P_\delta(k) \propto k^{n_{\mathrm{s}}} \cdot T^2(k), \tag{3.53}$$

where the transfer function is given by

$$T(q) = \frac{\ln(1 + 2.34q)}{2.34q} \left[1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4 \right]^{-\frac{1}{4}}. \tag{3.54}$$

Here, q is the wave vector k divided by the shape parameter Γ, as introduced by Efstathiou et al. (1992) for CDM models and extended to models with $\Omega \neq 1$ by Sugiyama (1995):

$$q = \frac{k/\mathrm{Mpc}^{-1}h}{\Gamma} \quad \text{with } \Gamma = \Omega_m h \exp\left(-\Omega_b \cdot \left[1 + \frac{\sqrt{2h}}{\Omega_m}\right]\right). \tag{3.55}$$

The power-spectrum $P_\delta(k)$ is normalized such that the variance of the density fluctuations δ on scales of $R = 8$ Mpc is given by the parameter σ_8. The variance of the density field on an arbitrary scale R is usually measured by applying a top-hat shaped filter function to the density field. This yields:

$$\sigma_R^2 = \frac{1}{2\pi^2} \int_0^\infty \mathrm{d}k \, k^2 \hat{W}^2(kR) P(k), \tag{3.56}$$
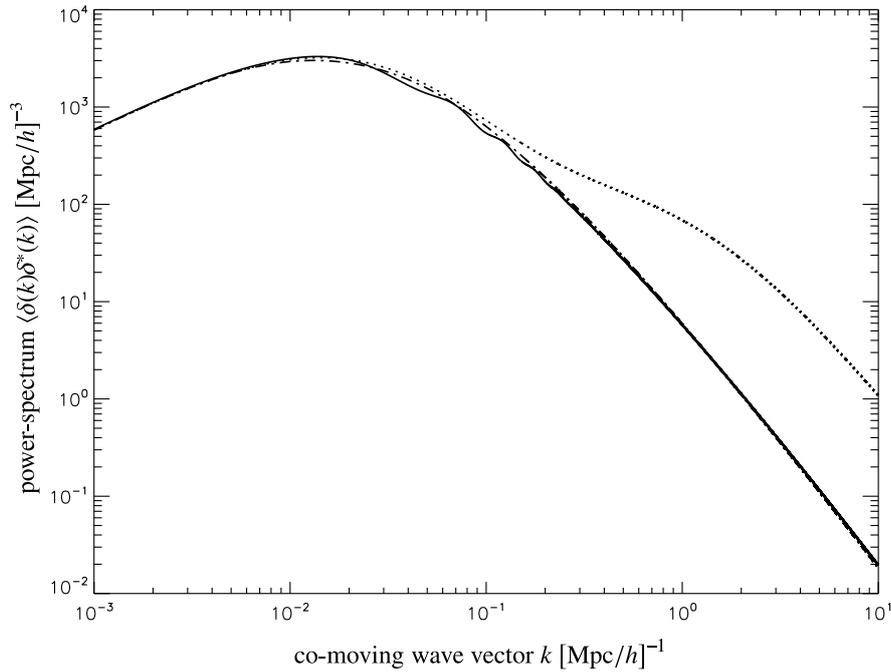
Figure 3.4.: Three different power-spectra $P_\delta(k)$ of the overdensity $\delta(\boldsymbol{x})$: The power-spectrum for an adiabatic cold dark matter model calculated according to eq. (3.54) (dash-dotted line), a linear power-spectrum for a universe with baryonic matter calculated according to the fitting formula provided by Eisenstein & Hu (1998) and Eisenstein & Hu (1999) (solid line) and the nonlinear fitting function of Smith et al. (2003) (dotted line).

with $W(x)$ being the top-hat filter. Its Fourier-transform is given by:

$$\hat{W}(y) = \frac{3}{y^3} \left[\sin(y) - y\cos(y) \right] = \frac{3}{y} J_1(y). \tag{3.57}$$

In Fig. 3.4 three different power-spectra $P_\delta(k)$ of the overdensity field $\delta(\boldsymbol{x})$ are plotted, for the case of a pure dark matter ΛCDM model, a ΛCDM model with baryonic matter, and the fully evolved nonlinear power-spectrum. Here, the set of parameters ($\sigma_8 = 0.9, n_s = 1$) was assumed.

3.3.3. Statistical description of the large scale structure: Lognormal random fields

The discussion in the previous section demonstrated that if the large scale density field in a homogeneous and isotropic universe were Gaussian distributed its statistical properties will be completely determined by the according matter power-spectrum. However, as soon as the density amplitudes approach values comparable to unity linear perturbation theory breaks down and the full gravitational dynamics have to be considered for the subsequent evolution of structures in the Universe. Unlike in the case of homogeneous gravitational interaction, non-linear gravitational effects will introduce mode coupling and phase correlations to the evolved density field. These effects have pronounced consequences for the statistical behavior of the density distribution. The first effect of mode coupling generally introduces off-diagonal elements in the according density covariance matrix and thus the matter power-spectrum, as defined in equation 3.49, alone is not sufficient to fully describe the statistical properties of the density field. Yet, by far more important is the second effect, which introduces phase correlations. Since a Gaussian random field is not able to represent phase correlations it is obvious that the evolved density field cannot obey Gaussian statistics. Also, gravitational interactions will yield very high positive density contrast amplitudes. In order to prevent $\delta(\boldsymbol{x}) < -1$ and at the same time preserving $\langle \delta(\boldsymbol{x}) \rangle = 0$ the resultant probability distribution must be strongly skewed (Peacock 1999). In principle, the probability distribution $\mathcal{P}(\delta(\boldsymbol{x})|t)$ at time t provides a complete characterization of the statistical properties of the cosmic mass distribution. However, the exact form of this distribution

can only be determined if the according dynamical equations are solved. In the absence of such a solution, one can only describe the statistical properties of the fully evolved non-linear density field by phenomenological probability distributions. A variety of such phenomenological models have been proposed in literature (see e.g. Hubble 1934, Saslaw 1985, Coles & Jones 1991, Gaztanaga & Yokoyama 1993). A particularly interesting model which meets many of the discussed prerequisites, is the lognormal probability distribution, which is also Gaussian in the limit of small density amplitudes. It has also been demonstrated to accurately represent the statistical behavior of the density fields calculated by n-body simulations from Gaussian initial conditions (Kayo et al. 2001). One therefore believes that the statistical properties of the evolved density field can be represented up to two point statistics by a multivariate lognormal distribution:

$$\mathcal{P}\left(\{\delta_i\} : i = 1, ..., N\right) = \frac{1}{\sqrt{2\pi det(Q)}} e^{-\frac{1}{2}\sum_{ij}(ln(1+\delta_i)+\mu_i)Q_{ij}^{-1}(ln(1+\delta_j)+\mu_j)} \prod_k \frac{1}{1+\delta_k}, \tag{3.58}$$

with Q being the covariance matrix of the lognormal distribution and $\mu_i = 1/2 \sum_{l,m} Q_{lm}$ is a constant mean field. In a homogeneous and isotropic universe the covariance for the lognormal distribution can be calculated from the cosmological covariance of the density distribution by relating both correlation functions with each other (see e.g. Coles & Jones 1991, Percival et al. 2004):

$$\xi_{lognormal}(r) = \ln(1 + \xi_{cosmo}(r)). \tag{3.59}$$

The lognormal distribution is convenient for many cosmological applications, such as fast galaxy mock catalog generation, since it resembles many physical properties of the true non-linear density field. However, it is still a phenomenological model and the exact probability distribution for the matter distribution is yet unknown. For some further remarks and discussions of the lognormal distribution the reader is referred to chapter 8.

4. The Sloan Digital Sky Survey

4.1. Introduction to the Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS)[1] is one of the most ambitious international projects in the history of astronomy (York et al. 2000). It is a joint project of many institutions from the USA, Japan, Korea and Germany, including the Max-Planck-Institute for Astronomy in Heidelberg and the Max-Planck-Institute for Astrophysics in Garching. The aim of this survey is to image 10^4 deg^2, an area of about a quarter of the northern sky, and also to conduct spectroscopy of various sources like galaxies, quasars and stars that are selected from the imaging data (Eisenstein et al. 2001). A dedicated 2.5 m telescope, located at Apache Point Observatory in New Mexico, USA, instrumented by a large-format mosaic CCD camera, takes high quality wide-field images in five wavelength bands. Its spectroscopic system is able to simultaneously obtain 640 spectra from various sources. The SDSS-I started in the year 2000 with the aim of mapping more than a quarter of the sky. It completed in summer 2005 and was followed by the SDSS-II, which finished operation in 2008. With the end of SDSS-II in total 1.6 million spectra, including $930,000$ galaxies, $120,000$ quasars and $460,000$ stars have been measured (Abazajian et al. 2009). The SDSS-II is followed by the SDSS-III, which commenced operation in 2008 and will run till 2014.

4.1.1. SDSS surveys

The Sloan Digital Sky Survey-II carried out three distinct surveys aiming at the construction of a large three dimensional map of the Universe, constraining cosmological parameters, exploring the structure, composition and kinetics of the Milky Way and identifying and measuring light curves for Type Ia supernovae (SNe Ia) (Abazajian et al. 2009).

- **Sloan Legacy Survey:**
  The Sloan Legacy Survey is primarily an extragalactic survey designed to probe the three dimensional large scale structure of the Universe. It covers a contiguous area of about 7500 deg^2 of the Northern Galactic Cap and additionally three stripes in the Southern Galactic Cap, yielding a total spectroscopy footprint of 8032 deg^2 (Abazajian et al. 2009). It contains about 230 million celestial objects and provides in particular $929,555$ spectra of galaxies. This information on the angular positions and redshift distances of galaxies permits to investigate the three dimensional large scale structure of the Universe, consisting of voids and filaments, to great accuracy. It also allows for statistical analyses in terms of cosmological power-spectra or higher order correlation functions and the determination of cosmological parameters. In this thesis the three dimensional density field is inferred from a spectroscopic galaxy sample of the the Sloan Legacy Survey (see chapter 9).

- **SEGUE**:
  The Sloan Extension for Galactic Understanding and Exploration (SEGUE) aims at retrieving a detailed three dimensional map of the Milky Way and providing information on the formation history, kinetics, dynamical and chemical evolution of our galaxy (Yanny et al. 2009). The SEGUE complements the Legacy survey with an additional *ugriz* imaging survey of 3500 deg^2 and provides $240,000$ stellar spectra in the disk and spheroid (Yanny et al. 2009). This data gives information on the formation and evolution of our galaxy as well as it reveals the age, composition and phase space distribution of stars within the Milky Way.

- **Sloan Supernova Survey:**
  The Sloan Supernova Survey was a project to search for Type Ia supernovae (SNe Ia) in a redshift range $0.05 \leq z \leq 0.35$ (Frieman et al. 2008). It identified and measured light curves for several hundred Type Ia supernovae using repeated five-band imaging over an area of 300 deg^2 along the celestial equator (Sako

---

[1] http://www.sdss.org/
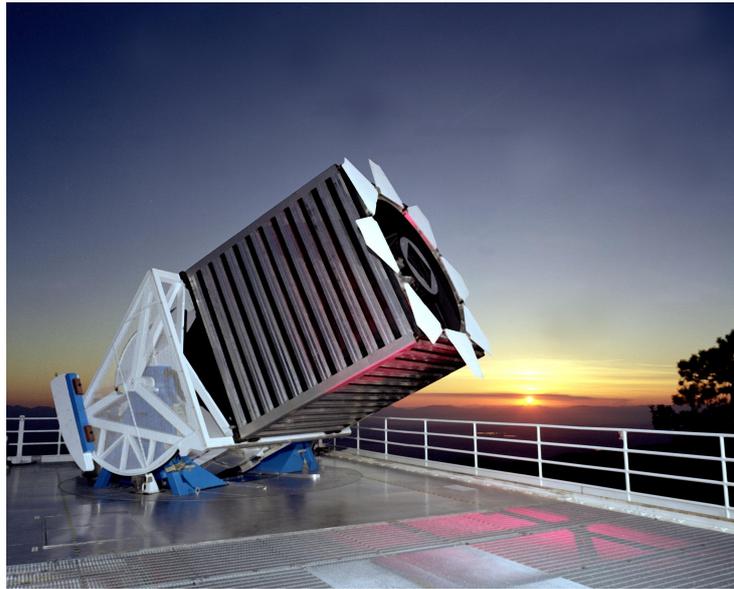
Figure 4.1.: The main 2.5-meter SDSS telescope at Apache Point Observatory in New Mexico, USA. Credit Sloan Digital Sky Survey (http://www.sdss.org/).

et al. 2008). The survey was conducted to address the paucity of SNe Ia data at intermediate redshifts and the minimization of systematics in previous SNe Ia samples (Frieman et al. 2008). A particular scientific goal of the Sloan Supernova Survey is the determination of the properties of dark energy and the cosmological parameters driving the currently observed accelerated expansion of the Universe.

4.1.2. Instrument description

The SDSS uses a special purpose telescope located at Apache Point Observatory in New Mexico, USA. This telescope consists of a 2.5 m $f/2.25$ primary and a 1.08 m secondary mirror. Additionally, it is equipped with a Gascoigne astigmatism corrector and a pair of interchangeable aspheric correctors to provide an undistorted focal plane (Gunn et al. 2006). This wide-field altitude-azimuth telescope is a modified Ritchey-Chrétien design giving a final focal ratio of $f/5$. The telescope enclosure is a rectangular frame structure mounted on wheels which rolls back at night for observing. Further, the telescope is encased in co-rotating wind and light baffles. The advantage of this design is that the telescope is protected from wind and stray light while no dome induced seeing is introduced. The 2.5 m main telescope is accompanied by a 0.5 m photometric telescope situated beside the main telescope. This photometric telescope operates in conjunction with the main telescope to calibrate the photometry of the survey. The 2.5 m SDSS telescope is depicted in Fig 4.1. The SDSS telescope is instrumented with modern digital detectors. A wide-area, multiband CCD camera, cooled by liquid nitrogen to a temperature $T \sim 193$ K is used to carry out the imaging survey. This camera consists of 24 2048×2048 pixel CCDs on the focal plane (Gunn et al. 1998).It is used in time-delay-and-integrate (TDI, or drift scan) mode to provide efficient sky coverage. Additionally, the spectroscopic system consists of a pair of fiber-fed double spectrographs. It can simultaneously obtain 640 spectra from stars, galaxies and quasars covering a broad wavelength range ($3800 - 9200$ Å).

4.1.3. Acknowledgement

5. Data processing

All truly wise thoughts have been thoughts already thousands of times; but to make them truly ours, we must think them over again honestly, till they take root in our personal experience.

Johann Wolfgang von Goethe

Abstract

This chapter provides an overview and discussion of the process of data acquisition, which is prior to any data analysis step. Data acquisition is the sampling of the real world to generate and store data that can be manipulated by a computer. This process irrevocably reduces the information content of a continuous physical system to that one represented by a finite and countable set of states, which can be represented in computer memory. In particular, we will address the sampling of cosmological galaxy distributions and the uncertainties and artifacts introduced by this process. Further, we will discuss the application of Fast Fourier transforms to sampled data systems. The analysis and discussion of the emergent sampling artifacts will then lead to the development of a supersampling method. This method, when applied to the problem of sampling a galaxy distribution, can greately decrease sampling artifacts, permitting highly accurate data analysis.

5.1. Introduction

Inherent to any measurement is the problem of accurately storing the obtained data, in order to hold it ready for further processing. Any known storage device, such as a fotofilm or a computer hard disk, stores information by manipulating only a finite and countable amount of states. Ideal storage of observed data would hence require a bijection of the observed states in our physical environment to the limited amount of states provided by our storage device. It is obvious that this requirement cannot be met in general. In particular, analyzing a continuous system would require a bijection from an uncountable set to a finite countable set of states, which mathematically cannot be achieved. For this reason, in practice, one confines the measured numbers to some finite precision in order to record them. In this fashion, the projection from the real physical system onto our storage device becomes a surjection, implicating that no unique inverse mapping exists. The requirement of storing the results of a measurement for further processing thus introduces a reduction of information, and hence uncertainty. It is important to distinguish this uncertainty from the intrinsic uncertainties of the measurement itself, such as noise. In example, the needle of an analog Voltmeter can represent any voltage within a given range. The result is continuous and consists of the true voltage plus some noise. Even in the absence of noise, the experimenter can only note down the measured voltage with finite precision. This example demonstrates that, although the needle of the Voltmeter is able to represent any continuous value, storing this value is only possible by confining the measurement to finite precision. It is thus, in this case, the process of storing, not the measurement itself, which introduces uncertainty. Also note, that since the process of storing the measurement is a surjection, unique recovery of the true underlying signal is not possible.

Modern data processing is usually carried out with computers, which process the stored measurement in order to extract optimal information on the physical system to be studied. Prior, to any data processing application therefore stands the problem of how to represent the measurements in computer memory and what kind of uncertainties will be introduced by a certain representation. In the case of three dimensional density field and power-spectrum estimation, as discussed in this thesis, we are particularly interested in accurately representing a galaxy observation in our computer, which at the same time permits efficient and fast data processing. Limited amount of computer memory and computational time are here the prime limiting factors for reducing the precision of galaxy measurements. In particular, since we want to use Fast Fourier Transforms (FFT), the spatial volume has to be discretized in a set of equidistant volume elements of finite size. The position of each galaxy within this discretized volume therefore can only be represented to finite precision, which introduces uncertainty to each galaxy position $\vec{x}$. If we assume equal

probability for a specific galaxy to be at any point within the nth volume element, then the position of the galaxy can be estimated as $\vec{x} = \vec{x}_0 + n\Delta\vec{x}/2 \pm \Delta\vec{x}/2$, where $\vec{x}_0$ is a constant offset, and $\Delta\vec{x} = (\Delta x_1, \Delta x_2, \Delta x_3)$ incorporates the side lengths of the volume element. Hence, the discretized measurement does not contain any fine details and becomes fuzzy at the scale of the spatial resolution. Finite spatial resolution directly implicates a truncation (band-limitation) of the Fourier space. This can be understood by considering a position measurement of a point particle with plane waves. In order to resolve the true position of the point particle one requires plane waves with infinite frequency, or accordingly infinitely short wave length. However, in practice one can only test particle positions with finite frequencies, introducing a cutoff at some maximal frequency in Fourier space. This example illustrates that finite spatial resolution implicates a bandlimited Fourier space or vice versa.

This is the basic idea behind Shannon's sampling theorem (Shannon 1948, 1949). Shannon considered the information content of sampled-data systems, and stated that exact information recovery from discretely sampled data is possible, if and only if the signal is band limited. However, many physical signals, such as a distribution of point particles, frequently adopted as a model for galaxy distributions (see e.g. Layzer 1956, Peebles 1980, Martínez & Saar 2002), are not band limited, as it requires infinite resolution to resolve the exact position of each particle. Even though galaxies are objects with finite extent, for many practical applications, such as power-spectrum estimation, the finite size of a galaxy is much smaller than the sampling resolution $\Delta\vec{x}$ of the galaxy density field, and hence is not band limited at frequencies according to the spatial resolution. In these cases, Shannon demonstrated that unique recovery of information from the sampled data is not possible. He also showed that optimal information recovery can be achieved by low-pass filtering and then sampling the continuous signal at discrete positions. Since low-pass filtering amounts to convolving with a sinus cardinal function in real space, it is obvious, that high resolution information of the underlying signal is lost in this procedure.

However, for many practical applications Shannon's theorem is still an idealized theory which strictly speaking is not applicable to real world data analysis. This is mainly due to the fact that ideal low pass-filtering cannot be accomplished in many technical applications. For this reason one usually relies on filter approximations of the ideal low-pass filter, which only have finite support in real-space and hence permit fast computation of the low-pass filtering convolution, for the expense of not perfectly band-limiting the Fourier space.

In cosmology some frequently used techniques to discretize a continuous distribution of point particles are Nearest Grid Point (NGP), Cloud In Cell (CIC) or Triangular Shaped Clouds (TSC) (Hockney & Eastwood 1988), which all attribute a weighted fraction of the individual point particle mass to the surrounding discrete grid positions. The process of sampling by relying on these filter approximation in general not only irrevocably reduces the information content of the continuous signal to that one represented by a set of discrete points, it also introduces sampling artifacts like aliasing or Gibbs ringing (Wolberg 1997, Marschner & Lobb 1994, Cui et al. 2008). These sampling artifacts usually make further processing of the sampled signal difficult. For example, estimation of higher-order spectra, like the bi- or tri-spectrum, with Fast Fourier Transform (FFT) techniques will be erroneous due to sampling artifacts. As mentioned in Cui et al. (2008), so far, there is no known approach to accurately correct for these sampling effects in measuring higher-order spectra like the bi-spectrum with FFT techniques. For the same reasons, these sampling artifacts pose a potential threat to iterative data analysis schemes, as presented in the following chapters. Due to these problems, the aim of signal processing technologies is to find low-pass filter approximations which sufficiently suppress these sampling artifacts, while at the same time still being computationally far less expensive than applying the full ideal low-pass filter. The Digital Signal Processing (DSP) literature provides plenty of possible approaches (see e.g. Smith 2002) which could be introduced to various cosmological signal processing problems.

For cosmological applications recently different approaches have been presented to deal with the problem of sampling artifacts in power-spectrum estimation (Jing 2005, Cui et al. 2008). Jing (2005) proposes to evaluate all aliasing sums of the NGP, CIC and TSC method. He finds that the sampled shot noise contribution to the power-spectrum can be expressed in a simple analytic way. To correct for further aliasing contributions, he introduces an iterative correction method for the power-spectrum, assuming a power-law behavior. However, Jing's method is only applicable to power-spectrum estimations in the absence of artificial mode-coupling, as introduced by a survey geometry and selection effects of a galaxy observation. The method proposed by Cui et al. (2008) utilizes the scale functions of the Daubechechies wavelet transformation as filter approximations. It has been demonstrated to yield better results than standard CIC or TSC methods.

However, most of these approaches rely on increasing the spatial support of the filter approximation to more closely represent the properties of the ideal low-pass filter. This approach is well known in the DSP literature, which favors the windowed sinc functions as optimal approximations to the ideal low pass filter (Theußl et al. 2000,

a)

b)
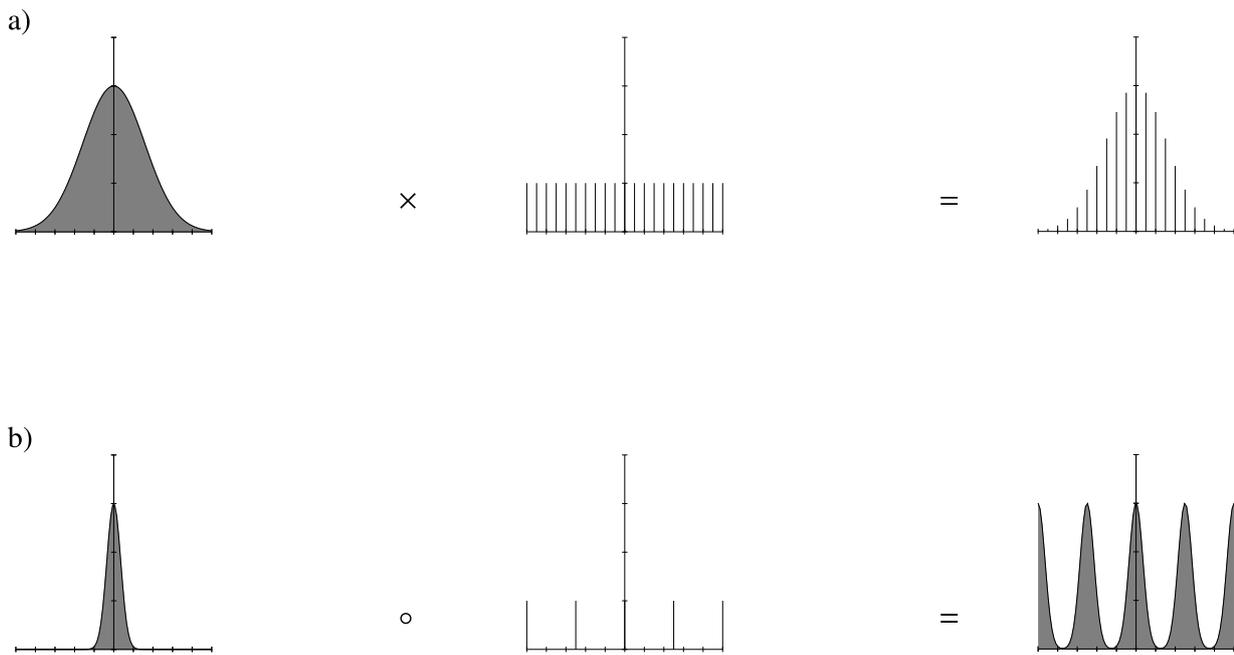
Figure 5.1.: Multiplication a) and convolution b) of a function with the Dirac comb.

Duan et al. 2003, Marschner & Lobb 1994, Smith 2002). These approaches, though, become computationally more expensive as the spatial support of the filter approximation grows, and tends towards evaluating the ideal low-pass filtering convolution. Many applications, like for instance particle mesh codes, however, rely on fast but nevertheless accurate low-pass filtering procedures. Hence, sampling a continuous signal like a galaxy distribution on a discrete grid is always a trade off between computational efficiency and quality.

It is important to point out, that data-analysis schemes, which incorporate convolutions, in order to correct for artificial mode-coupling as introduced by survey geometries or selection effects can be drastically affected by sampling artifacts. The application of a convolution to a sampled data set will distribute false aliased power over broad regions of the reconstructed signal. For this reason, post processing corrections of the sampling artifacts, as proposed in Jing (2005), will not be applicable, as they would require to carefully trace the distribution of all sampling artifacts.

In addition, efficient data analysis algorithms often rely on the aid of discrete Fourier Transforms (DFT) and FFTs as approximations to the continuous Fourier transform. Since DFTs and FFTs assume periodicity, their application to a signal will introduce a second discretization step in Fourier space.

The above discussion demonstrates, that prior to any data analysis step, one carefully has to understand the process of data aquisition, and estimate its implications for further data processing and interpretation of final results. Therefore, in the following, we are going to discuss the implications of sampled-data systems and the application of FFTs. The theoretical discussion will then lead to the introduction of a new supersampling technique to the problem of sampling a three dimensional galaxy distribution. This method is also frequently applied as an anti-aliasing technique in 3D computer graphics (see e.g. Wolberg 1997, Goss & Wu 2000). This method is a two step filtering process in which the signal first is sampled to a discrete grid with super resolution via the CIC or TSC method, then low-pass filtered with the ideal discrete low-pass filter, and finally resampled at target resolution.

In this fashion, by using simple low-pass filter approximation and utilizing FFTs, we provide a fast method to accurately calculate a low-pass filtering procedure, which greately suppresses sampling artifacts.

5.2. The requirements of FFTs

As already mentioned in the introduction, the FFT is a valuable tool in processing huge cosmological datasets due to its computational efficiency. This allows us to apply many mathematical operations like convolutions or deconvolutions to the data in an computational feasible way. However, using FFTs relies on two strong requirements as the function has to be discrete not only in real-space but also in Fourier-space.

It is a well known fact that a natural physical signal, like a galaxy distribution, is neither living in a discretized real nor a discretized Fourier-space. The reduction of such a signal to a set of finite and discrete sample points, thus, introduces an enormous loss of information.

As will be demonstrated below, the sampling theorem by Claude Shannon (Shannon 1948, 1949) requires a function to be band limited, meaning, that the function has to be sufficiently smooth in order to have its support extending only up to a maximal frequency in Fourier-space. If this criterion would be fulfilled, then the complete information of the continuous signal can be represented by a set of discrete sample points in real-space. However, physical theories about the dark matter distribution and the power-spectrum tell us, that the power-spectrum possibly extends over all frequencies in Fourier-space, meaning that the matter and galaxy distributions cannot be sampled on a discrete grid without loss of information.

In addition, as already mentioned above, discreteness in real-space is not enough, and must be understood only as a necessary requirement for using FFTs. Using FFTs additionally enforces the Fourier-space to be discrete. This is in agreement with the fact that the FFT requires the real-space signal to be periodic, meaning the signal can be represented by a finite set of Fourier waves.

For these reasons we claim the requirement of real-space and Fourier-space discreteness to be strong criteria in hindsight of information conservation, as in general no physical signal will fulfill these requirements in a natural way.

In the following we will discuss how to optimally sample a continuous signal in order to conserve as much information as possible. In doing so, we will arrive at the ideal instrument response function of our computer, which allows us further intuitive insight into the problem of representing a real physical quantity on a grid of finite sample points.

5.3. Discretizing the real-space

As already mentioned a necessary requirement for the application of FFTs is the real-space discreteness of a signal. However, for the case discussed in this thesis the real physical signal is continuous in real-space and hence has to be discretized. This is usually achieved by dividing the continuous signal into small regions and associating a number with each of those. As this process also involves a loss of spatial resolution, and therefore requires to soften sharp features, information gets lost. This process of converting a continuous signal into a discrete representation can cause several artifacts like aliasing or Gibbs ringing.

A good way to understand, analyze and eliminate these artifacts is through Fourier analysis. In the following we will discuss the necessary requirements for real-space discretization, as was demonstrated by Claude Shannon in his sampling theorem (Shannon 1948, 1949).

5.3.1. Sampling theorem

In order to discretize a continuous function, we will represent a point sample as a set of scaled Dirac impulse function.

With this definition, sampling a signal is equivalent to multiplying it by a grid of impulses, one at each sample point $x_j = j\Delta x$, where j is an integer and Δx is the grid spacing (Marschner & Lobb 1994), as illustrated in Figure 5.1.

Let $f(x)$ be a continuous function, then its sampled version $g(x)$ can be expressed by:

$$g(x) = \Pi(x) f(x),\qquad(5.1)$$

where the sampling function $\Pi(x) = \sum_{j=-\infty}^{\infty} \delta^D(x - j\Delta x)$ is a Dirac comb or impulse train. With this definition we

yield the discrete function at the sample positions g_j as:

$$g_j \equiv f(j\Delta x),\qquad(5.2)$$

However, in order to demonstrate the origin of aliasing effects in the following we will focus on the Fourier transform of equation (5.1). By making use of the convolution theorem A.1.1, which states that the product of two functions in real-space yields a convolution in Fourier-space, the Fourier transform of the sampled function $\hat{g}(p)$ can be expressed as:

$$\hat{g}(p) = \left(\hat{\Pi} \circ \hat{f}\right)(p),\qquad(5.3)$$

where theˆ-symbol denotes the Fourier transform of a function and the circle symbol ∘ denotes a convolution.

As demonstrated in Appendix A.1.2 the Fourier transform of the Dirac comb $\hat{\Pi}(p)$ is again a Dirac comb and is given by:

$$\hat{\Pi}(p) = p_s \sum_{j=-\infty}^{\infty} \delta^D\left(p - j\,p_s\right),\qquad(5.4)$$

where $p_s = 2\pi/\Delta x$ is the repetition length in Fourier-space, which is often called the Nyquist rate. Since $\hat{\Pi}(p)$ is a Dirac comb the convolution in equation (5.3) amounts to duplicating $\hat{f}(p)$ at every point of $\hat{\Pi}(p)$:

$$\begin{aligned}
\hat{g}(p) &= \left(\hat{\Pi} \circ \hat{f}\right)(p) \\
&= p_s \int_{-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \delta^D\left(p' - j\,p_s\right) \hat{f}(p - p')dp' \\
&= p_s \sum_{j=-\infty}^{\infty} \hat{f}\left(p - j\,p_s\right),\qquad(5.5)
\end{aligned}$$

as displayed in Figure 5.1. This demonstrates that the Fourier representation of the sampled function $\hat{g}(p)$ is a superposition of the true continuous Fourier transform $\hat{f}(p)$ and all its replicas at positions $j\,p_s \equiv j(2\pi)/\Delta x$. We call the copy of $\hat{f}(p)$ at $j = 0$ the primary spectrum and all other copies alias spectra. This result reveals that the sampling operation has left the original input spectrum $\hat{f}(p)$ intact, just replicating it periodically in the Fourier domain with a spacing of p_s (Wolberg 1997).

This suggests to rewrite the sampled spectrum $\hat{g}(p)$ as a sum of two terms, the low-frequency (baseband), and the high frequency (aliasing) components;

$$\begin{aligned}
\hat{g}(p) &= \hat{f}(p) + \sum_{j=-\infty, j\neq 0}^{\infty} \hat{f}\left(p - jp_s\right) \\
&= \hat{f}(p) + \hat{F}_{HF}(p)\qquad(5.6)
\end{aligned}$$

The baseband spectrum is exactly $\hat{f}(p)$, and the high-frequency components, $\hat{F}_{HF}(p)$, consists of the remaining replicated versions of $\hat{f}(p)$ that constitute harmonic versions of the sampled signal (Wolberg 1997).

A crucial observation in the study of sampled-data systems is that exact signal reconstruction from the sampled data requires to discard the replicated spectra $\hat{F}_{HF}(p)$, leaving only $\hat{f}(p)$, the spectrum of the signal we seek to recover (Wolberg 1997).

If the different replicas $\hat{f}(p)$ do not overlap with the baseband spectrum, then we can recover $\hat{f}(p)$ from $\hat{g}(p)$, by simply multiplying $\hat{g}(p)$ with a function which is one inside the baseband and zero elsewhere, and therefore eliminates the high frequency contributions, see Figure 5.2 a). On the other hand, if the high-frequency contribution $\hat{F}_{HF}(p)$ has some overlap with the baseband spectrum, there is no way to uniquely recover the original signal $\hat{f}(p)$ from its sampled version $\hat{g}(p)$, see Figure 5.2 b). Therefore, the only provision for exact sampling is that $\hat{f}(p)$ must be undistorted due to the overlap with $\hat{F}_{HF}(p)$ (Wolberg 1997). For this to be true, two conditions must hold:

1. The signal must be bandlimited. This avoids functions $\hat{f}(p)$ with infinite extent that are impossible to replicate without overlap.

2. The sampling frequency $p_s = (2\pi)/\Delta x$ must be greater than twice the maximum frequency p_{max} present in the signal. This can be understood by looking at Figure 5.2. This minimum frequency, known as the Nyquist
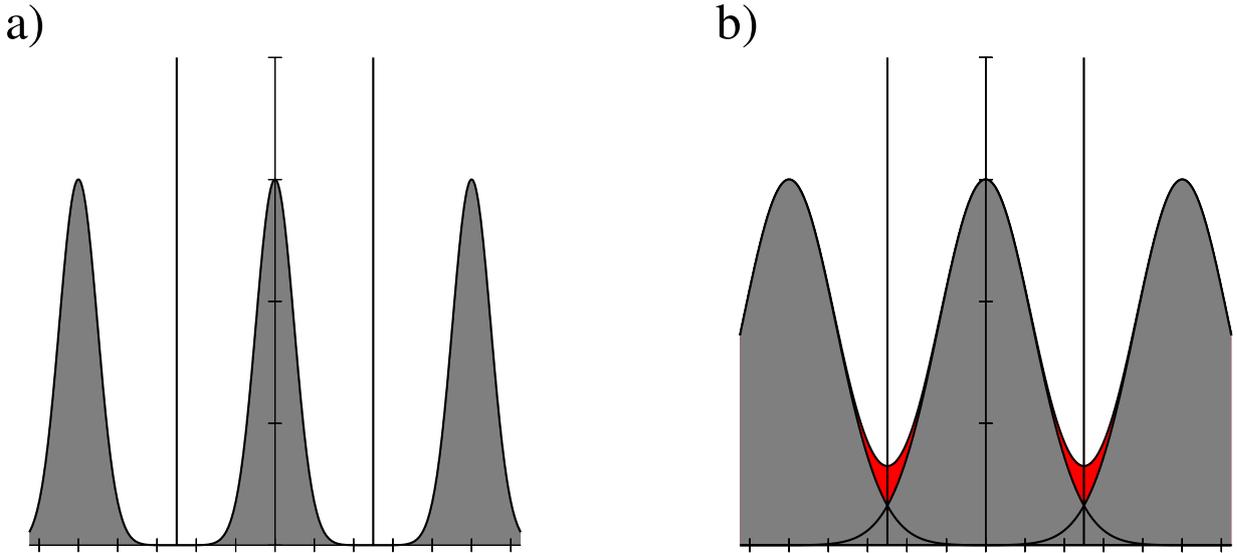
Figure 5.2.: a) Undistorted sampled function and b) sampled function with aliasing.

rate, is the minimum distance between the spectra copies, each with bandwidth p_s.

The first condition merely ensures that a sufficiently large sampling frequency p_s exists that can be used to separate replicated spectra from each other (Wolberg 1997). The second condition provides an answer to the problem of the sufficiency of data samples to exactly recover the continuous input signal. It states that the signal can only be recovered exactly when $p_s > 2 \, p_{max}$, with $p_{max} = p_{Nyquist}$ being the Nyquist frequency, not to confuse with the Nyquist rate (Wolberg 1997).

This is the sampling theorem as pioneered by Claude Shannon in his papers (Shannon 1948, 1949).

5.3.2. Low-Pass-Filtering

As seen in the previous section, when a signal is being sampled it must be band-limited if we are to recover its information content correctly. Natural signals, however, are not generally band-limited, and so must be low-pass filtered before they are sampled or equivalently, the sampling operation must include some form of local averaging.

It is therefore required to apply a filter to the signal, which cuts away all Fourier modes higher than a certain maximal frequency $p_{max} = p_s/2$. Such a filter is called an ideal low pass filter, and it's Fourier representation is given as:

$$\hat{W}(p) = \begin{cases} 1 & \text{for } |p| < p_{max} \\ 0 & \text{for } |p| \geq p_{max} \end{cases} . \tag{5.7}$$

It is ideal in the sense, that it has unity gain in the pass-band region, hence not introducing any attenuation of the Fourier modes to pass, and that it perfectly suppresses all the power in the stop-band region. Applying this ideal low-pass filter to the signal by multiplication in Fourier-space yields the low-pass filtered signal:

$$\hat{h}(p) = \hat{W}(p) \, \hat{f}(p), \tag{5.8}$$

which according to the convolution theorem A.1.1 translates to its real-space representation as:

$$h(x) = (W \circ f)(x). \tag{5.9}$$

The real-space representation $W(x)$ of the ideal low pass filter is calculated in Appendix A.1.3 and yields a sinc
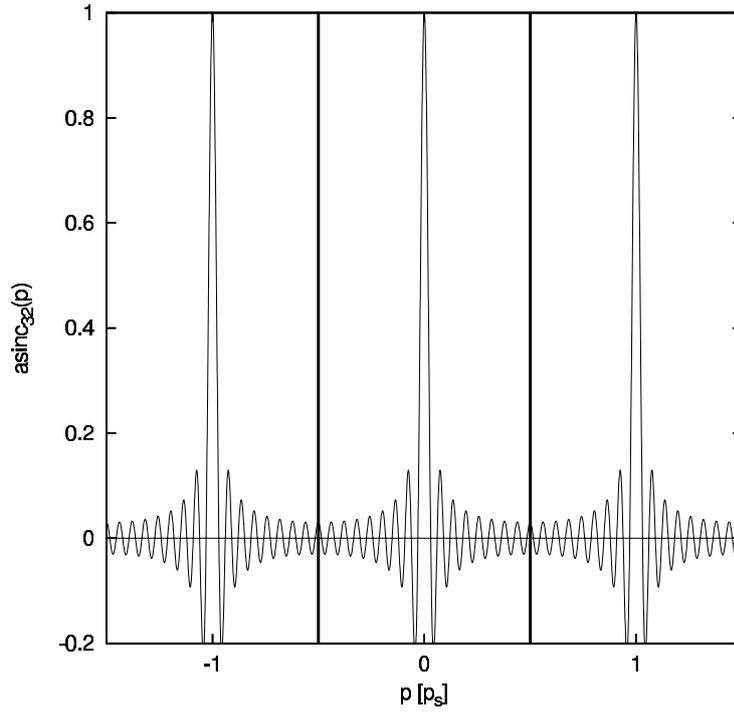
Figure 5.3.: Aliased sinc function for N=32. The vertical lines limit the base-band.

function:

$$W(x) = \frac{p_{max}}{\pi} \, sinc(p_{max}x) \,. \tag{5.10}$$

Hence, low-pass filtering is equivalent to convolving with a sinc function in real-space.

As the sinc function has not such a sharp localization in space as a Dirac delta distribution, any sharp feature will be smeared out. In this sense low-pass filtering reduces spatial information, which cannot be uniquely restored.

The low-pass filtered function $h(x)$ now meets the requirements of the sampling theorem, and can be discretized in real-space with a grid spacing $\Delta x < \pi/p_{max}$. With these requirements the sampling procedure for a natural signal turns into:

$$g(x) = \Pi(x) \, h(x) = \Pi(x) \, (W \circ f) \, (x) \,. \tag{5.11}$$

This process is graphically represented in Figure 5.5 a).

5.3.3. Sampling on a finite real-space domain

We have demonstrated above, that a band-limited signal can be uniquely recovered from its discretized approximation. This is due to the fact that there exists a sufficiently high sampling frequency to separate the aliased copies from the base-band spectrum and that the sampling operation has left the original spectrum $\hat{f}(p)$ intact. However, this is closely related to the fact that the samples are distributed over the infinite domain of real-space as the sum in the sampling operator $\Pi(x)$ extends from minus to plus infinity since only then the Fourier transform of the sampling operator is again a Dirac comb. Unfortunately, this is not the case in real world applications, especially since it is generally not possible to evaluate the infinite sampling sum with a computer. Therefore, one generally wants to restrict the investigation to a finite subset of samples. For technical reasons one usually chooses a rectangular sub domain. Therefore the sampling operator can be split in a zero centered sub domain and the surrounding rest as

follows:

$$
\begin{aligned}
\Pi(x) &= \sum_{j=-\infty}^{\infty} \delta^D(x - j\Delta x) \\
&= \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \delta^D(x - j\Delta x) + \sum_{j=\frac{N}{2}+1}^{\infty} \left[\delta^D(x - j\Delta x) + \delta^D(x + j\Delta x) \right] \\
&= \Pi_N(x) + \Xi_N(x),
\end{aligned}
\tag{5.12}
$$

where we introduced the finite sampling operator $\Pi_N(x)$ and the trans-domain sampling operator (Jordan operator) $\Xi_N(x)$. The sampled signal on the sub-domain is then given by:

$$
g_N(x) = g(x) - R_N(x),
\tag{5.13}
$$

with $g_N(x) = \Pi_N(x)f(x)$ and the rest term $R_N(x) = \Xi_N(x)f(x)$. As demonstrated above $g(x)$ contains all information to uniquely recover the band-limited signal. Thus, by reducing the set of signal samples to a finite subset information gets lost and the signal cannot be reconstructed uniquely. This can be easily seen by Fourier analysis.

According to Appendix A.1.4 the Fourier transform of the finite sampling operator is given as:

$$
\hat{\Pi}_N(p) = \mathrm{asinc_N}(p),
\tag{5.14}
$$

with the aliased sinc function $\mathrm{asinc_N}(p)$ being:

$$
\mathrm{asinc_N}(p) = \frac{\sin\left(\frac{p\Delta x}{2}(N+1)\right)}{\sin\left(\frac{p\Delta x}{2}\right)}.
\tag{5.15}
$$

The aliased sinc function is plotted in Fig. 5.3. It is obvious that the Fourier transform of the finite sampling operator is not a Dirac comb anymore, though it still exhibits the property of having resonance peaks at the repetition length p_s. Sampling on a finite domain therefore amounts to convolving the Fourier spectrum of the true continuous signal with the aliased sinc function. This operation however does not leave the base-band spectrum intact, and, unlike as in the Sampling theorem, unique recovery of the underlying signal is not possible anymore. It is also obvious that due to the shape of the aliased sinc function the base-band spectrum will always be affected by the higher order spectra, which are overlapping with the baseband. However, low pass filtering will not completely eliminate these aliasing contributions but it can alleviate them.

Therefore, in general it is not possible to uniquely restore information from signals which have been sampled on finite domains, and interpretations drawn from these sampled signals will always be afflicted with some uncertainty.

5.4. Discretizing The Fourier-space

In the previous section we discussed the theory of discretizing the real-space representation of a continuous signal. We demonstrated that in real world applications due to the finite sampling operator the sampling theorem is not applicable anymore in a strict sense. However, discretizing the real-space representation of the signal is usually not enough for data processing purposes. It can for instance be easily demonstrated that the Fourier transform of such a discrete signal is still a continuous function in Fourier-space. Nevertheless, especially applications in which FFT techniques are used require also discrete Fourier-space representations of the signal. Therefore, in the following we will derive a sampling method, designed to be used with FFTs.

5.4.1. FFTs and the Fourier-space representation

FFTs, since they are based on the Fourier-series, require the function to be periodic on the observational interval L. This assumes that the observed signal can be decomposed into eigen-modes of a resonator with length L. This resonator has a ground frequency of $p_0 = 2\pi/L$ and all other harmonic frequencies can be obtained by multiplying with an integer $p_i = p_0 i$. Thus, the Fourier domain observation obtained via an FFT is discrete. It is worth

mentioning, that this discreteness in Fourier-space arises from applying the FFTs and therefore implicitly assuming the function to be periodic on the observational interval. In doing so we introduced a second discretizing process, but this time the Fourier-space is being discretized. Thus, FFTs should be considered as an additional filter for the true underlying signal.

As a consequence the question arises how the Fourier modes obtained by an FFT of the observed signal are related to the Fourier modes of the true natural signal. This is of special interest in the case of cosmological matter power-spectrum estimation. To demonstrate this we will follow the usual ideal approach to discretize a signal, by first low-pass filtering and then sampling it to grid positions. The filtered and gridded signal can then be represented by:

$$g_j = \int_{-\infty}^{\infty} dx\, W(\Delta x\, j - x)\, f(x)\,, \tag{5.16}$$

where $W(x)$ is the ideal low-pass filter as given by equation (5.10). Applying the FFT as defined in Appendix A.1.6 yields:

$$
\begin{aligned}
\hat{g}_k &= \hat{C} \sum_{j=0}^{N-1} g_j\, e^{-2\pi jk \frac{\sqrt{-1}}{N}} \\
&= \hat{C} \sum_{j=0}^{N-1} e^{-2\pi jk \frac{\sqrt{-1}}{N}} \int_{-\infty}^{\infty} dx\, W(\Delta x\, j - x)\, f(x) \\
&= \frac{\hat{C}}{2\pi} \int_{-\infty}^{\infty} dp\, \hat{f}(p) \sum_{j=0}^{N-1} e^{-2\pi jk \frac{\sqrt{-1}}{N}} \int_{-\infty}^{\infty} dx\, W(\Delta x\, j - x)\, e^{\sqrt{-1}\, p\, x} \\
&= \frac{\hat{C}}{2\pi} \int_{-\infty}^{\infty} dp\, \hat{f}(p) \sum_{j=0}^{N-1} e^{-2\pi jk \frac{\sqrt{-1}}{N}} \left(e^{\sqrt{-1}\, p\, \Delta x j} \right) \int_{-\infty}^{\infty} dx'\, W(x')\, e^{-\sqrt{-1}\, p\, x'} \\
&= \frac{\hat{C}}{2\pi} \int_{-\infty}^{\infty} dp\, \hat{f}(p) \hat{W}(p) \sum_{j=0}^{N-1} e^{-2\pi jk \frac{\sqrt{-1}}{N}} \left(e^{\sqrt{-1}\, p\, \Delta x j} \right) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp\, \hat{f}(p)\, \hat{W}(p)\, \hat{C} \sum_{j=0}^{N-1} e^{-2\pi jk \frac{\sqrt{-1}}{N}} \left(e^{\sqrt{-1}\, p\, \Delta x j} \right) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp\, \hat{f}(p)\, \hat{W}(p)\, \hat{C} \sum_{j=0}^{N-1} \left(e^{-\sqrt{-1}\Delta x j(\Delta p k - p)} \right) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp\, \hat{f}(p)\, \hat{W}(p)\, \hat{U}(\Delta p k - p)\,,
\end{aligned}
$$

$$\tag{5.17}$$

where we have introduced the Fourier-space lattice interval $\Delta p = 2\pi/(\Delta x N) = 2\pi/L$ and the mode coupling function $\hat{U}(\Delta p k - p)$:

$$
\begin{aligned}
\hat{U}(\Delta p k - p) &= \hat{C} \sum_{j=0}^{N-1} e^{-\sqrt{-1}\Delta x j(\Delta p k - p)} \\
&= \hat{C} \sum_{j=0}^{N-1} \left(e^{-\sqrt{-1}\Delta x(\Delta p k - p)} \right)^j \\
&= \hat{C}\, \frac{1 - e^{-\sqrt{-1}\Delta x(\Delta p k - p) N}}{1 - e^{-\sqrt{-1}\Delta x(\Delta p k - p)}} \\
&= \hat{C}\, e^{-\sqrt{-1}\frac{\Delta x}{2}(\Delta p k - p)(N-1)}\, \frac{e^{\sqrt{-1}\frac{\Delta x}{2}(\Delta p k - p) N} - e^{-\sqrt{-1}\frac{\Delta x}{2}(\Delta p k - p) N}}{e^{\sqrt{-1}\frac{\Delta x}{2}(\Delta p k - p)} - e^{-\sqrt{-1}\frac{\Delta x}{2}(\Delta p k - p)}}
\end{aligned}
$$
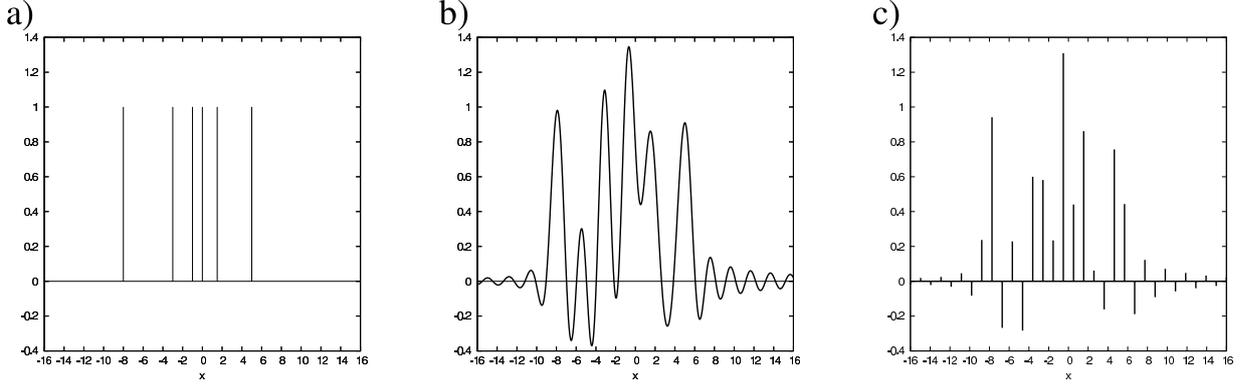
Figure 5.4.: Sampling process of a galaxy distribution. The galaxy distribution displayed in panel a) is low-pass filtered in panel b) and finally sampled at discrete positions in panel c).

$$
= \hat{C}\, e^{-\sqrt{-1}\frac{\Delta x}{2}(\Delta pk - p)(N-1)} \frac{\sin\left(\frac{\Delta x}{2}(\Delta pk - p)\,N\right)}{\sin\left(\frac{\Delta x}{2}(\Delta pk - p)\right)}
$$

$$
= \hat{C}\, e^{-\sqrt{-1}\frac{\Delta x}{2}(\Delta pk - p)(N-1)}\, asinc(\Delta pk - p)\,. \tag{5.18}
$$

This immediately demonstrates that the spectral representation obtained by the FFT is a filtered version of the true Fourier transform of the signal. It can be easily seen, that if the true signal is periodic on the observational domain L then the convolution with the filter kernel $\hat{U}(p)$ vanishes. Hence, in this case we obtain the true Fourier representation of the signal. However, as natural signals are in general not periodic on the observational domain they cannot be decomposed in a discrete set of Fourier waves, and hence some Fourier smoothing must be introduce, in order to represent the non-periodic function by a periodic function. It is also worth noticing, that the mode coupling function does not only affect the amplitude of certain modes, but also introduces a phase shifting factor. This is due to the fact, that the FFT applies a window which is not centered on the origin of the real-space domain. For these reasons it is not possible to uniquely recover the true Fourier spectrum from the FFT Fourier representation of the signal.

5.4.2. Sampling in Fourier-space

In the previous section we demonstrated that applying an FFT or in more general a DFT to a low pass filtered non-periodic signal yields a Fourier representation, which deviates from the true Fourier representation by a convolution with the mode coupling function $\hat{U}(p)$. The continuous signal therefore has to be filtered in such a way that its Fourier representation has a discrete spectrum. This can be achieved by sending the signal through a resonator of length L, which effectively means the convolution with a Dirac comb in real-space. The proof of this can trivially been shown by making use of the convolution theorem, and knowing that the Fourier transform of the Dirac comb is again a Dirac comb. We introduce the realspace replication operator $\Pi_R(x)$ as

$$
\Pi_R(x) = \sum_{j=-\infty}^{\infty} \delta^D(x - jL)\,, \tag{5.19}
$$

with its Fourier transform, according to Appendix (A.1.2), being:

$$
\hat{\Pi}_R(p) = \frac{2\pi}{L} \sum_{j=-\infty}^{\infty} \delta^D(p - j\,\Delta p)\,. \tag{5.20}
$$

The replication operator allows us therefore to formulate the necessary step of discretizing the Fourier-space representation of the continuous signal, by evaluating the following convolution:

$$f_R(x) = (\Pi_R \circ f)(x), \tag{5.21}$$

where $f_R(x)$ is the replicated signal in real-space. The according Fourier space representation then reads:

$$\hat{f}_R(p) = \hat{\Pi}_R(p)\,\hat{f}(p) = \frac{2\pi}{L} \sum_{j=-\infty}^{\infty} \delta^D(p - j\Delta p)\,\hat{f}(p). \tag{5.22}$$

This immediately demonstrates that $\hat{f}_R(p)$ is discrete in Fourier-space. Forcing a continuous non-periodic function to be periodic, or sending it through a resonator, is a filtering process, which discretizes the Fourier space representation. We can now replace $\hat{f}(p)$ by the discretized function $\hat{f}_R(p)$ in equation (5.17) to yield:

$$
\begin{aligned}
\hat{g}_k &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp\,\hat{f}_R(p)\,\hat{W}(p)\,\hat{U}(\Delta p k - p) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} dp\,\frac{2\pi}{L} \sum_{l=-\infty}^{\infty} \delta^D(p - l\Delta p)\,\hat{f}(p)\,\hat{W}(p)\,\hat{U}(\Delta p k - p) \\
&= \frac{1}{L} \sum_{l=-\infty}^{\infty} \hat{f}(l\Delta p)\,\hat{W}(l\Delta p)\,\hat{U}(\Delta p\,(k-l)).
\end{aligned} \tag{5.23}
$$

Note, that the mode coupling function now only depends on two integers, and therefore is a matrix in discrete Fourier-space. The mode coupling function is then:

$$
\begin{aligned}
\hat{U}(\Delta p\,(k-l)) &= \hat{C} \sum_{j=0}^{N-1} e^{-\sqrt{-1}\Delta x \Delta p j (k-l)} \\
&= \hat{C} \sum_{j=0}^{N-1} e^{-2\pi\sqrt{-1}j\frac{(k-l)}{N}} \\
&= \hat{C}\,N\,\delta_{lk}^{K},
\end{aligned} \tag{5.24}
$$

where for the last equality we refer to Appendix A.1.7.

Using this result in equation (5.23) then yields:

$$\hat{g}_k = \frac{\hat{C}\,N}{L}\,\hat{f}(k\,\Delta p)\,\hat{W}(k\,\Delta p). \tag{5.25}$$

This is a remarkable result. As the ideal low-pass filter $\hat{W}(p)$, as given by equation (5.7), has unity gain in the pass band, it is possible to uniquely recover individual Fourier modes of the true physical continuous and non-periodic signal.

It is therefore clear, that the use of FFT techniques requires to sample the function twice, once in real-space, and once in Fourier-space. This can simply be achieved by replicating the continuous and non-periodic signal and low pass filtering it. According to Appendix A.1.5 the ideal discretization filter kernel $\Psi(x)$ would then be an aliased sinc function:

$$\Psi(x) = \frac{1}{L}\mathrm{asinc}_N(x). \tag{5.26}$$

This filtering procedure can be accomplished in a two step filtering process, where one first applies the replication operator and then the low-pass filter, or vice versa, to the continuous non-periodic signal. From this two step filtering it is also clear that there exist two parameters to adjust the sampled field. The first parameter is the cut-off frequency p_{max} of the low-pass filter, which controls the real-space resolution, the second parameter is the resonator length L which controls the Fourier-space resolution, or the spatial domain under consideration.

a)
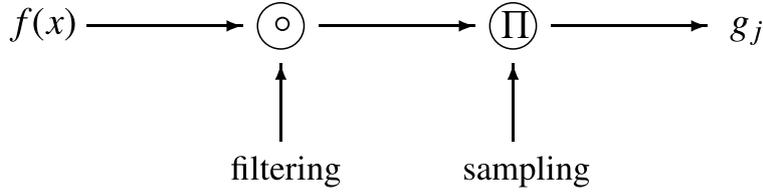
b)
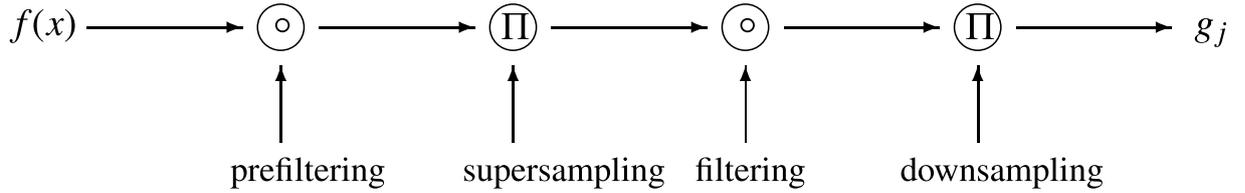
Figure 5.5.: Ideal sampling scheme a) and two stage supersampling scheme b).

5.4.3. The instrument response function of our computer and the loss of information

In the previous sections we demonstrated that in order to process information via FFT techniques on a computer the function has to be band limited and discrete in Fourier-space meaning the function has to be periodic on the observational interval and can be represented by only a finite amount of Fourier waves. As natural signals in general do not obey these requirements processing the data with a computer has to be understood as an additional observational and filtering step, which modifies the input data.

Therefore, in order to make inferences from the information stored on a computer towards the information about a real physical observable, one must take into account the systematics introduced by the sampling process and the use of FFTs or DFTs. The discrete representation on a computer is only an approximation to the real continuous signal, and hence cannot represent all properties of the true original physical observable. This will be demonstrated now by the case of the strictly positive galaxy density field.

In many cosmological applications, like for instance power-spectrum estimation, galaxies are considered to be point particles represented by a Dirac delta distribution. With this definition a galaxy distribution consisting of M galaxies can be expressed as:

$$\rho(x) = \sum_{p=1}^{M} \delta^D(x - x_p),$$

(5.27)

with x_p being the position of a galaxy. As can be seen easily this density field is strictly positive by definition. However, applying the ideal low pass filter yields:

$$
\begin{aligned}
\rho'(x) &= \int_{-\infty}^{\infty} dx' \, W(x - x')\rho(x') \\
&= \sum_{p=1}^{M} W(x - x_p).
\end{aligned}
$$

(5.28)

This can be understood as replacing the Dirac delta distribution by the ideal low-pass filter kernel and therefore giving each galaxy the shape of a sinc function.

In Figure 5.4 we display a simple example of a low-pass filtered galaxy distribution with positions $x_p \in \{-8, -3, -1, 0, 1, 5\}$ to be sampled on a grid with grid spacing $\Delta x = 1$. As already suggested above, the low-pass filtering introduces

some sort of spatial smoothing due to the finite width of the sinc kernel. Hence, two galaxies being closer together than the grid spacing cannot be resolved independently as can be seen in the cases of the galaxies at position $x_p = -1$ and $x_p = 0$. Another important thing to remark is that due to the oscillatory nature of the sinc function, the superposition of sinc functions will lead to positive and negative interference. This might introduce density peaks at positions where no peak would be observed in the true natural signal, which is demonstrated in Figure 5.4 at position $x = -5.5$.

As the sinc function is not strictly positive, the low pass filtered galaxy density given in equation (5.28) does not possess the physical property of being strictly positive as the original observable, demonstrated as well by Figure 5.4. This property will only be restored in the limit of infinite resolution, when the sinc function approaches the Dirac delta distribution.

Being a strictly positive density field thereby is a physical property which cannot be represented by the sampled galaxy density field. This result will also be true for other physical properties which will only be recovered in the limit towards infinite resolution.

However, it is worthwhile mentioning that though the galaxy density field may have negative contributions all integral quantities like total number of galaxies or total mass, are identical to the ones which could be obtained by integration over the true natural signal. This is due to the fact that the low-pass filter conserves the zeroth Fourier mode of the true natural signal. For this reason, it is also not possible to fix the negative contributions by cutting them away or taking the absolute values of the low-pass filtered signal, as this operations will not conserve the zeroth Fourier mode and therefore violates conserved quantities like the total number or total mass of galaxies.

5.5. Sampling 3d galaxy distributions

In the sections above we examined the theory of discretizing continuous signals and demonstrated some of the problems coming with this approach. For simplicity we discussed only the 1d case, but all results can be extended straight forwardly to the 3d case. In this section we will discuss the practical implementation of sampling methods in order to process galaxy distributions via FFT techniques.

5.5.1. Ideal sampling procedure

As already described in section 5.4.3 a galaxy distribution can be expressed as a set of point particles. With this definition a 3d galaxy distribution consisting of M galaxies can be expressed as:

$$\rho(\vec{x}) = \sum_{p=1}^{M} \delta^D(\vec{x} - \vec{x_p}),$$ (5.29)

where $\vec{x_p}$ is the three dimensional position vector of a galaxy.

Note, that assuming a galaxy to be a point particle is a strong prior. Therefore in more general it is possible to allow for the galaxies to have a shape $s(\vec{x})$, which describes the physical extend of a galaxy into space. The new density field $\rho'(\vec{x})$ is then obtained by simple convolution of equation (5.29) with the galaxy shape $s(\vec{x})$:

$$\rho'(\vec{x}) = (s \circ \rho)(\vec{x}) = \sum_{p=1}^{M} s(\vec{x} - \vec{x_p}),$$ (5.30)

However, such a modification of the galaxy shape would only be necessary if the grid resolution is sufficient enough to resolve individual galaxies and their inner structure, which is not the usual case for applications as discussed in this thesis. Therefore, in the following a galaxy distribution will always be considered to be a set of point particles.

The galaxy distribution as given by equation (5.29) can be sampled to a grid according to the ideal sampling scheme displayed in Figure 5.5, which consists of two steps. In the first step the high frequency contributions of the galaxy distribution are eliminated by applying the ideal low-pass filter:

$$\varrho(\vec{x}) = (W \circ \rho)(\vec{x}) = \sum_{p=1}^{M} W(\vec{x} - \vec{x_p}),$$ (5.31)

a)

b)
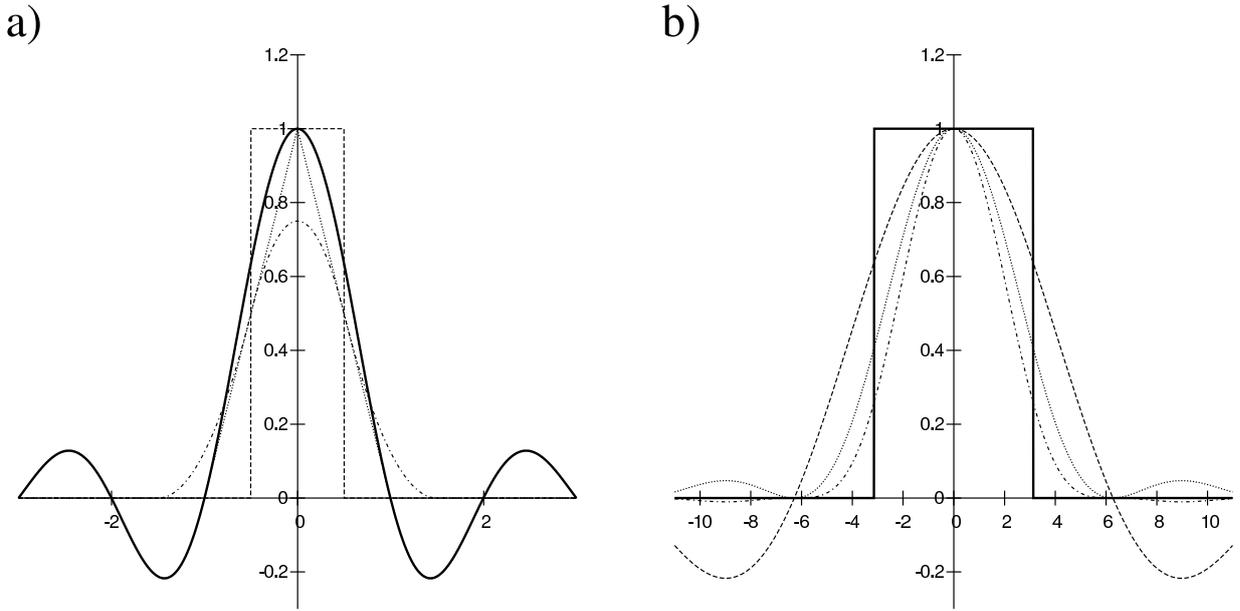
Figure 5.6.: Comparison of the real (a) and Fourier-space (b) representations of the NGP kernel (dashed curve), the CIC kernel (dotted curve) and the TSC kernel (dashed dotted curve) to the ideal low-pass filter (solid curve).

which simply states that the galaxies have been given the artificial shape of the 3d sinc function $W(\vec{x})$. In a following step the low-pass filtered distribution will be sampled at discrete positions on the grid. For practical implementation these two steps will usually be done in one step, by simply convolving the galaxy distribution to the discrete gridpoints:

$$\varrho_{\vec{n}} = \sum_{p=1}^{M} W(\Delta x \, \vec{n} - \vec{x_p}), \tag{5.32}$$

where Δx is the grid spacing and $\vec{n}$ is the triplet of integers $\vec{n} = (i, j, k)$. The sum in equation (5.32) describes the ideal sampling procedure, which allows to have the best information conservative representation of the continuous signal in discrete form. However, as the ideal low-pass filter kernel extends over all space, the sum over all particles has to be evaluated for each individual voxel of the discrete grid, which makes this procedure impractical for real world applications.

5.6. Practical sampling

Due to the infinite support of the ideal low-pass filter kernel in real-space the ideal sampling method (by convolving with such a filter) is in general not feasible. For many applications it is computationally too expensive. For this reason one usually chooses a practical approach by approximating the ideal sampling operator, by a less accurate, but faster calculable sampling operator. This usually means to approximate the low pass filter, by a function with compact support in real-space. As a result the convolution can be calculated faster, for the prize of not completely suppressing the aliasing power in the stop-bands. Approximating the ideal sampling operator therefore is always a trade off between accuracy and computational speed.

In the literature of Digital Signal Processing this problem is well known and studied for many years already. Especially the literature of modern 3D computer graphics provides a lot of practical solutions to the problem of sampling and filtering. Many very detailed studies about the optimal filter approximations have been made, and can be found in the signal processing literature (Theußl et al. 2000, Mitchell & Netravali 1988, Marschner & Lobb

1994, Wolberg 1997).

In the following we will discuss some sampling techniques as commonly used in cosmology and display their strength and weaknesses.

5.6.1. Filter approximation

There are several known practical issues when dealing with filter approximation (Marschner & Lobb 1994). Using filter or filter approximation will in general lead to a variety of effects like smoothing ("blurring"), which refers to the removal of rapid variations in the signal by spatial averaging. Or, one will observe a reshaping of the Fourier-space by applying a filter which has not unity gain at every mode in Fourier-space. In addition, low-pass filtering step discontinuities will result in oscillations or ringing, just before or after discontinuities (Marschner & Lobb 1994). This is the Gibbs phenomenon, and is due to the fact, that step discontinuities in the signal cannot be represent by a finite superpostion of Fourier waves.

In cosmology usually the sub-class of separable low-pass filter approximations are used (Hockney & Eastwood 1988). These filter obey the relation (Marschner & Lobb 1994):

$$W(x, y, z) = W_s(x) \, W_s(y) \, W_s(z) \,, \tag{5.33}$$

where $W_s(x)$ is the separated 1D filter kernel. This separation allows for fast computational speed in performing the 3D convolution (Ferguson 2001).

The simplest filter approximation frequently used in cosmology is the Nearest Grid Point (NGP) kernel. This zero-order kernel provides the simplest and fastest interpolation method (Duan et al. 2003). Each galaxy in the input data is assigned to the value of the nearest sample point in the output data. The nearest neighbor kernel is defined as:

$$W_s(x) = \begin{cases} 1 & \text{for } -0.5 < x \le 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{5.34}$$

The Fourier response of this kernel is a sinc function which has a poor localization and pass-band selectivity. This property typically leads to low-quality interpolated data with blocking effects for signals with high frequency contents like sharp intensity variations or high noise level as might be expected in sampling individual point particles (Duan et al. 2003).

Another commonly used filter is the linear filter. This first-order kernel linearly interpolates between adjacent points of the input data along each dimension (Duan et al. 2003). It is defined as:

$$W_s(x) = \begin{cases} 1 - |x| & \text{for } 0 \le |x| < 1 \\ 0 & \text{for } 1 \le |x| \end{cases} \tag{5.35}$$

In Cosmology this filter is often referred to as the Cloud in Cell (CIC) scheme (Hockney & Eastwood 1988). It is popular for prefiltering as it provides a good tradeoff between filtering quality and computational cost (Duan et al. 2003). Nevertheless, a significant amount of spurious stop-band aliasing components continues to leak into the pass-band, contributing to some aliasing (Wolberg 1997).

The next order kernel is the so called Triangular Shaped Cloud (TSC) kernel which is defined as:

$$W_s(x) = \begin{cases} 0.75 - |x|^2 & \text{for } 0 \le |x| < 0.5 \\ 0.5 \, (1.5 - |x|)^2 & \text{for } 0.5 \le |x| < 1.5 \\ 0 & \text{for } 1.5 \le |x| \end{cases} \,, \tag{5.36}$$

(Hockney & Eastwood 1988). This filter kernel has better stop-band behaviour and is therefore usually preferred to the CIC kernel.

This can be seen in Figure 5.6 where the Fourier transforms of the NGP, CIC and the TSC filter kernel are compared to the ideal low-pass filter. In addition to the stop-band leakage of all filters they all introduce pass-band attenuation, meaning the power of the Fourier modes is suppressed by the filter. This is of special relevance in all applications where one is interested in the correct spectral representation of the signal, in particular in the case of the matter power-spectrum estimation.

This reshaping of the Fourier-space due to the application of an imperfect low-pass filter can be undone by
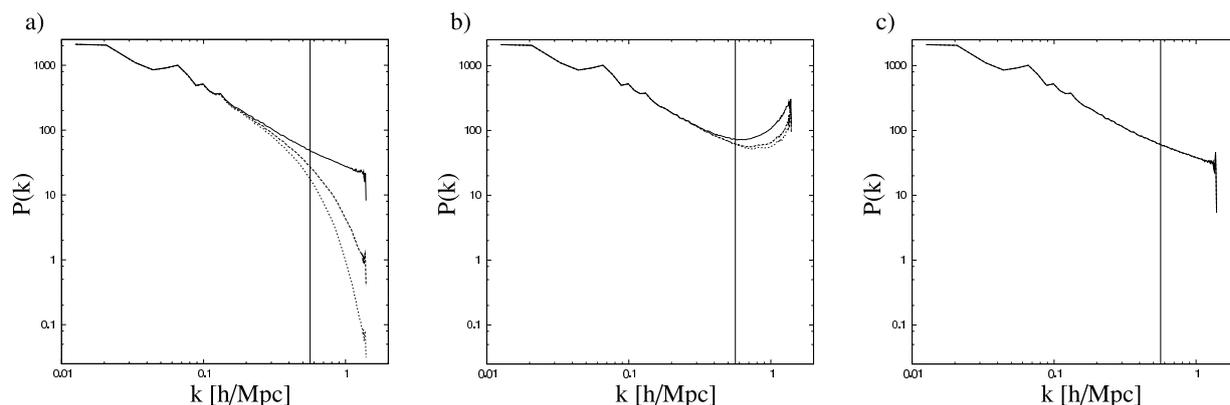
Figure 5.7.: Power spectra calculated from a galaxy mock catalog created by De Lucia & Blaizot (2007) with the three different filter methods NGP (solid line), CIC (long-dashed line) and TSC (short-dashed line). Panel a) displays the power-spectrum obtained without further processing, while panel b) demonstrates the effect of deconvolving with the filter kernel. In panel c) we show the power-spectrum taken using the supersampling technique, prefilterd with CIC kernel (solid line) and TSC kernel (dashed line) which cannot be distinguished visually. The vertical line indicates the critical Fourier mode k_c below which the spectrum usually is trusted.

deconvolving with the Fourier transform of the corresponding filter kernel. This procedure in fact returns the correct power at most Fourier modes, as can be seen in Figure 5.7, where we demonstrate the power spectra obtained after applying these filter approximations to a galaxy mock catalog created by De Lucia & Blaizot (2007). The left panel displays the power-spectrum without deconvolving by the filter kernel while the right shows the same power spectra but deconvolved by the corresponding Fourier transforms of the filter kernels. As it can be seen in Figure 5.7 most of the modes now seem to have correct power, and the different filter approximations agree to each other at the lower modes. Nevertheless, this procedure will also enhance the aliased contribution leaking into the pass-band region, which will severely affect the power-spectrum towards the highest Fourier modes. In general all Fourier modes of the sampled signal will be affected by aliasing. However, natural signal Fourier amplitudes tend to drop off with higher frequency and therefore the sampled signal Fourier modes will be affected less and less towards the lower frequencies.

For this reason the power-spectrum is usually trusted only up to Fourier modes of $k_c = 0.7\pi N/L$ (Cui et al. 2008),in cosmology. Figure 5.7 also demonstrates that the NGP method performs poorly and should not be considered to be used in precision applications.

Another important thing to remark is that although TSC couples 3^3 cells, while CIC only couples only 2^3, and therefore has a broader spatial support, it does not perform much better than the CIC scheme. This is due to the fact, that the overall amplitude of the ideal low-pass filter drops off in space, and hence the main contribution of the corrections will come from cells close to the origin. As the amplitude of the ideal low-pass filter decreases rapidly it requires more and more cells to make as strong corrections as were obtained by only considering the closer cells to the kernel origin. This demonstrates that in order to make further corrections to the filter approximation, one requires to couple more and more cells, while at the same time corrections will be smaller and smaller.

5.7. Supersampling

As already pointed out above, correct sampling is in practice generally not feasible, and some approximation must necessarily be made. However, for most natural signals, like the galaxy distribution, an approximation method can take advantage of the fact that signal content generally declines as the frequency increases. This observation resulted in the rule of thumb, that the matter power-spectrum can be trusted at frequencies below k_c (Cui et al. 2008). The most common anti-aliasing technique in 3D computer graphics, supersampling, takes advantage of this by sampling the signal at a frequency higher than the desired sample rate (Goss & Wu 2000). A low-pass filter is

then applied to the supersamples which attenuates or eliminates the frequency content above a threshold so that the signal resampled at the target rate exhibits fewer aliasing artifacts.

As supersampling methods are very successful in reducing sampling artifacts in 3D computer graphics, we present an adapted supersampling method to be used in cosmological applications.

5.7.1. Super resolution and downsampling

The supersampling method consist of two main steps, the supersampling step, in which the signal is sampled at high resolution, and the downsampling process, in which the high resolution samples are sampled to the target resolution. In our approach we make use of FFTs to allow for pass-band attenuation correction and for fast and efficient calculation of the overall supersampling method. This two stage filtering process is illustrated in Figure 5.5, and consists of the following steps:

1. supersampling: The continuous signal is sampled to a grid with a resolution n times larger than the target resolution. This is achieved by applying the CIC or TSC method to allow for fast and efficient computation of the high resolution samples.

2. downsampling: The high resolution samples are corrected for pass-band attenuation and low-pass filtered. The high resolution low-pass filtered samples are then resampled at the target resolution.

While the supersampling step is in principle identical to the method as described in section 5.6 the downsampling process requires some more explanation.

As already described in section 5.6, using an imperfect filter approximation usually leads to pass-band attenuation which has to be corrected.

This is achieved by applying a FFT to the supersampled signal, and then dividing by the Fourier transform of the according filter approximation. Reducing the supersampled signal to the final resolution requires to low-pass filter the supersamples and sample them again at lower resolution.

In our approach low-pass filtering and downsampling is done in one step. The low-pass filter step can easily be achieved in Fourier-space by discarding all frequencies higher than a given threshold. Introducing the ideal low pass-filter in discrete Fourier-space as:

$$\hat{W}_k = \begin{cases} 1 & \text{for } -\left(\frac{N_{SS}}{2n} - 1\right) \leq k \leq \frac{N_{SS}}{2n} \\ 0 & \text{otherwise} \end{cases}, \tag{5.37}$$

with N_{SS} being the number of supersampling cells, yields the low-pass filtered supersampled signal g_j as:

$$\begin{aligned} g_j &= C_{SS} \sum_{k=-\left(\frac{N_{SS}}{2}-1\right)}^{\frac{N_{SS}}{2}} \hat{W}_k \, \hat{f}_k \, e^{2\pi jk\frac{\sqrt{-1}}{N_{SS}}} \\ &= C_{SS} \sum_{k=-\left(\frac{N_{SS}}{2n}-1\right)}^{\frac{N_{SS}}{2n}} \hat{f}_k \, e^{2\pi jk\frac{\sqrt{-1}}{N_{SS}}} , \end{aligned} \tag{5.38}$$

where $\hat{f}_k$ is the pass-band attenuation corrected Fourier transform of the supersampled signal and C_{SS} is the FFT normalization constant according to the super resolution Fourier transform. Downsampling can now easily be achieved by introducing the number of cells of the target resolution $N_{DS} = N_{SS}/n$ and using $j = n \, j'$, which simply results from the fact that the grid spacing of the target resolution grid is n times smaller than the super resolution grid spacing.

$$g_j = C_{SS} \sum_{k=-\left(\frac{N_{DS}}{2}-1\right)}^{\frac{N_{DS}}{2}} \hat{f}_k \, e^{2\pi j'k\frac{\sqrt{-1}}{N_{DS}}} . \tag{5.39}$$

Sampling the supersampled and low-pass filtered signal to the target resolution grid can be done by applying the Kronecker delta function $\delta^k_{jj'}$ to yield the downsampled signal:

$$
\begin{aligned}
g_{j'} &= C_{SS} \sum_{k=-\left(\frac{N_{DS}}{2}-1\right)}^{\frac{N_{DS}}{2}} \hat{f}_k\, e^{2\pi j' k \frac{\sqrt{-1}}{N_{DS}}} \\
&= C_{DS} \sum_{k=-\left(\frac{N_{DS}}{2}-1\right)}^{\frac{N_{DS}}{2}} \frac{C_{SS}}{C_{DS}} \hat{f}_k\, e^{2\pi j' k \frac{\sqrt{-1}}{N_{DS}}} ,
\end{aligned}
\tag{5.40}
$$

with C_{DS} being the FFT normalization according to the target resolution.

Hence, equation (5.40) provides an easy prescription for the low-pass filtering and downsampling process. Downsampling can be achieved by sorting the Fourier modes of the supersampled signal up to the Nyquist mode of the target resolution in the target resolution grid in Fourier-space, rescale it to the correct Fourier normalization and apply an inverse FFT to obtain the real-space representation for the signal.

The interpretation of this process is easy. As the FFT assumes the signal to be periodic and to consist of a finite superposition of Fourier waves, the resolution of the supersampled signal can simply be reduced by discarding the high frequency waves, resulting in a smoother real-space representation of the sampled signal. The result of this procedure is a sampled signal at target resolution which is greatly cleaned from aliasing effects.

The ability of this supersampling method is demonstrated with Figure 5.7. Here we used a super resolution factor $n = 2$ and calculated the power-spectrum from the supersampled signal. For the supersampling step we applied the CIC and the TSC method. As can be seen in Figure 5.7 the two results for CIC and TSC cannot be distinguished. Also it is worth noticing that the aliasing contribution had been corrected far beyond the usual limit of trust $k_c = 0.7\pi N/L$.

In Figure 5.8 we contrast the supersampling method to the TSC method by comparing slices through a sampled 3D mock galaxy catalogue in a 500 Mpc box which has been sampled regularly at 128^3 pixels. At first glance, visually there is not much difference between the two samples. However, a closer inspection of the high density regions reveals that the density field obtained with the pure TSC method is more blocky, meaning it can have sudden density jumps from one pixel to the next. In the case of the supersampled density field these high density pixels are surrounded by pixels with varying densities. In this way the sharp edges are softened, as should be expected by an anti-aliasing technique. Additionally, one can observe, that the filaments are better connected in the case of the supersampled density field. This is due to the additional super resolution information which was taken into account in the supersampling process.

5.8. Discussion and conclusion

In this chapter we reviewed and discussed the implications of data acquisition and the use of FFT techniques in cosmological signal processing applications. As described in section 5.2 the application of FFTs requires the signal to be discrete in real and Fourier-space. As natural signals are in general continuous in either spaces, they must be approximated by discretized representations with the constrained of conserving as much information of the true natural signal as possible. According to Shannon's theorem this can be achieved by low-pass filtering and sampling the continuous signal at discrete positions.

However, we also demonstrated that Shannon's theorem is a necessary but not sufficient requirement, for processing a signal via FFTs or in more general DFTs. The application of FFTs or DFTs additionally requires the signal to be periodic on a bounded spatial domain, or in other words, the signal must be decomposable into a finite amount of Fourier waves. Thus, the application of FFT or DFT techniques implicitly assumes the discreteness of Fourier-space which is generally not given by natural signals. This introduces an additional filtering procedure for the true continuous Fourier modes, as shown in section 5.4.1.

In section 5.4.3 we demonstrated that, due to discretizing, the physical property of being a positive density field is not reflected by the discretized representation of the continuous galaxy distribution. This is due to the convolution with the ideal low-pass filter kernel which is a sinus cardinal function and thereby not positive definite. The loss of physicality is thus an expression of the loss of information due to discretizing the continuous signal, and hence is a
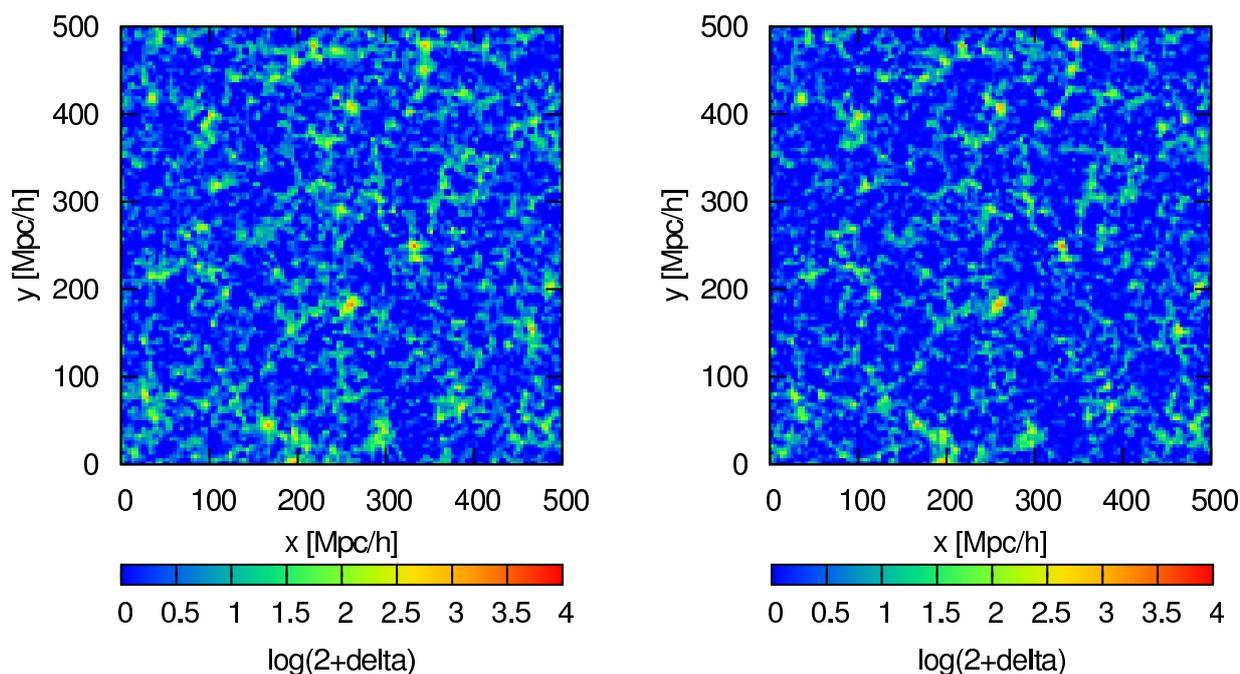
Figure 5.8.: Comparison of slices through sampled mock galaxy distribution using the supersampling method (left panel) and the standard TSC method (right panel).

fundamental problem of any data processing procedure.

In addition to this fundamental problem, we discussed that ideal sampling is in general not applicable to real world problems.

Signal processing in cosmology or any other field is a technologically challenging problem governed by the requirement of computational feasibility of the sampling procedure. As ideal sampling is in general computationally too expensive to be used for real world signal processing applications one usually introduces low-pass filter approximations to allow for faster computation for the expense of introducing sampling artifacts like aliasing.

Due to these artifacts further processing and analysing the data may be erroneous (see e.g. Cui et al. 2008). Especially, iterative signal reconstruction processes, which utilize convolutions, deconvolutions and other signal processing operations, may enhance or distribute false aliased power into regions of the signal formerly unaffected by aliasing.

The aim for signal processing technologies therefore is to find low-pass filter approximations, which sufficiently suppress these sampling artifacts, while at the same time still being computationally far less expensive than the ideal sampling procedure. The Digital Signal Processing literature provides plenty of possible approaches (see e.g. Smith (2002)) which could be introduced to various cosmological signal processing problems.

Recently, many authors presented new methods to reduce aliasing effects in cosmological power-spectrum estimation due to imperfect filter approximation Jing (2005), Cui et al. (2008). However, these method either rely on post processing correction or on increasing the spatial support of the ideal low-pass filter approximation. Enlarging the spatial support of the filter kernel will lead to better approximations of the ideal low-pass filter, for the expense of being computationally less efficient. This approach is well known in the DSP literature, which favors the windowed sinc functions as optimal approximations to the ideal low pass filter (Theußl et al. 2000, Duan et al. 2003, Marschner & Lobb 1994, Smith 2002).

However, these methods are either computationally too demanding or not applicable for iterative data analysis algorithms as presented in the remainder of this thesis.

For this reasons we proposed a new supersampling technique frequently applied as an anti-aliasing technique in 3D computer graphics (Wolberg 1997, Goss & Wu 2000). As described in section 5.3.1, aliasing is a result

of sampling, or more specifically, the lack of a sufficient sampling rate. Supersampling, as the name suggests, solves the aliasing problem by taking more samples than would normally be the case with usual particle assignment schemes as CIC or TSC. By taking more samples at sub-pixels, we are able to more accurately capture the details of the natural continuous signal. The target pixel value is then obtained by averaging the values of the sub-pixel reducing the aliasing edge effects in the signal. Hence, supersampling reduces aliasing by band-limiting the input signal and exploiting the fact that the signal content generally declines as the frequency increases.

The supersampled signal still contains artifacts due to aliasing, but the artifacts are less prominent than if the signal was sampled directly at the target rate. Nevertheless, there are two problems associated with supersampling. The first problem is, as already mentioned, that the newly designated Nyquist frequency of the superresolution samples continues to be fixed (Wolberg 1997). Hence, there will always be sufficiently higher frequencies that will alias. The second problem is cost of memory (Wolberg 1997). To sample a three dimensional signal at double resolution requires eight times more memory. On the other hand, if memory is no issue, the supersampling method is computationally less expensive than introducing a next order low-pass filter approximation, while at the same time providing better aliasing suppression. In addition, the supersampling procedure, as proposed in this chapter, incorporates pass-band attenuation corrections, which otherwise would have to be corrected for in an additional separate step, when applying just a low-pass filter approximation.

This supersampling method is especially useful for Wiener-filter reconstructions of the cosmic density field, since it greatly reduces sampling artifacts as required for signal deconvolutions. In particular, it was employed to prepare Sloan Digital Sky Survey data for the application of a Wiener-filter (Kitaura et al. 2009). Some of the results taken from this work are presented in Fig. 5.9.

The supersampling technique can be understood as being complementary to the approach of finding better low-pass filter approximations, since the combination of both methods will naturally increase the quality of the sampled signal.

Nevertheless, the problem of data acquisition and data processing via FFT techniques is a complex one. Beside being subject to fundamental issues of discritizing physical quantities, it mainly is a technologically challenging problem, which requires to efficiently apply a low-pass filter to an observed continuous signal. The optimal solution for each individual application may also vary from case to case, as it can be dependent on the individual properties of the underlying signal and scientific application (e.g high accuracy power-spectrum estimation).
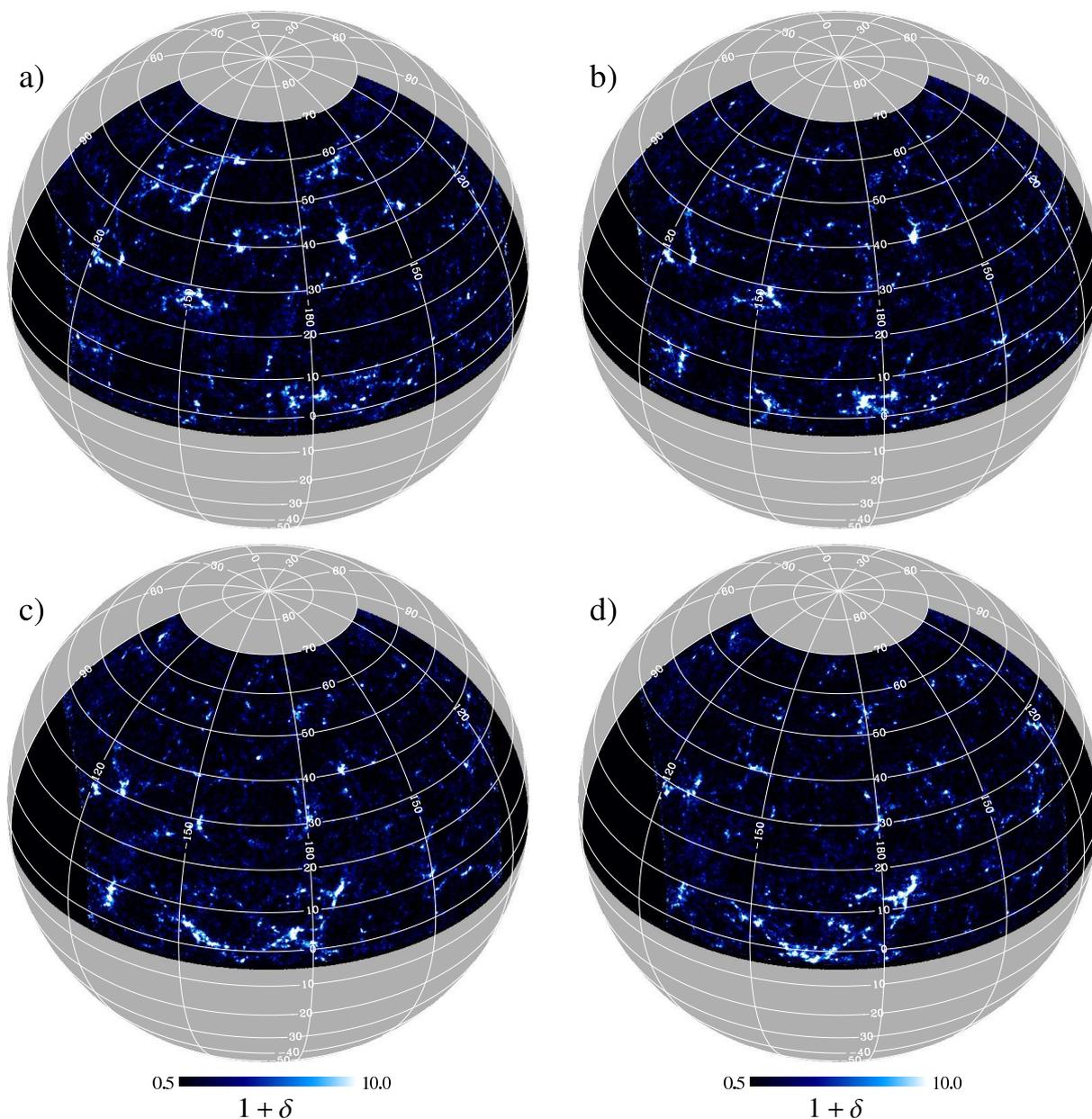
Figure 5.9.: Different projections of the estimated three dimensional Wiener filtered density field on the sky. The plots depict different radial slices with thickness of 20 Mpc around the Sloan Great Wall. Different panels correspond to different co-moving distances d from the observer: $d = 290 - 310$ Mpc (panel a), $d = 300 - 320$ Mpc (panel b), $d = 310 - 330$ Mpc (panel c) and $d = 320 - 340$ Mpc (panel d). Courtesy of Francisco S. Kitaura, for details see Kitaura et al. (2009).

6. Bayesian inference

The world is full of obvious things which nobody by any chance ever observes.

Sir Arthur Conan Doyle (The Hound of the Baskervilles)

Abstract

In this chapter basic ideas and motivations for Bayesian inference with respect to cosmological data analysis are presented. Further, the role of the prior in Bayesian data analysis is discussed and its use is justified by referring to the 'No free lunch theorems' and demonstrating that prior information is required for any empirical learning process. In addition this chapter contains a brief introduction to the subject of Markov Chain Monte Carlo methods, to the degree as is required for this thesis.

6.1. Introduction

To further our knowledge about the real physical world we always have to perform either experiments or conduct observations to study the behavior of physical systems. The obtained measurements are then used twofold, either to identify new unknown phenomena or to check the validity of existing theories by means of the observations. For this reason, precise data analysis is of crucial importance to any natural science which tries to establish a scientific model about the world we are living in. The process of data analysis thus is meant to extract optimal information, about the physical system to be studied, by abstracting all uncertainties introduced by the measurement or observation. In the particular case of galaxy observations, as discussed in this thesis, there are several classes of uncertainties and systematics, which can hamper the detection of an underlying signal. These uncertainties and systematics are either intrinsic to the nature of the observable itself or are generated in the process of observation. Here we are going to list some of the uncertainties frequently encountered in galaxy redshift observations.

1. Cosmic variance:

   In our present understanding, the observable universe is the consequence of the interplay of some physical laws, such as quantum mechanics and general relativity. However, some of these processes are completely random, like the generation of initial density fluctuations. For this reason, the matter field or galaxy distribution in the Universe can only be described statistically, which prevents a derivation from first principles. Our Universe hence represents a specific realization of such a random process. As we have only one such realization at our disposal, it is difficult to draw statistical conclusions about its significance. This intrinsic uncertainty is referred to as cosmic variance, and is the direct consequence of observing only one universe. (For a more detailed discussion in the case of power-spectrum estimation see 7.6)

2. Galaxy bias:

   Luminous matter, such as galaxies, does not trace the mass in the Universe directly. Various observations and theories indicate a difference in spatial distribution between luminous objects and dark matter. Confronting theoretical model predictions for the dark matter distribution, hence requires to relate the observed luminous matter to the density field of dark matter. This bias is related to the complicated processes of galaxy formation, which are non-linear, and probably non-local. Various studies indicate that the bias factor is not universal, as it depends on galaxy type, formation time and redshift (see e.g. Xia et al. 1987, Boerner et al. 1991, Norberg et al. 2002, Li et al. 2006, Meneux et al. 2008). Several other studies even suggest a stochastic behavior of the non-linear galaxy bias, which would only allow for a statistical relation to the underlying dark matter density field (Pen 1998, Dekel & Lahav 1999, Tegmark & Bromley 1999).

3. Redshift space distortions:

The proper physical distance from the observer to a galaxy cannot be measured directly. One rather has to rely on redshift measurements as distance indicators. If a galaxy would be at rest with respect to the Hubble flow then the exact co-moving distance could be calculated via the redshift distance relation as discussed in chapter 3. However, the co-moving motions of galaxies are usually superimposed by systematic (linear) and random (nonlinear) peculiar motions in the gravitational potentials of their host halos. These peculiar velocities yield additional Doppler effects in the redshift measurement making the exact localization of the galaxy impossible. In the absence of knowledge about these random peculiar velocities these redshift distortion effects introduce statistical uncertainties on the exact location of galaxies in addition to any spectroscopic measurement uncertainty. (For a detailed overview of redshift distortions and their relevance for cosmography projects see the discussion in 9.3.4.)

4. Noise:

A galaxy redshift survey consists of a set of discrete objects at certain positions in three dimensional configuration space. This set of galaxies represents a discrete sample of the underlying continuous density field, and hence introduces uncertainties in the form of noise. A frequent assumption is that this noise contribution can be modeled as Poisson noise also referred to as shot noise. This shot noise is of statistical nature and is particularly important in regions, which have only been sparsely sampled by galaxies, such as voids. (For a more detailed discussion see 7.4.2)

5. Selection effects and survey geometry:

Any observation introduces systematic effects due to the characteristics of the detector. Therefore, the particular response of a detector to a specific signal has to be modeled, in order to infer information from the observation. For galaxy redshift surveys, this requires to model the characteristics of the telescope and the survey strategy. Since galaxies only have finite brightness, fainter galaxies have lesser probability of being detected the more distant they are from the observer. This effect can be modeled by a radial selection function, which takes into account that the amount of observed galaxies with a given luminosity generally declines with distance. The angular selection function or survey geometry defines the regions in the sky which have been observed. Usually, deep galaxy redshift surveys have no 4π-geometry, but observe only certain patches on the sky. In addition, one has to model the exposure or completeness and also weather conditions, as it influences the seeing of the telescope. The above effects are the main cause for the systematic uncertainty of large scale power-spectrum estimation, as they distort the power-spectrum shape, and drastically decreases the visibility of features, such as the baryonic wiggles (Tegmark 1995, Ballinger et al. 1995, Cole et al. 2005). Ignoring these effects in density estimations will yield strongly biased estimates. Therefore, in general careful treatment of these effects is required (Zaroubi et al. 1995).

The given examples of statistical and systematic uncertainties clarify that the task of identifying new phenomena and testing theories with observations is by far not trivial. In addition to the observational uncertainties, as listed above, there are the systematic uncertainties of the data acquisition procedure as discussed in the previous chapter. All these effects must be carefully taken into account when drawing conclusions from observations. Already in the previous chapter we demonstrate that the procedure of data acquisition does generally not permit unique reconstruction of the measured signal. This statement can be strengthened, as the statistical properties of some observational uncertainties will generally not allow for unique inversion of the observational process. For this reason a data analysis cannot yield a single and unique answer, but rather has to provide a set of possible solutions, all consistent with the observations within the limits of statistical uncertainty. Hence, the means of data analysis have to be statistical. Below, we will give an introduction to Bayesian statistics to the degree required for this work.

6.2. Conventional vs. Bayesian statistics

When discussing statistical data analysis one has to review in principle two different viewpoints of statistics in metrology, the conventional (frequentist) and Bayesian approach. These two viewpoints crucially differ in the interpretation of probability, when using probability distributions, and their consequences for determining characteristic values and testing hypotheses. The first and most important epistemological difference is the way of treating probabilities. Although conventional as well as Bayesian statistics have their foundations in probability theory they differ

fundamentally in the understanding of the concept of the probability $\mathcal{P}(A)$ with ($0 \leq \mathcal{P}(A) \leq 1$) of a random event A, which can or cannot occur in a well defined experiment.

In conventional or frequentist statistics one defines the probability $\mathcal{P}(A)$ as the relative frequency with which the event A occurs in repeated experiments with nominally identical test conditions. In Bayesian statistics, on the other hand, the probability $\mathcal{P}(A)$ represents the degree of believe that A will occur in an experiment under consideration of all available information, before the experiment has been conducted. This is the classical concept of probability as introduced by Jakob Bernoulli and Pierre-Simon Laplace. It is the same kind of probability by which a dice player assumes that the probability to roll a "one" is 1/6 for an ideal dice.

Statements of conventional statistics are always of the inner form: "Event A occurs with probability p when hypothesis H is true". Translated to a measurement this means: " The measured value x occurs with a probability p if the measurand X has the true value ξ." This means that conventional statistics can only answer questions of the following kind: "Given the true value ξ of the measurand X, what is the probability distribution of the measured values x?" In terms of conditional probabilities

Bayesian statistics, in contrast, also allows for statements of the kind:" Given the observed event A the hypothesis H may be true with probability q. For an actual measurement this means:"Given the measured value x the measurand X has the true value ξ with probability q." Therefore, Bayesian statistics also answers the question: "Given the observed measured value x, what is the probability that the true value of X is ξ?" For this reason, Bayesian statistics answers the underlying question to every measurement problem, of how to estimate the true value of the measurand from observations, while conventional statistics does not.

6.3. Conditional probabilities

The two approaches can be formulated mathematically by introducing the conditional probability $\mathcal{P}(A|B)$, which describes the probability of event A occurring with the additional condition that event B already occurred or will occur. To conditional probabilities applies the famous Bayesian theorem, which relates conditional probabilities by

$$\mathcal{P}(A|B)\mathcal{P}(B) = \mathcal{P}(B|A)\mathcal{P}(A) = \mathcal{P}(A, B) . \tag{6.1}$$

This means that the probability for event A given event B times the probability for B is equal to the product of the probability for event B given event A with the probability for A. In equation (6.1), $\mathcal{P}(A, B)$ denotes the probability for the event "A and B". Two events are independent of each other if

$$\mathcal{P}(A|B) = \mathcal{P}(A) \quad \text{and} \quad \mathcal{P}(B|A) = \mathcal{P}(B) \tag{6.2}$$

holds. If there exist several events B_j, of which only exactly one will occur in an experiment, the probability for event $\mathcal{P}(A)$ can be obtained by marginalizing over all B_j

$$\mathcal{P}(A) = \sum_j \mathcal{P}(A|B_j)\mathcal{P}(B_j) . \tag{6.3}$$

Translated to an actual measurement problem $\mathcal{P}(x|\xi, y)$ denotes the probability for measuring the value x given the true value ξ of the measurand and additional constraints y of the experiment. In contrast, $\mathcal{P}(\xi|x, y)$ is the probability for the true value ξ given the actual measured value x and the additional constraints y of the experiment. These definitions are independent from the different statistical approaches but demonstrate the difference between the conventional and Bayesian statistics. Conventional statistics provides probabilities of the form $\mathcal{P}(x|\xi, y)$ while Bayesian statistics makes statements about $\mathcal{P}(\xi|x, y)$. In an actual experiment the true value ξ of the measurand is generally unknown and the true task of an experiment is to provide statements about $\mathcal{P}(\xi|x, y)$, the probability that the measurand has the true value ξ given the measured value x and additional constraints y of the experiment. In summary, the fundamental difference between the two approaches are, that conventional statistics assumes the measurement to be uncertain and the measurand known, while Bayesian statistics assumes the observation to be known and the measurand uncertain.

6.4. The prior distribution and the mechanism of empirical learning

According to the Bayes theorem, given in equation (6.1), $\mathcal{P}(\xi|x,y)$ can be described by separating the information obtained from the current experiment from all other anticipated information on the measurand. Then $\mathcal{P}(\xi|x,y)$ can be written as a product

$$\mathcal{P}(\xi|x,y) = \mathcal{P}(\xi)\frac{\mathcal{P}(x|\xi,y)}{\mathcal{P}(x)} \, , \tag{6.4}$$

where $\mathcal{P}(x|\xi,y)$ is the likelihood, which describes the process of assimilating new information from observations, $\mathcal{P}(\xi)$ is the prior incorporating our a priori knowledge about the measurand and $\mathcal{P}(x)$ is often referred to as the evidence, which normalizes the probability distribution. Since in a Bayesian framework probabilities describe states of knowledge, equation (6.4) describes a learning procedure in which the prior probability is updated in the light of new experiences or observations.

Bayesian inference has always been criticized to be subjective due to the notion of such a prior probability which needs to be specified prior to any analysis (Gelman 2008). In order to ensure objectivity, critics often demand to use non-informative priors, i.e. flat priors. This however, can be regarded as inappropriate from a Bayesian point of view, especially when dealing with sparsely sampled data. In addition, it is not always clear in respect to which coordinates the flat prior should be formulated. Introducing a simple coordinate transformation will often reveal an informative prior in such a new coordinate system. For an actual inference problem a perfectly flat prior requires that all mathematically possible solutions to a problem have equal probability of occurring in an experiment. In this situation, the 'No free lunch theorems' in the field of mathematical optimization state that on average no search strategy can be expected to perform better (or worse) than any other (Wolpert & Macready 1997). Therefore, a general-purpose universal optimization strategy is theoretically impossible (Ho & Pepyne 2002). The only way one search strategy can outperform another is if it is specialized to the specific problem under consideration. Therefore, one can conclude that optimal search strategies should be adapted to specific classes of problems by incorporating prior knowledge, rather than using general search strategies such as genetic algorithms or simulated annealing (Ho & Pepyne 2002). The Bayesian prior marks certain regions in solution space, and therefore renders inferences with prior information more efficient for a given class of problems than those without.

Also note, that learning without prior knowledge is impossible. As mentioned above, in Bayesian terms probabilities describe our knowledge states which can be updated by new experiences or observations. Learning, therefore, can be regarded as continuous update of our prior knowledge via the prescription given in equation (6.4).

A simple example of measuring distances to some objects demonstrates that Bayes theorem truly describes an empirical learning procedure. Let us consider there exists some measuring device, such as a ruler or laser ranger, which returns the true value of the distance plus some additive normally distributed random noise component. Prior to any measurement the knowledge state must not prefer certain values in solution space, in this case $\mathbb{R}$. Therefore, each value of $\mathbb{R}$ must have equal probability of occurring in a measurement. Since, probabilities have to be normalized to give unity when integrated over, it is clear that the flat prior is also an ill posed prior in the sense, that it cannot be properly normalized. For this reason, flat priors are also frequently referred to as improper priors. Updating the knowledge state with new information from a measurement can then be described by

$$\mathcal{P}(\xi|x_1,\sigma) = \mathcal{P}(\xi)\frac{\mathcal{P}(x_1|\xi,\sigma)}{\mathcal{P}(x_1)} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_1-\xi)^2}{2\sigma^2}} \, . \tag{6.5}$$

Through incorporation of the measurement x_1 the knowledge state evolved from a flat distribution towards a Gaussian distribution around the measured value x_1. From a Bayesian point of view it seems logical to use this new knowledge state as prior information when inferring the true value for the distance ξ from a second measurement x_2. One then yields the new posterior distribution

$$\mathcal{P}(\xi|x_1,x_2,\sigma) = \mathcal{P}(\xi)\frac{\mathcal{P}(x_1|\xi,\sigma)}{\mathcal{P}(x_1)}\frac{\mathcal{P}(x_2|\xi,\sigma)}{\mathcal{P}(x_2)} = \frac{\sqrt{2}}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{2\left(\xi-\frac{1}{2}(x_1+x_2)\right)^2}{\sigma^2}} \, , \tag{6.6}$$

using that $\mathcal{P}(x_2|x_1,\xi,\sigma) = \mathcal{P}(x_2|\xi,\sigma)$ due to statistical independence of the measurements. Note, that in this inference step, the Gaussian posterior of the previous step was used as a prior distribution. In this fashion, successive
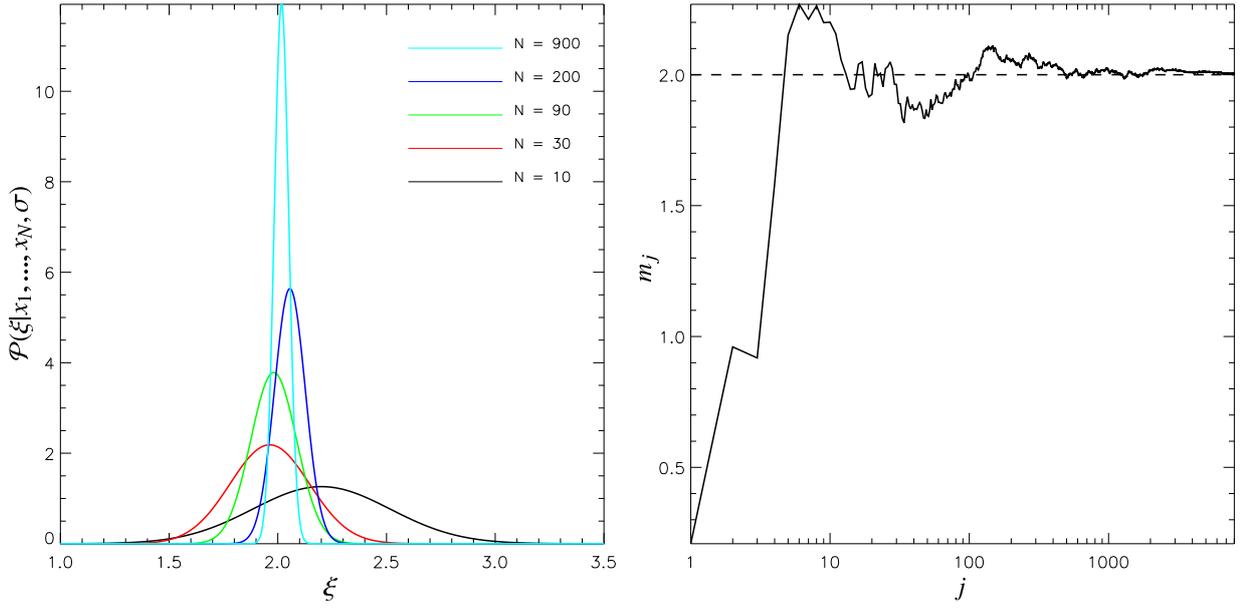
Figure 6.1.: Evolution of the distance posterior in the process of successive assimilation of new observational information. The posterior distributions tend towards a Dirac delta distribution around the true value $\xi = 2$.

accumulation of knowledge states yields the Gaussian posterior distribution

$$\mathcal{P}(\xi|x_1, ..., x_N, \sigma) = \mathcal{P}(\xi) \prod_{i=1}^{N} \frac{\mathcal{P}(x_i|\xi, \sigma)}{\mathcal{P}(x_i)} \quad = \quad \frac{\sqrt{N}}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{N\left(\xi - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2}{\sigma^2}}$$

$$= \quad \mathcal{P}(\xi) \frac{\mathcal{P}(x_N|\xi)}{\mathcal{P}(x_N)} \, , \tag{6.7}$$

where the last line is given by the prior $\mathcal{P}(\xi) \propto \frac{\sqrt{\tilde{\sigma}^2 + \sigma^2}}{\sqrt{2\pi}\,\tilde{\sigma}^2\,\sigma^2} e^{-\frac{1}{2} \frac{(\xi - \tilde{\mu})^2}{\tilde{\sigma}^2}}$, the likelihood $\mathcal{P}(x_N|\xi) \propto e^{-\frac{1}{2} \frac{(\xi - x_N)^2}{\sigma^2}}$ and the evidence $\mathcal{P}(x_N) \propto e^{-\frac{1}{2} \frac{(\tilde{\mu} - x_N)^2}{\tilde{\sigma}^2 + \sigma^2}}$. Here, N is the number of conducted measurements. In the last line of of equation (6.7) we rewrote the posterior distribution in the familiar Bayesian form as the product of a Gaussian prior, with mean $\tilde{\mu} = 1/(N-1) \sum_{i=1}^{N-1} x_i$ and the variance $\tilde{\sigma} = \sigma/\sqrt{N-1}$, and a Gaussian likelihood normalized by the evidence. Especially, the prior variance $\tilde{\sigma}$ demonstrates that with increasing N the knowledge about the true distance ξ gets refined. In Fig. 6.1 some successive posterior distributions are presented to demonstrate the evolution of the knowledge state. With an increasing number of measurements the posterior distribution of the distance to the object ξ gets sharper and tends towards a Dirac delta distribution. The above example demonstrates that previously obtained information can greatly improve the inference process for a single measurement. If on the other hand, we had ignored previously obtained knowledge in each inference step, i.e. assuming a flat prior, our initial posterior distribution given in equation (6.5) would not at all evolve and learning would not have taken place. The successive posterior distributions in Fig. 6.1 also nicely demonstrate, that inference will yield different results, depending on the current knowledge state, especially in cases when little data is available. This can be demonstrated by estimating the maximum a posteriori value for each of the above learning steps as depicted in Fig. 6.1. The maximum a posteriori distance m_j is given by

$$m_j = \frac{\sum_{i=1}^{j} d_i}{N} \, , \tag{6.8}$$

which is obtained by extremizing equation (6.6). The maximum a posteriori value clearly approaches the true value of the distance $\xi = 2$ for high numbers of experiments j. Figure (6.1), also demonstrates that the results obtained

can strongly depend on the knowledge state j especially in cases when only little observational data is available. The results are therefore subjective, since two experimenters of knowledge states $j - k$ and j would generally not agree on the parameters $\tilde{\mu}$ and $\tilde{\sigma}$ for their prior distributions. To the author the criticism of subjectivity in the above example sounds absurd, since experimenters who have better knowledge about the system to study also deserve to get better, i.e. more precise, results compared to those who know less or even nothing. This is clearly one of the rationale behind strict data policies of modern cosmological surveys, who want to press home their advantage. This empirical learning procedure also clarifies, that prior to any scientific discussion there has to be a transfer of knowledge, i.e. exchange of prior distributions, in order to equalize different knowledge states. In practice, in science, this is achieved by publishing articles and books or releasing data. In the light of the above example, different prior states do not really pose a threat to scientific progress, since different researchers will sooner or later agree, once a sufficient amount of information has been gathered. Together with the 'No free lunch theorems' and the notion that learning is not possible without prior information, the use of an informative prior distribution is nothing to demonize in general.

The above learning procedure is purely experience or data driven, and thus will always yield satisfactory results for any natural science in which theories are tested against observations. Therefore, a problem with Bayesian inference in natural sciences only appears when prior information is used which does not thoroughly reflect the actual knowledge state as proposed by all gathered data on the system to be studied. This only happens when using prior information derived from purely theoretical considerations, which may not necessarily agree with the true behavior of the natural system. In this context, the criticism of Bayesian methods being subjective should be rephrased as the 'fear of wrong assumptions' in the inference process. Using a wrong prior assumption hence is not a shortcoming of the Bayesian method itself but is a mistake of the researcher conducting such an analysis. However, the advantage of Bayesian methods is to explicitly state the prior assumptions and therefore clearly exposing them to scientific criticism. To summarize, the 'No free lunch theorems' state that if any inference method or search algorithm is to perform better than any other search method then it necessarily has to be specialized to a specific class of problems through the incorporation of prior information. Hence, prior information is required for any learning process.

6.5. Markov Chain Monte Carlo methods

As discussed in the previous sections, in Bayesian statistics the object of interest is the posterior distribution as it encodes our entire knowledge about the system to study and all statistical inference can be deduced by reporting appropriate statistical summaries. Representing such a posterior distribution, especially in high dimensions, however, poses serious numerical and technological challenges. For this reason, the breakthrough of Bayesian methods had to wait until the 1980s and was mostly owed to the development of computer hard- and software (Wolpert 2004). In particular, Markov chain Monte Carlo (MCMC) methods removed many of the computational integration problems previously encountered with Bayesian methods. These MCMC techniques, especially the Metropolis algorithm, have been placed among the ten algorithms that have had the most influence on the development of science and technology in the 20th century (Beichl & Sullivan 2000). In the following, the basic ideas and techniques of MCMC methods will be presented (for a more complete introduction to MCMC techniques and their applications to Bayesian inference see e.g. Neal 1993, Andrieu et al. 2003, Robert & Casella 2005, Gamerman & Lopes 2006).

6.5.1. The Monte Carlo principle

The basic idea of Monte Carlo simulations is to approximate the target probability distribution $\mathcal{P}(\xi)$, defined on a high dimensional space χ, by a set of N independent and identically distributed (i.i.d.) samples $\{\xi^{(i)}\}_{i=1}^{N}$ drawn from this target distribution (Andrieu et al. 2003). The set of N samples can then be used to approximate the probability density with the following point particle distribution

$$\mathcal{P}_N(\xi) = \frac{1}{N} \sum_{i=1}^{N} \delta^D(\xi - \xi^{(i)}), \tag{6.9}$$

where $\delta^D(\xi)$ is the Dirac delta distribution. Given such a representation of the target density any statistical summary can be approximated with tractable sums, rather than integrals, as

$$I_N(f) = \frac{1}{N} \sum_{i=1}^{N} f(\xi^{(i)}) \xrightarrow[N \to \infty]{\text{a.s.}} I(f) = \int_{\chi} f(\xi) \mathcal{P}(\xi) \, d\xi. \tag{6.10}$$

This means that the estimate $I_N(f)$ is unbiased and will almost surely (a.s.) converge to $I(f)$ by the Strong Law of Large Numbers (Andrieu et al. 2003, Robert & Casella 2005). In principle, the integral in equation (6.10) could also be estimated by deterministic integration methods. However, the advantage of the Monte Carlo integration arises from the fact that it positions the samples densely in regions of high probability, and sparsely in regions of low probability. In this fashion, the Monte Carlo integrator adapts to the specific geometry of the problem under consideration, and preferably maps out those regions which predominantly contribute to the integral. Therefore, the N samples allow also for the determination of the maximum of the target distribution $\mathcal{P}(\xi)$ which according to Andrieu et al. (2003) can be obtained by

$$\tilde{\xi} = \underset{\xi^{(i)}; i=1,\dots,N}{\arg \max} \mathcal{P}(\xi^{(i)}). \tag{6.11}$$

Especially when dealing with problems of high dimensionality, as discussed in this thesis, Monte Carlo methods easily outperform any other numerical integration method known so far. However, the success of the method strongly depends on the ability to efficiently draw random samples from the target distribution $\mathcal{P}(\xi)$. For standard form distributions, e.g. normal distributions, this task can be solved straightforwardly. If this is not the case, more sophisticated Markov Chain Monte Carlo methods have to be employed.

6.5.2. Markov Chain Monte Carlo algorithms

As stated above, approximations of integrals of the type given in equation (6.10) require a sampled representation of the target distribution $\mathcal{P}(\xi)$. Such representations can be efficiently obtained with the aid of MCMC algorithms which generate samples $\xi^{(i)}$ while exploring the state space χ via a Markov chain mechanism (Andrieu et al. 2003). A Markov chain is a discrete random process, generating a sequence of random variables $\{\xi^{(i)}\}_{i=1}^{N}$, which obeys the Markov property. This means, that the probability distribution at the next step only depends on the current step, but is conditionally independent of all previous steps. Then the sequence of random variables $\{\xi^{(i)}\}_{i=1}^{N}$ is Markovian if the following holds

$$\mathcal{P}\left(\xi^{(i+1)}|\xi^{(0)}, \xi^{(1)}, \dots, \xi^{(i)}\right) = \mathcal{P}\left(\xi^{(i+1)}|\xi^{(i)}\right) = \mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right), \tag{6.12}$$

for all i, for all $\xi^{(0)}, \xi^{(1)}, \dots, \xi^{(i)}$ and $\xi^{(i+1)} \in \chi$. Here, $\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)$ is the transition probability of jumping from the current state i to the following state $i+1$. Then the probability for the $\xi^{(i+1)}$ at state $i+1$ can be obtained by marginalizing over all $\xi^{(i)}$ of the current step i

$$\mathcal{P}_{i+1}\left(\xi^{(i+1)}\right) = \sum_{\xi^{(i)}} \mathcal{P}_i\left(\xi^{(i)}\right) \mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right). \tag{6.13}$$

Therefore, given the initial probability $\mathcal{P}_0\left(\xi^{(0)}\right)$ the behavior of the chain is determined at all times (Neal 1993). Within the scope of this thesis, particularly invariant, or stationary, distributions over the states of the Markov chain are of importance, since they persist forever once reached. Formally speaking, a distribution given by the probabilities $Q\left(\xi^{(i+1)}\right)$ is stationary if, for all i,

$$Q\left(\xi^{(i+1)}\right) = \sum_{\xi^{(i)}} Q\left(\xi^{(i)}\right) \mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right). \tag{6.14}$$

In the case of Bayesian inference, one seeks to construct Markov chains where the invariant distribution $Q(\xi)$ is given by the according posterior density. Then all obtained samples will be samples from the correct posterior distribution.

Constructing such a Markov chain can most easily be achieved by requiring the chain to be homogeneous, i.e. to satisfy detailed balance. This means, that the transition from state $\xi^{(i)}$ to $\xi^{(i+1)}$ must be equally likely as the transition from $\xi^{(i+1)}$ to $\xi^{(i)}$. More formally, one requires

$$Q\left(\xi^{(i+1)}\right)\mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right) = Q\left(\xi^{(i)}\right)\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right) . \tag{6.15}$$

This directly implies that $Q(\xi)$ is an invariant distribution, since

$$\sum_{\xi^{(i)}} Q\left(\xi^{(i)}\right)\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right) = \sum_{\xi^{(i)}} Q\left(\xi^{(i+1)}\right)\mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right) = Q\left(\xi^{(i+1)}\right)\sum_{\xi^{(i)}}\mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right) = Q\left(\xi^{(i+1)}\right) . \tag{6.16}$$

Note, that it is also possible for a distribution to be invariant without requiring the chain to be homogeneous (Neal 1993). However, for the purpose of this thesis it is not enough to construct a Markov chain with respect to which the distribution we wish to sample from is stationary. We also require, that for any initial probability $\mathcal{P}_0(\xi)$ the probabilities at step i, $\mathcal{P}_i(\xi)$, will converge to the stationary distribution $Q(\xi)$ as $i \to \infty$. Such a Markov chain is called *ergodic* and can have only one invariant distribution, frequently referred to as equilibrium distribution (Neal 1993). It can be shown, that homogeneous Markov chains satisfy the requirement of *ergodicity*, i.e. independence of initial conditions (for a proof see e.g. Kemeny & Snell 1960, Neal 1993, Robert & Casella 2005). Hence, homogeneous Markov chains provide an efficient strategy to draw samples from the desired target distribution $\mathcal{P}(\xi)$.

6.5.3. Metropolis-Hastings algorithm

As described above, the construction of an ergodic Markov chain requires the definition of an appropriate transition probability $\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)$ which meets the requirement of homogeneity given in equation (6.15). Possibly the most generic approach to this problem was proposed by Metropolis et al. (1953) and subsequently generalized by Hastings (1970). A very intuitive description of this Metropolis-Hastings algorithm was given by Chib & Greenberg (1995), which will be outlined in the following.

The entire problem of constructing an ergodic Markov chain is reduced to finding an adequate transition probability. Clearly, if the transition probability $\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)$ satisfies equation (6.15), the search is over. However, generally one is concerned with problems where the transition probability does not meet the requirement of homogeneity, especially in cases where the target distribution has no standard form. For example, for some $\xi^{(i)},\xi^{(i+1)}$ one might find

$$Q\left(\xi^{(i+1)}\right)\mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right) < Q\left(\xi^{(i)}\right)\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right) . \tag{6.17}$$

This means that the process moves more frequently from $\xi^{(i)}$ to $\xi^{(i+1)}$ than from $\xi^{(i+1)}$ to $\xi^{(i)}$ violating the requirement of detailed balance. In order to achieve a balance of transitions, one can introduce the probability $\alpha\left(\xi^{(i+1)}|\xi^{(i)}\right) < 1$ of moving to the next state. If the transition is not accepted the process again returns $\xi^{(i)}$ as a value from the target distribution, thus, balancing the transitions. This allows for the definition of a new transition probability given by

$$\mathcal{T}_{MH}\left(\xi^{(i)}|\xi^{(i+1)}\right) = \mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right)\alpha\left(\xi^{(i)}|\xi^{(i+1)}\right) , \tag{6.18}$$

for $\xi^{(i)} \neq \xi^{(i+1)}$ and $\alpha\left(\xi^{(i)}|\xi^{(i+1)}\right)$ is yet to be determined. Since in the above example transitions from $\xi^{(i)}$ to $\xi^{(i+1)}$ are too frequent the probability for the backwards transition $\alpha\left(\xi^{(i)}|\xi^{(i+1)}\right)$ must be as large as possible, namely 1. Then the condition of detailed balance given in equation (6.15) can be written as

$$\begin{aligned}Q\left(\xi^{(i)}\right)\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)\alpha\left(\xi^{(i+1)}|\xi^{(i)}\right) &= Q\left(\xi^{(i+1)}\right)\mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right)\alpha\left(\xi^{(i)}|\xi^{(i+1)}\right) \\ &= Q\left(\xi^{(i+1)}\right)\mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right) . \end{aligned} \tag{6.19}$$

Through simple algebraic transformation this yields

$$\alpha\left(\xi^{(i+1)}|\xi^{(i)}\right) = \frac{Q\left(\xi^{(i+1)}\right)\mathcal{T}\left(\xi^{(i)}|\xi^{(i+1)}\right)}{Q\left(\xi^{(i)}\right)\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)} . \tag{6.20}$$

If the inequality in equation (6.17) is reversed an analogues result for $\alpha\left(\xi^{(i)}|\xi^{(i+1)}\right)$ is obtained. Thus, the probabilities $\alpha\left(\xi^{(i)}|\xi^{(i+1)}\right)$ and $\alpha\left(\xi^{(i+1)}|\xi^{(i)}\right)$ are introduced to guarantee $\mathcal{T}_{MH}\left(\xi^{(i)}|\xi^{(i+1)}\right)$ satisfies the homogeneity requirement. More precisely, the probability of moving to the next state must be set to

$$\alpha\left(\xi^{(i+1)}|\xi^{(i)}\right) = \begin{cases} \min\left[\frac{Q(\xi^{(i+1)})\mathcal{T}(\xi^{(i)}|\xi^{(i+1)})}{Q(\xi^{(i)})\mathcal{T}(\xi^{(i+1)}|\xi^{(i)})}, 1\right] & \text{if } Q\left(\xi^{(i)}\right)\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right) > 0 \\ 1 & \text{otherwise} \end{cases} . \tag{6.21}$$

The complete definition of the Metropolis-Hastings transition probability further requires to take into account the possibly nonzero probability that the process remains in its current state. Therefore, the complete Metropolis-Hastings transition kernel is given as

$$\mathcal{T}_{MH}\left(\xi^{(i+1)}|\xi^{(i)}\right) = \mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)\alpha\left(\xi^{(i+1)}|\xi^{(i)}\right) + \left[1 - \sum_{\xi^{(l)}}\mathcal{T}\left(\xi^{(l)}|\xi^{(i)}\right)\alpha\left(\xi^{(l)}|\xi^{(i)}\right)\right]\delta^K_{\xi^{(i+1)},\xi^{(i)}}, \tag{6.22}$$

with $\delta^K_{i,j}$ being the Kronecker delta. Because $\mathcal{T}_{MH}\left(\xi^{(i+1)}|\xi^{(i)}\right)$ satisfies detailed balance by construction the Metropolis-Hastings transition kernel has $Q(\xi)$ as its invariant distribution.

Several things should be remarked about the Metropolis-Hastings algorithm at this point. First, the candidate-generating probability $\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)$ completely specifies the algorithm (Chib & Greenberg 1995). Also, in contrast to rejection sampling algorithms, if a proposed value is rejected the current state is accepted as the new state (Neal 1993). Another very important feature is, that the calculation of $\alpha\left(\xi^{(i+1)}|\xi^{(i)}\right)$ does not depend on the normalization constant of $Q(\xi)$, as can be seen in equation (6.21). This is of special importance for many Bayesian applications since the normalization of the target distribution is generally not known. The major advantage of the Metropolis-Hastings algorithm is, that it allows the construction of an ergodic Markov chain from nearly any proposal distribution $\mathcal{T}\left(\xi^{(i+1)}|\xi^{(i)}\right)$. In particular, it allows to construct transition probabilities from a set of base transition probabilities $\mathcal{B}_1, ..., \mathcal{B}_s$ so that the complete transition probability leaves the desired distribution invariant. This is especially convenient for high dimensional problems, where each $\mathcal{B}_k$ might only change a subset of parameters (Neal 1993).

A special kind of a high dimensional Metropolis-Hastings sampler, the Hybrid Monte Carlo sampler, and its application to cosmological observations are presented in chapter 8 and 9.

6.5.4. Gibbs sampling

The Gibbs sampler, as proposed by Geman & Geman (1984) and Gelfand & Smith (1990), is a special case of the Metropolis Hastings algorithm presented above. It is applicable when the joint distribution is not known explicitly, but the conditional distribution of each individual variable is known. Suppose one needs to generate samples from a multivariate joint probability distribution $\mathcal{P}(\xi_1, \xi_2, ..., \xi_M)$ but one only knows how to sample each individual value from conditional probabilities $\mathcal{P}(\xi_q|\{\xi_p\} : p \neq q)$ for the value ξ_q given all other values. In this situation the transition probabilities can be built from a set of base transition probabilities $\mathcal{B}_1, \mathcal{B}_2, ..., \mathcal{B}_M$, where each of these changes only one single parameter, while leaving all other values unchanged (Neal 1993). More formally, a base transition probability can be written as

$$\mathcal{B}_q\left(\xi^{(i+1)}|\xi^{(i)}\right) = \mathcal{P}(\xi_q^{i+1}|\{\xi_p^i\} : p \neq q) \cdot \prod_{p \neq q}\delta^K_{\xi_p^{(i+1)},\xi_p^{(i)}}. \tag{6.23}$$

This means $\mathcal{B}_q$ draws a new value ξ_q^{i+1} from its distribution conditional on the current values of all other components, while leaving all these other components unchanged. These base transitions can then be applied in sequence to yield the complete transition probability

$$\mathcal{T}_{Gibbs}\left(\xi^{(i+1)}|\xi^{(i)}\right) = \mathcal{B}_M\left(\xi^{(i+1)}|\alpha^{(M-1)}\right) \circ \mathcal{B}_{M-1}\left(\alpha^{(M-1)}|\alpha^{(M-2)}\right) \circ ... \circ \mathcal{B}_1\left(\alpha^{(1)}|\xi^{(i)}\right), \tag{6.24}$$

where the symbol $\circ$ denotes the sequential application of the random processes and $\alpha^{(l)}$ denotes intermediate transition states. In order for the Gibbs algorithm to work, the transition probability $\mathcal{T}_{Gibbs}\left(\xi^{(i+1)}|\xi^{(i)}\right)$ must leave the desired target distribution invariant. Already intuitive considerations clarify that this is the case for the Gibbs

sampling algorithm. First, $\mathcal{B}_q$ leaves all values except for the qth value unchanged. Hence the desired marginal distribution for all these components is certainly invariant. In addition, the conditional probability for $\xi_q^{(i+1)}$ in the new state given all other components is defined to be the desired target distribution. This ensures, that if one started from the desired target distribution one still has the target distribution after the application of $\mathcal{B}_q$ (Neal 1993). This statement can be more formally proved by testing the homogeneity condition

$$
\begin{aligned}
\sum_{\xi^{(i)}} \mathcal{P}\left(\xi^{(i)}\right) \mathcal{B}\left(\xi^{(i+1)}|\xi^{(i)}\right) &= \sum_{\xi^{(i)}} \mathcal{P}(\xi_q^{(i)}|\{\xi_p^{(i)}\} : p \neq q) \mathcal{P}(\{\xi_p^{(i)}\} : p \neq q) \cdot \mathcal{P}(\xi_q^{(i+1)}|\{\xi_p^{(i)}\} : p \neq q) \cdot \prod_{p \neq q} \delta_{\xi_p^{(i+1)}, \xi_p^{(i)}}^K \\
&= \mathcal{P}(\xi_q^{(i+1)}|\{\xi_p^{(i+1)}\} : p \neq q) \mathcal{P}(\{\xi_p^{(i)}\} : p \neq q) \sum_{\xi_q^{(i)}} \mathcal{P}(\xi_q^{(i)}|\{\xi_p^{(i+1)}\} : p \neq q) \\
&= \mathcal{P}(\xi^{(i+1)}).
\end{aligned}
\tag{6.25}
$$

The Gibbs sampler therefore is homogeneous and has the desired target density as its invariant distribution. Provided that all $\mathcal{B}_q$ are non-zero it is also guaranteed that there exists a non-zero probability for the chain to move from any state, to any other state in M steps, with each step changing one component (Neal 1993). If there also exists a non-zero probability for the chain remaining in its current state, the Markov chain is irreducible and ergodic.

The major advantage of the Gibbs sampler therefore is, that given there exists a procedure to sample from the conditional probabilities, one can sample from the joint probability in M steps, without the need of rejecting any samples. An example of the Gibbs sampler and its application to cosmological data analysis is described in chapter 7.

6.6. Summary

In this chapter, the basic concepts and ideas of Bayesian statistics and the according mathematical framework have been discussed. After presenting a motivation for using statistical inference procedures in analyses of cosmological data sets, i.e. galaxy surveys, the two different viewpoints of Bayesian and conventional statistics in metrology are compared. These two different approaches not only differ in the notion of probability but also in the implications for the inference process. In particular, Bayesian statistics answers the underlying question to every measurement of how to estimate the probability distribution of possible signals compatible with the observations, while conventional statistics does not. Since Bayesian methods are frequently criticized for being subjective due to the incorporation of prior knowledge, we discussed and justified the use of such a prior probability. The 'No free lunch theorems' in the field of mathematical optimization clearly state that optimal search strategies require the incorporation of prior information in order to perform better than any other method. Further, we described an empirical prior learning process, demonstrating that learning without prior information is not possible. Nevertheless, prior distributions have to be carefully chosen, and particularly in natural sciences the prior distribution should be given by all a priori available measurements, rather than by purely theoretical considerations, which may be wrong. In Bayesian statistics the object of interest is the posterior distribution. It contains all our knowledge about the measurand and its uncertainties. Especially, in cases of high dimensionality or difficult functional shapes, as discussed in this thesis, the evaluation of the posterior distribution or its summaries is not trivial. For this reason, these posterior distributions can be approximated by a discrete set of random signal samples generated by Markov Chain Monte Carlo methods. We discussed two possible methods to generate such Markov samples, the first being the famous Metropolis-Hastings algorithm and the second is the Gibbs sampler. These two algorithms generate homogeneous and ergodic Markov chains, ensuring that the invariant distribution of the Markov chain is the desired posterior distribution. These two algorithms build the fundament for the cosmological data analysis methods as described in the remainder of this thesis.

7. ARES - Joint inference of the three dimensional density field and its power-spectrum

True reality lies beyond immediate sensation and the objects we see every day.

Hegel

Abstract

This chapter describes the development and implementation of a new Bayesian computer algorithm for the analysis of galaxy redshift surveys. The resultant computer algorithm ARES particularly aims at the joint inference of the three dimensional matter field and its power-spectrum, under the reasonable assumption that the long wavelength Fourier components are Gaussian distributed. Based on a very efficient Gibbs sampling procedure, ARES does not only provide a single estimate such as mean, mode or variance but samples from the joint posterior of the power-spectrum and density field conditional on a set of galaxy observations. This allows for reporting any desired statistical summary, in particular joint uncertainty information can be provided. The method is applied to mock catalogs, with highly structured observational masks and selection functions, in order to demonstrate its ability to infer the power-spectrum from real data sets, while fully accounting for any mask induced mode coupling.

7.1. Introduction

In the previous chapters we discussed the necessary key concepts of signal processing with digital computers and the mathematical framework of Bayesian statistics for cosmological data analysis. Based on these foundations, in this chapter, a new computer algorithm for joint analyses of three dimensional density fields and their power-spectra will be developed and implemented. The scientific aim of this computer algorithm is to provide detailed cosmographic descriptions of the three dimensional matter distribution as well as a statistical characterization of the large scale structure in terms of power-spectra and joint uncertainties. The analysis of the large scale structure is of particular interest for cosmology, since it contains a wealth of information on the origin and evolution of the Universe. The formation process of this structure involved a lot of interesting physics ranging from quantum field theory, general relativity to the dynamics of collisionless dark matter particles or the behavior of the baryonic sector, e.g. galaxy formation. Harvesting this information from probes of the large scale structure, such as large galaxy surveys, therefore is an important scientific task to further our knowledge and to establish a conclusive cosmological picture. In recent years great advances have been made, both in retrieving huge amounts of data and increasing sensitivity in galaxy redshift surveys. Especially the recent galaxy surveys, the 2dF Galaxy Redshift Survey (Colless et al. 2001) and the Sloan Digital Sky Survey (York et al. 2000) provide sufficient redshifts to probe the 3D galaxy distribution on large scales. In particular, the two point statistics of the matter distribution contains valuable information to test standard inflation and cosmological models, which describe the origin and evolution of all observed structure in the Universe. Measuring the power-spectrum from galaxy observations therefore has always attracted great interest. Precise determination of the overall shape of the power-spectrum can for instance place important constraints on neutrino masses, help to identify the primordial power-spectrum, and break degeneracies for cosmological parameter estimation from CMB data by measuring the parameter combination Ω_m/h (e.g. Hu et al. 1998, Spergel et al. 2003, Hannestad 2003, Efstathiou et al. 2002, Percival et al. 2002, Spergel et al. 2003, Verde et al. 2003). In addition, several characteristic length scales have been imprinted on the matter distribution throughout cosmic history, which can serve as new standard rulers to measure the Universe. A prominent example of these length scales is the sound

horizon, which yields oscillatory features in the power-spectrum, the so called baryon acoustic oscillations (BAO) (e.g. Silk 1968, Peebles & Yu 1970, Sunyaev & Zeldovich 1970). Since the physics governing these oscillatory features is well understood, precise measurements of the BAO will allow us to establish a new, precise standard ruler to measure the Universe through the distance redshift relation (Blake & Glazebrook 2003, Seo & Eisenstein 2003).

Unfortunately, contact between theory and observations cannot be made directly, since observational data is subject to a variety of systematic effects and statistical uncertainties. Such systematics and uncertainties arise either from the observational strategy or are due to intrinsic clustering behavior of the galaxy sample itself (Sánchez & Cole 2008). As already discussed in chapter 6, some of the most prominent uncertainties and systematics are:

- survey geometry and selection effects

- close pair incompleteness due to fiber collisions

- galaxy biases

- redshift space distortions

The details of galaxy clustering, and how galaxies trace the underlying density field are in general very complicated. The bias between galaxies and mass density is most likely non-linear and stochastic, so that the estimated galaxy spectrum is expected to differ from that of the mass (Dekel & Lahav 1999). Even in the limit where a linear bias could be employed, it still differs for different classes of galaxies (see e.g. Cole et al. 2005). In addition, the apparent density field, obtained from redshift surveys, will generally be distorted along the line-of-sight due to the existence of peculiar velocities.

However, the main cause for the systematic uncertainties in large scale power-spectrum estimations is the treatment of the survey geometry (Tegmark 1995, Ballinger et al. 1995). Due to the survey geometry the raw power-spectrum yields an expectation value for the power-spectrum, which is the true cosmic power-spectrum convolved with the survey mask (Cole et al. 2005). This convolution causes an overall distortion of the power-spectrum shape, and drastically decreases the visibility of the baryonic features.

The problems, mentioned above, have been discussed extensively in literature, and many different approaches to power-spectrum analysis have been proposed. Some of the main techniques to recover the power-spectrum from galaxy surveys are Fourier transform based, such as the optimal weighting scheme, which assigns a weight to the galaxy fluctuation field, in order to reduce the error in the estimated power (see e.g. Feldman et al. 1994, Tegmark 1995, Hamilton 1997a, Yamamoto 2003, Percival et al. 2004). Alternative methods rely on Karhunen-Loève decompositions (Tegmark et al. 1997, Tegmark & et al. 2004, Pope et al. 2004) or decompositions into spherical harmonics, which is especially suited to address the redshift space distortions problematic (Fisher et al. 1994, Heavens & Taylor 1995, Tadros et al. 1999, Percival et al. 2004, Percival 2005). In addition, to these deconvolution methods there exists a variety of likelihood methods to estimate the real space power-spectrum (Ballinger et al. 1995, Hamilton 1997a,b, Tadros et al. 1999, Percival 2005). In order to not just provide the maximum likelihood estimate but also conditional errors, Percival (2005) proposed a Markov Chain Monte Carlo method to map out the likelihood surface.

Nevertheless, as the precision of large scale structure experiments has improved, the requirement on the control and characterization of systematic effects, as discussed above, also steadily increases. It is of critical importance to propagate properly the uncertainties caused by these effects through to the matter power-spectrum and cosmological parameters estimates, in order to not underestimate the final uncertainties and thereby draw incorrect conclusions on the cosmological model.

This chapter describes a new alternative Bayesian approach to power-spectrum inference from three dimensional large scale structure datasets. This method already conceptionally differs from previous likelihood methods due to the different notion of probability between the Bayesian and the frequentist viewpoint (see e.g. Michel & Kirchhoff 1999, Trotta 2008). As already discussed in chapter 6, this thesis relies on Bayesian methods rather than conventional likelihood methods, since they yield more general and profound statements about measurements. In example, conventional likelihood methods can only answer questions of the inner form like :" Given the true value s of a signal, what is the probability distribution of the measured values d?" A Bayesian method, on the other hand, answers questions of the type :"Given the observations d, what is the probability distribution of the true underlying signal s?" For this reason, Bayesian statistics directly answers the underlying question to every measurement problem, of how to infer the true value of the signal from observations, while conventional likelihood methods do not (Michel

& Kirchhoff 1999). The main task of the Bayesian inference method, proposed in this thesis, is to infer the power-spectrum posterior distribution $\mathcal{P}(\{P(k_i)\}|\{d_i\})$, with $P(k_i)$ being the power-spectrum coefficients of the k_ith mode and $d_i = d(\vec{x}_i)$ is an observation at position $\vec{x}_i$ in three dimensional configuration space. In this fashion the method permits a fully global analysis of the power-spectrum, taking into account all systematic effects and statistical uncertainties. This posterior distribution would then contain all information on the two point statistics supported by the data. In order to explore this posterior distribution we employ a Gibbs sampling method, previously applied to CMB data analysis (see e.g. Wandelt et al. 2004, Wandelt 2004, O'Dwyer et al. 2004, Eriksen et al. 2004, Jewell et al. 2004, Larson et al. 2007, Eriksen et al. 2007, Jewell et al. 2009).

Since direct sampling from $\mathcal{P}(\{P(k_i)\}|\{d_i\})$ is impossible or at least difficult, they propose instead to draw samples from the full joint posterior distribution $\mathcal{P}(\{P(k_i)\}, \{s_i\}|\{d_i\})$ of the power-spectrum coefficients $P(k_i)$ and the 3D matter density contrast amplitudes s_i conditional on a given set of data points $\{d_i\}$. This is achieved by iteratively drawing density samples from a Wiener-posterior distribution and power-spectrum samples via an efficient Gibbs sampling scheme (see Fig. 7.1 for an illustration). Here, artificial mode coupling, as introduced by survey geometry and selection function, is resolved by solving the Wiener-filtering equation, which naturally regularizes inversions of the observational response operator in unobserved regions. In this fashion, we obtain a set of Monte Carlo samples from the joint posterior, which allows us to compute any desired property of the joint posterior density, with accuracy only limited by the sample size. In particular, we obtain the power-spectrum posterior $\mathcal{P}(\{P(k_i)\}|\{d_i\})$ by simply marginalizing the joint posterior $\mathcal{P}(\{P(k_i)\}, \{s_i\}|\{d_i\})$ over the auxiliary density amplitudes s_i, which is trivially achieved by ignoring the s_i samples.

The Gibbs sampler also offers unique capabilities for propagating systematic uncertainties end-to-end. Any effect, for which we can define a sampling algorithm, either jointly with or conditionally on other quantities, can be propagated seamlessly through to the final posterior.

It is worth noting, that our method differs from traditional methods of analyzing galaxy surveys in a fundamental aspect. Traditional methods consider the analysis task as a set of steps, each of which arrives at intermediate outputs which are then fed as inputs to the next step in the pipeline. Our approach is a truly global analysis, in the sense that the statistics of all science products are computed jointly, respecting and exploiting the full statistical dependence structure between various components.

In this chapter ARES (Algorithm for REconstruction and Sampling), a computer algorithm to perform a full Bayesian data analysis of 3D redshift surveys, is presented. In section 7.3 we give an introduction to the idea of the large scale structure Gibbs sampling approach, by first reviewing the Gibbs sampling algorithm as a special case of Markov Chain Monte Carlo methods and then giving an overview of its specific application to Bayesian power-spectrum inference. The following sections 7.4 and 7.5, describe the details of our algorithm. Here we will derive the Wiener posterior from a Bayesian point of view and demonstrate how to generate three dimensional density samples. Further, we will demonstrate that the power-spectrum can efficiently be drawn from an inverse gamma distribution. The choice of the prior and the relevance for the cosmic variance are discussed in section 7.6. Details concerning the numerical implementation , such as random number generation or parallelization, are discussed in section 7.7. We then test ARES thoroughly in section 7.8, particularly focussing on the treatment of survey masks and selection functions. In section 7.9 we demonstrate the running median filter, and use it as an example to demonstrate how uncertainties can be propagated to all inferences based on the set of Gibbs samples. Finally we conclude in section 7.10 by discussing the results of the method and giving an outlook for future extensions and applications.

7.2. Notation

In this section, we describe the basic notation used throughout this work. Let the quantity $\rho_i = \rho(\vec{x}_i)$ be the field amplitude of the three dimensional field $\rho(\vec{x})$ at position $\vec{x}_i$. Then the index i has to be understood as a multi index, which labels the three components of the position vector:

$$\vec{x}_i = [x_i^1, x_i^2, x_i^3], \tag{7.1}$$

where x_i^j is the jth component of the ith position vector. Alternatively one can understand the index i as a set of three indices $\{r, s, t\}$ so that for an equidistant grid along the three axes the position vector can be expressed as:

$$\vec{x}_i = \vec{x}_{r,s,t} = [\Delta x\, r, \Delta y\, s, \Delta z\, t], \tag{7.2}$$
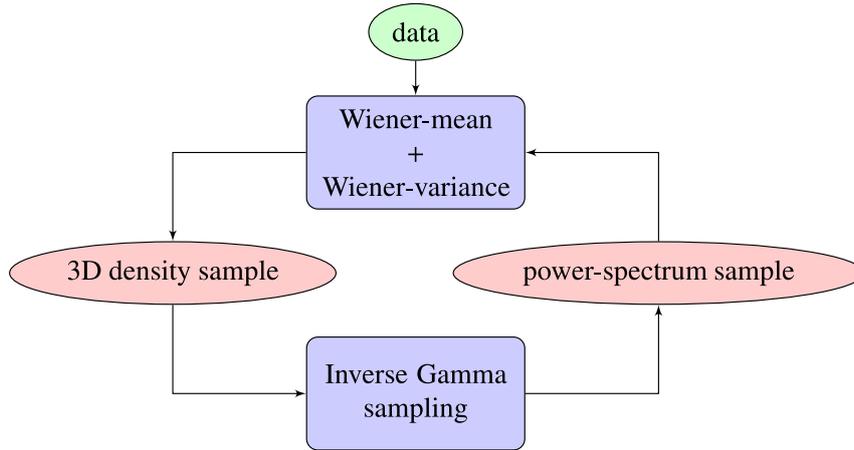
Figure 7.1.: Flow-chart depicting the two step iterative Gibbs sampling procedure. First a density sample is drawn from the Wiener posterior distribution, conditional on which a power-spectrum sample is generated in the second Gibbs sampling step.

with Δx, Δy and Δz being the grid spacing along the three axes. With this definition we yield:

$$\rho_i \equiv \rho_{r,s,t} \,. \tag{7.3}$$

Also note that any summation running over the multi index i is defined as the three sums over the three indices r, s and t:

$$\sum_i \equiv \sum_r \sum_s \sum_t \,. \tag{7.4}$$

Further, we will frequently use the notation $\{\rho_i\}$, which denotes the set of field amplitudes at different positions $\vec{x}_i$. In particular:

$$\{\rho_i\} \equiv \{\rho_0, \rho_1, \rho_2, ..., \rho_{N-1}\} \,, \tag{7.5}$$

where N is the total number of position vectors.

7.3. The Large scale structure Gibbs sampler

7.3.1. Gibbs sampling

The aim of the Bayesian analysis method, as presented here, is to map out the joint posterior distribution of the three dimensional density field and the corresponding matter power-spectrum. In principle, this task could be solved by dividing the parameter space into an equidistant grid and estimating the posterior values at each grid node. However, since the number of grid points required for such an analysis scales exponentially with the number of free parameters, this approach cannot be realized efficiently. In such situations Markov Chain Monte Carlo (MCMC) methods are known to generally outperform any known deterministic approach to this problem (see e.g. Neal 1993, Andrieu et al. 2003, Robert & Casella 2005). The basic ideas of MCMC methods is to approximate the target probability distribution $\mathcal{P}(\xi)$, defined on a high dimensional space χ, by a set of N independent and identically distributed (i.i.d) samples $\{\xi^{(i)}\}_{i=1}^N$ drawn from the target distribution (Andrieu et al. 2003). Here, the Markov chain, usually realized as a discrete random process, efficiently generates samples $\xi^{(i)}$ while exploring the state space χ. Given such a set of samples it is possible to approximate the target posterior with the following point particle distribution:

$$\mathcal{P}_N(\xi) = \frac{1}{N} \sum_{i=1}^N \delta^D(\xi - \xi^{(i)}) \,, \tag{7.6}$$

with $\delta^D(x)$ being the Dirac delta distribution. Given such a representation of the target density any statistical summary can be approximated with tractable sums, rather than integrals, as:

$$I_N(f) = \frac{1}{N} \sum_{i=1}^{N} f(\xi^{(i)}) \xrightarrow[N \to \infty]{\text{a.s.}} I(f) = \int_{\chi} f(\xi) \mathcal{P}(\xi) \, \mathrm{d}\xi \, . \tag{7.7}$$

This means that the estimate $I_N(f)$ is unbiased and will almost surely (a.s.) converge to $I(f)$ by the Strong Law of Large Numbers (Andrieu et al. 2003, Robert & Casella 2005). The Gibbs sampler, as proposed by Geman & Geman (1984) and Gelfand & Smith (1990), is a special case of the MCMC algorithm. Its special importance to Bayesian data analysis arises from the fact that it allows for very efficient parameter space exploration if the problem under consideration can be formulated accordingly. The Gibbs sampler is generally applicable to problems where the joint distribution $\mathcal{P}(\xi_1, \xi_2, ..., \xi_M)$ of parameters is not known explicitly, but the conditional distribution $\mathcal{P}(\xi_q | \{\xi_p\} \, : \, p \neq q)$ of each individual variable ξ_q is known (Neal 1993). In this case, the theory of Gibbs sampling (Gelfand & Smith 1990, Tanner 1996, O'Hagan 2000) states, that iterative random draws from the individual conditional distributions will yield samples from the joint target distribution. More explicitly, the iteration of the following random draws:

$$
\begin{aligned}
&1) \quad \xi_1^{(j+1)} \curvearrowleft \mathcal{P}(\xi_1 | \xi_2^j, ..., \xi_M^j) \\
&2) \quad \xi_2^{(j+1)} \curvearrowleft \mathcal{P}(\xi_2 | \xi_1^{j+1}, \xi_3^j, ..., \xi_M^j) \\
& \quad . \\
& \quad . \\
& \quad . \\
&M) \quad \xi_M^{(j+1)} \curvearrowleft \mathcal{P}(\xi_M | \xi_1^{j+1}, \xi_2^{j+1}, ..., \xi_{M-1}^{j+1}) \, ,
\end{aligned}
$$

will yield samples from the joint multivariate density $\mathcal{P}(\xi_1, \xi_2, ..., \xi_M)$. Since the Gibbs sampler is an homogeneous Markov chain by definition and is also ergodic, the sequence of samples are proven to converge to samples from the desired target distribution (Neal 1993). For a more detailed introduction to MCMC methods and the Gibbs sampler the reader is referred to the literature (see e.g. Kemeny & Snell 1960, Hastings 1970, Neal 1993, Andrieu et al. 2003, Robert & Casella 2005, Gamerman & Lopes 2006).

7.3.2. Joint power-spectrum and density field inference

The advantages of Gibbs sampling procedures for Bayesian inference of cosmological density fields and power-spectra have already been previously demonstrated in the case of CMB (see e.g. Wandelt et al. 2004, Wandelt 2004, Eriksen et al. 2004, Jewell et al. 2004). Here we extend their approach to be applicable to three dimensional galaxy surveys. The entire purpose of the Gibbs sampler in our approach is to efficiently map out the joint posterior distribution $\mathcal{P}(\{P(k_i)\}, \{s_i\} | \{d_i\})$ of the power-spectrum coefficients $P(k_i)$ and the 3D matter density contrast amplitudes s_i given a set of observations $\{d_i\}$ via a MCMC method. The corresponding two step Gibbs sampling procedure can therefore be written as follows:

$$
\begin{aligned}
&1) \quad \{s_i\}^{(j+1)} \curvearrowleft \mathcal{P}(\{s_i\} | \{P(k_i)\}^{(j)}, \{d_i\}) \\
&2) \quad \{P(k_i)\}^{(j+1)} \curvearrowleft \mathcal{P}(\{P(k_i)\} | \{s_i\}^{(j+1)}, \{d_i\}) \, ,
\end{aligned}
\tag{7.8}
$$

where $\mathcal{P}(\{s_i\} | \{P(k_i)\}^{(j)}, \{d_i\})$ is the conditional probability of the three dimensional density field conditional on a given power-spectrum and the data and $\mathcal{P}(\{P(k_i)\} | \{s_i\}^{(j+1)}, \{d_i\})$ is the conditional probability of the power-spectrum conditional on a previously sampled three dimensional density field and the data. Iteration of the Gibbs sampling steps (7.8) will therefore yield a set of MCMC samples and therefore an approximation of the joint posterior:

$$
\mathcal{P}_{N_{Gibbs}}(\{P(k_i)\}, \{s_i\} | \{d_i\}) \; = \; \frac{1}{N_{Gibbs}} \sum_{k=1}^{N_{Gibbs}} \delta^D(\{s_i\} - \{s_i\}^{(k)}) \, \delta^D(\{P(k_i)\} - \{P(k_i)\}^{(k)}) \, ,
$$

$$\tag{7.9}$$

where N_{Gibbs} is the number of Gibbs samples. The overall Gibbs sampling procedure is depicted in Fig. 7.1. According to the Gibbs sampling procedure (7.8) we first generate a density field realization from the Wiener posterior via the procedure described in section 7.4. Given such a density field sample, in the second step a power-spectrum sample is drawn from an inverse gamma distribution as discussed in section 7.5.1. Iteration of these sampling steps yields a sampled representation based on which any desired statistical summary can be reported. Further, it allows for estimating normalization factors required for Bayesian model comparisons and calculation of odds factors. In particular, it is possible to provide an analytic description of the full power-spectrum posterior $\mathcal{P}(\{P(k_i)\}|\{d_i\})$ by marginalizing over the three dimensional density field samples. As will be demonstrated in section 7.5.2 such a procedure yields a Blackwell-Rao estimate for the power-spectrum posterior. Also note, that additional sampling steps such as the joint inference of biases or peculiar velocities can be added to the Gibbs sampling procedure (7.8), allowing for a fully global analysis of all these quantities.

7.4. Sampling the signal maps

As discussed above, the first step in the Gibbs sampling procedure is to obtain three dimensional density samples. To do this we assume the conditional probability $\mathcal{P}(\{s_i\}|\{P(k_i)\}^{(j)}, \{d_i\})$ to be a multivariate Gaussian distribution, the so called Wiener posterior. A random density sample drawn from this Wiener posterior can then be obtained by infering the Wiener mean and adding a fluctuation term consistent with the Wiener variance, as described below. The numerically most demanding procedure in this sampling step is the calculation of the Wiener mean, which amounts to three dimensional Wiener filtering. Wiener filtering in three dimensions usually involves the inversion of huge matrices, generally on the order of $10^6 \times 10^6$ or more entries. Since the aim of the Gibbs sampling chain is to produce as many samples as possible in a given time, this matrix inversion step also represents a bottleneck for our method. This problem has been abolished, by implementing a very efficient operator based Krylov method for the matrix inversion (see e.g. Wandelt et al. 2004, Wandelt 2004, Eriksen et al. 2004, Jewell et al. 2004, Kitaura & Enßlin 2008). The efficiency of this algorithm allows for the generation of many thousands of samples from the Wiener posterior in computationally feasible times. In the following sections we will describe in more detail the process of generating density samples from the Wiener posterior.

7.4.1. The Wiener filter

As already described above, the main task for the signal sampling step is to derive the Wiener mean, which in the case of a Gaussian distribution coincides with the maximum a posteriori values for the signal amplitudes s_i. In the following we are going to discuss the derivation of the Wiener posterior and the generation of signal samples from a Bayesian point of view.

According to Bayes' theorem the conditional signal posterior can be written as the product of a signal prior and a likelihood normalized by the so called evidence. Further, here we will use the signal covariance matrix S rather than the power-spectrum $\{P(k_i)\}$. It is well known, that the power-spectrum is just the Fourier transform of the signal covariance in configuration space. Since the Fourier transform is a basis transformation with a unitary transformation matrix, the signal covariance matrix S and the power-spectrum $\{P(k_i)\}$ can be used interchangeably for a normalized Fourier transform (see section 7.5.1 and Appendix A.2 for more details).

We can therefore write the signal posterior as:

$$
\begin{aligned}
\mathcal{P}(\{s_i\}|S, \{d_i\}) &= \frac{\mathcal{P}(S)}{\mathcal{P}(\{d_i\}, S)} \mathcal{P}(\{s_i\}|S) \mathcal{P}(\{d_i\}|\{s_i\}, S) \\
&= \frac{1}{\mathcal{P}(\{d_i\}|S)} \mathcal{P}(\{s_i\}|S) \mathcal{P}(\{d_i\}|\{s_i\}),
\end{aligned}
\tag{7.10}
$$

where we assume that the data amplitudes d_i are conditionally independent of the signal covariance matrix S, once the signal amplitudes s_i are given. Following Bardeen et al. (1986), we describe the signal prior for the large scale matter distribution as a multivariate Gaussian, with zero mean and the signal covariance S. We can then write:

$$
\mathcal{P}(\{s_i\}|S) = \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2} \sum_i \sum_j s_i S_{ij}^{-1} s_j}.
\tag{7.11}
$$

The Fourier transform of the signal covariance matrix S has an especially appealing form in Fourier space. It is well known, that in an homogeneous and isotropic universe the Fourier transform of the signal covariance is a diagonal matrix, with the diagonal elements being the power-spectrum. Hence, we can express the Fourier representation of the signal covariance as:

$$\hat{S}_{kl} = \delta_{kl}^{K} P_k \tag{7.12}$$

where the ˆ-symbol denotes a Fourier transform, δ_{ij}^{K} is the Kronecker delta and $P_k = P(k_k)$ is the power-spectrum coefficient at the Fourier mode $\vec{k}_k$ in three dimensional pixel space (see e.g. Padmanabhan 1993, Lahav & Suto 2004).

The choice of the Gaussian prior can be justified by inflationary theories, which predict the matter field amplitudes to be Gaussian distributed in the linear regime approximately on scales $k \lesssim 0.15\,\mathrm{h/Mpc}$ (Peacock & Dodds 1994, Percival et al. 2001).

On nonlinear scales the Gaussian prior does not represent the full statistical behavior of the matter field anymore. During nonlinear gravitational structure formation the statistics of the initial density field evolves from a Gaussian distribution towards a distribution which can be approximated to one point statistics by a log normal distribution, as commonly assumed in literature (Coles & Jones 1991, Colombi 1994, Kayo et al. 2001).

However, note that in this case the Gaussian prior still describes the two point statistics of the underlying density field even in the nonlinear regime. The Gaussian prior should therefore be regarded as our a priori knowledge of the matter distribution, which is only formulated up to two point statistics. Next, we discuss the likelihood $\mathcal{P}(\{d_i\}|\{s_i\})$ given in equation (7.10).

As we seek to recover the maximum a posteriori signal s_i from the set of observations d_i we must assume a model which relates these two quantities. The most straightforward data model is linear, and can be written as:

$$d_i = \sum_k K_{ik}\, s_k + \epsilon_i\,, \tag{7.13}$$

where K_{ij} is an observation response operator and ϵ_i is an additive noise contribution, which will be defined in more detail in the next section. If we assume the noise ϵ_i to be Gaussian distributed, with zero mean and covariance N, we can express the likelihood as:

$$\mathcal{P}(\{d_i\}|\{s_i\}) = \frac{1}{\sqrt{\det{(2\pi\mathrm{N})}}} e^{-\frac{1}{2}\left(\sum_i \sum_j \left[d_i - \sum_k K_{ik}\, s_k\right] N_{ij}^{-1} \left[d_j - \sum_l K_{jl}\, s_l\right]\right)}\,, \tag{7.14}$$

where we simply inserted the data model given in equation (7.13) into the Gaussian noise distribution.

With these definitions the signal posterior is a multivariate Gaussian distribution and can be written as:

$$\mathcal{P}(\{s_i\}|S, \{d_i\}) \quad \propto \quad e^{-\frac{1}{2}\left(\sum_i \sum_j s_i\, S_{ij}^{-1}\, s_j + \left[d_i - \sum_k K_{ik}\, s_k\right] N_{ij}^{-1} \left[d_j - \sum_l K_{jl}\, s_l\right]\right)}\,, \tag{7.15}$$

where we omitted the normalization factor, which is of no interest in the following.

Note, that omitting the normalization of the likelihood, requires that the additive noise term is independent of any signal contribution, as otherwise the noise covariance matrix would carry signal information and could not be neglected in the following. This assumption, however, is in general not true for the Poissonian noise, as described, in the next section.

The maximum of this signal posterior can then easily be found by either completing the square in the exponent of equation (7.15), or by extremizing $\mathcal{P}(\{s_i\}|S, \{d_i\})$ with respect to the signal amplitudes s_i. The latter approach allows us to directly read off the Wiener filter equation from equation (7.15), by simply differentiating the exponent with respect to s_i and setting the result to zero. The result is the famous Wiener filter equation which is given as:

$$\sum_j \left[S_{ij}^{-1} + \sum_m \sum_l K_{mi} N_{ml}^{-1} K_{lj} \right] m_j = \sum_m \sum_l K_{mi} N_{ml}^{-1} d_l\,, \tag{7.16}$$

where we denoted the variable m_j as a Wiener mean signal amplitude, to clarify that this reconstruction is the mean

and not a typical sample of the distribution. The solution of this equation requires inverting the matrix:

$$D_{ij} = S_{ij}^{-1} + \sum_m \sum_l K_{mi} N_{ml}^{-1} K_{lj} \,, \tag{7.17}$$

which leads to the solution for the signal amplitudes

$$m_i = \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} d_l \,. \tag{7.18}$$

This result demonstrates that inferring the maximum a posteriori values m_i for the signal amplitudes s_i involves inversions of the Wiener filter operator D. Therefore, the signal-sampling operation is by far the most demanding step of our Gibbs sampling scheme, as it requires the solution of a very large linear system.

Formally speaking, in practice, this corresponds to inverting matrices of order $\sim 10^6 \times 10^6$ or larger, which clearly is not computationally feasible through brute-force methods. For example, matrix inversion algorithms, based on usual linear algebra methods, have a numerically prohibitive $O(N_{pix}^3)$ scaling, in order to transform to the eigenspace of the system, which bars sampling from the signal posterior.

This is the situation in which Kitaura & Enßlin (2008) proposed a particular operator based inversion technique to allow for computationally efficient calculation of the Wiener filter equation in three dimensional space.

In this implementation, the system of equations (7.18) can be solved by means of conjugate gradients (CGs). The computational scaling of this method is thus reduced to the most expensive step for applying the operator on the right-hand side of the equations, which in our case is the Fast Fourier transform, which scales as $O(N_{pix} \log(N_{pix}))$.

7.4.2. The galaxy data model

In order to adapt the Wiener filter procedure for the specific application to galaxy observations, we are going to present the galaxy data model together with the corresponding Poissonian noise covariance matrix.

It is possible to model the observed galaxy distribution as a realization of an inhomogeneous Poissonian process (Martínez & Saar 2002). We then assume the observed galaxy numbers $N_i^O = N^O(\vec{x}_i)$ at position $\vec{x}_i$ in three dimensional configuration space to be drawn from a Poissonian distribution (Martínez & Saar 2002, Kitaura et al. 2009).

$$N_i^O \curvearrowleft \mathcal{P}(N_i^O | \lambda_i^O) = \frac{\lambda_i^{O^{N_i^O}} e^{-\lambda_i^O}}{N_i^O!} \,, \tag{7.19}$$

where the arrow denotes a random draw from the probability distribution and λ_i^O is the mean observable galaxy number at position $\vec{x}_i$. We can then write the observed galaxy numbers at discrete positions as:

$$N_i^O = \langle N_i^O \rangle + \epsilon_i^O = \lambda_i^O + \epsilon_i^O \,, \tag{7.20}$$

where the noise term ϵ_i^O denotes the difference between the observed galaxy number and the mean observable galaxy number. The Poissonian noise covariance matrix can then easily be obtained by:

$$N_{ij}^P = \langle \epsilon_i^O \epsilon_j^O \rangle = \langle [N_i^O - \langle N_i^O \rangle][N_j^O - \langle N_j^O \rangle] \rangle = \delta_{ij}^K \langle N_i^O \rangle = \delta_{ij}^K \lambda_i^O \,, \tag{7.21}$$

where we simply calculated the Poissonian variance for the observed galaxy number assuming the galaxies to be independent and identically distributed (i.i.d). The mean observable galaxy number can be related to the true mean galaxy number λ_i by applying the observation response operator R_{ij} as:

$$\lambda_i^O = \sum_j R_{ij} \lambda_j \,, \tag{7.22}$$

The true mean galaxy number, on the other hand, can be related to the dark matter over density field, the signal s_i, by introducing a physical model in the form of a bias operator B_{ij}, e.g. a scale dependent bias:

$$\lambda_i = \bar{\lambda} \left(1 + \sum_j B_{ij} s_j \right) . \tag{7.23}$$

By inserting equations (7.22) and (7.23) into equation (7.20) and applying trivial algebraic conversions, we yield the data model:

$$d_i \quad = \quad \frac{N_i^O}{\bar{\lambda}} - \sum_j R_{ij} = \sum_j R_{ij} \sum_k B_{jk} s_k + \frac{\epsilon_i^O}{\bar{\lambda}} \, , \tag{7.24}$$

For the case of galaxy redshift surveys the response operator R_{ij} is the product of the sky mask and the selection function, which are both local in configuration space, and hence the response operator turns to:

$$R_{ij} = \delta_{ij}^K M_i F_i \, , \tag{7.25}$$

where M_i is the value of the sky mask and F_i is the value of the selection function at position i. We therefore arrive at the data model already described in equation (7.13), which reads:

$$
\begin{aligned}
d_i \quad &= \quad M_i F_i \sum_k B_{ik} s_k + \frac{\epsilon_i^O}{\bar{\lambda}} \\
&= \quad \sum_k K_{ik} s_k + \epsilon_i \, ,
\end{aligned}
\tag{7.26}
$$

where we introduced the effective observation response operator $K_{ij} = M_i F_i B_{ij}$ and the noise contribution $\epsilon_i = \epsilon_i^O / \bar{\lambda}$. This is the galaxy data model which we derived from the assumption of the Poissonian distribution of galaxies.

The Wiener filter operator requires the definition of the noise covariance matrix N, which for the Poissonian noise can be expressed as:

$$N_{ij} = \langle \epsilon_i \epsilon_j \rangle = \frac{\langle \epsilon_i^O \epsilon_j^O \rangle}{\bar{\lambda}^2} = \delta_{ij}^K \frac{\lambda_i^O}{\bar{\lambda}^2} \, , \tag{7.27}$$

where we used the Poissonian noise covariance matrix given in equation (7.21).

However, introducing equation (7.22) and (7.23) yields the noise covariance matrix:

$$N_{ij} = \delta_{ij}^K \frac{1}{\bar{\lambda}} \left[\sum_k R_{ik} \left(1 + \sum_l B_{kl} s_l \right) \right] \, , \tag{7.28}$$

which immediately reveals, that there is a correlation between the underlying signal amplitudes s_i and the level of shot noise produced by the discrete distribution of galaxies (see e.g. Seljak 1998).

Nevertheless, as pointed out in the previous section, the Wiener filter relies on the fact, that the additive noise contribution is uncorrelated with the signal. Hence, we have to assume the noise covariance to be uncorrelated with the signal, but it may have some structure.

Therefore, we provide two approaches to effectively approximate the noise covariance matrix given in equation (7.28).

In the first approach we calculate an effective noise covariance matrix by averaging the noise covariance matrix given in equation (7.28) over the signal. We then obtain:

$$
\begin{aligned}
\bar{N}_{ij} \quad &= \quad \langle N_{ij} \rangle_s \\
&= \quad \delta_{ij}^K \frac{1}{\bar{\lambda}} \left[\sum_k R_{ik} \left(1 + \sum_l B_{kl} \langle s_l \rangle_s \right) \right] \\
&= \quad \delta_{ij}^K \frac{1}{\bar{\lambda}} \left[\sum_k R_{ik} \right] \, ,
\end{aligned}
\tag{7.29}
$$

where we used the fact, that the ensemble mean of the signal amplitudes for the density contrast vanishes. Note, that this model also arises when persuing a least squares approach to matter field reconstructions rather than the Bayesian approach as described in this thesis (for details see Kitaura & Enßlin 2008).
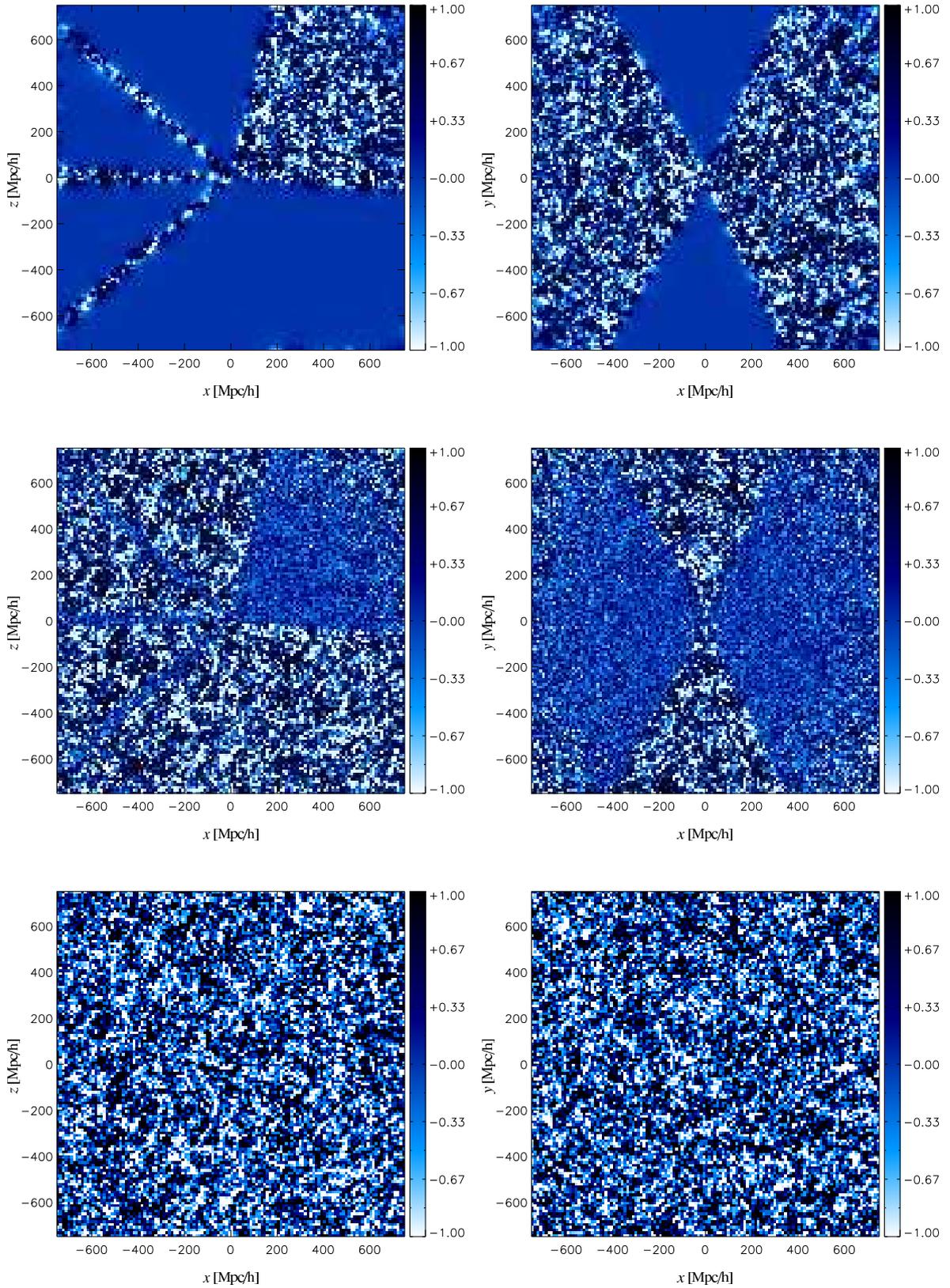
**Figure 7.2.:** Slices through a signal sample produced in one Gibbs sampling step. The right (left) panels depict slices through the three dimensional box at $z = 0$ ($y = 0$). The upper panels show the Wiener filtered mean signal, the middle panels present the fluctuation term, and the lower panels show the full, noiseless Gibbs sample. The color map encodes the amplitude of the density contrast.

In the other approach we introduce a noise structure function n_i^{SF} given as:

$$n_i^{SF} = \frac{\lambda_i^O}{\bar{\lambda}^2} . \tag{7.30}$$

With this definition the noise is approximated as being uncorrelated to the signal, but nonuniform. The noise covariance matrix then reads:

$$N_{ij}^{SF} = \delta_{ij}^K n_i^{SF} . \tag{7.31}$$

In order to use this noise structure function we have to infer λ_i^O from the observed galaxy numbers N_i^O. By applying Bayes' theorem to the Poissonian distribution given in relation (8.3) we yield:

$$\mathcal{P}(\lambda_i^O|N_i^O) = \mathcal{P}(\lambda_i^O) \frac{\mathcal{P}(N_i^O|\lambda_i^O)}{\mathcal{P}(N_i^O)} . \tag{7.32}$$

In the absence of any further a priori knowledge on λ_i^O we assume a flat prior and search for the maximum of:

$$\mathcal{P}(\lambda_i^O|N_i^O) = \frac{\lambda_i^{O N_i^O} e^{-\lambda_i^O}}{\Gamma(N_i^O + 1)} , \tag{7.33}$$

which is normalized to yield unity when integrated over all λ_i^O.

The noise structure function n_i^{SF} can then be inferred by searching the most likely value for λ_i^O from equation (7.33). This yields:

$$n_i^{SF} = \frac{N_i^O}{\bar{\lambda}^2} . \tag{7.34}$$

Another estimator for n_i^{SF} is based on evaluating the mean of the probability distribution given in equation (7.33). The ensemble mean is calculated as:

$$\langle \lambda_i^O \rangle = \int_0^\infty d\lambda_i^O \lambda_i^O \frac{\lambda_i^{O N_i^O} e^{-\lambda_i^O}}{\Gamma(N_i^O + 1)} = N_i^O + 1 . \tag{7.35}$$

in this case the noise structure function n_i^{SF} can be written as:

$$n_i^{SF} = \frac{N_i^O + 1}{\bar{\lambda}^2} . \tag{7.36}$$

A more thorough discussion on Poissonian noise models and their numerical implications for matter field reconstructions can be found in Kitaura et al. (2009).

7.4.3. Drawing signal samples

In the previous sections, we described the Wiener filter and the galaxy data model, which are required to infer the mean of the signal posterior. However, this mean signal is no sample from the signal posterior yet, neither does it represent a physical density field, as it lacks power in the low signal to noise regions. To create a true sample from the signal posterior, one must therefore add a fluctuation term y_i, which compensates the power lost due to noise filtering. The signal sample can then be written as the sum of the signal mean and a fluctuation term:

$$s_i = m_i + y_i . \tag{7.37}$$

In our approach we realize the fluctuation term by generating a mock signal s_i^* and a mock observation d_i^* consistent with the data model given in equation (7.26). This kind of mock observation generation is well known in literature and has been applied to various scientific applications, as for instance the generation of constrained initial conditions for Nbody simulations (see e.g. Bertschinger 1987, Hoffman & Ribak 1991, Ganon & Hoffman 1993, Kitaura &

Enßlin 2008). The fluctuation term can then simply be calculated as:

$$y_i = s_i^* - \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} d_l^* . \tag{7.38}$$

The interpretation of this equation is simple. In the high signal to noise regime, the Wiener filter is nearly a pass-through operator, meaning the reconstructed signal is nearly identical to the true underlying signal. Therefore, as the variance is low, the fluctuation term tends towards zero. In the low signal to noise regime, on the other hand, the Wiener filter will block, and no signal can be reconstructed. The fluctuation term will therefore be nearly identical to the underlying mock signal s_i^*.

In this fashion we add the correct power to the Wiener mean reconstruction. The effect of adding the fluctuation term to the Wiener mean is presented in Fig. 7.2, where we see the Wiener mean reconstruction, the fluctuation term and the sum of both.

The mock data d_i^* is generated to obey the data model described in equation (7.26) and the Wiener variance.

We therefore first draw a mock signal s_i^* with correct statistics from the multivariate Gaussian signal prior given in equation (7.11). Such a mock signal is best generated in Fourier space following the description of Martel (2005). One first draws two Gaussian random numbers, χ_a and χ_b, with zero mean and unit variance and then calculates the real and imaginary part of the signal in Fourier space as:

$$
\begin{aligned}
RE(\hat{s}_k) &= \sqrt{\frac{P_k}{2}} \chi_a \\
IM(\hat{s}_k) &= \sqrt{\frac{P_k}{2}} \chi_b ,
\end{aligned}
\tag{7.39}
$$

where P_k is the power-spectrum coefficient at the kth position in Fourier space. Note, that the mock signal s_i^* is supposed to be a real quantity, and therefore hermiticity has to be imposed in Fourier space before performing the inverse Fourier transform (for details see Martel 2005).

Next we have to generate the additive noise contribution. In order to draw a noise term with the correct Poissonian statistics, we first draw a random number N_i^* from the Poissonian distribution:

$$N_i^* \curvearrowleft \mathcal{P}(N_i^*|\lambda_i^*) , \tag{7.40}$$

where we choose the mean observed galaxy number to be $\lambda_i^* = n_i^{SF} \bar{\lambda}^2$. According to equations (7.20) and (7.26) the mock noise term ϵ_i^* can be calculated as:

$$\epsilon_i^* = \frac{N_i^* - n_i^{SF} \bar{\lambda}^2}{\bar{\lambda}} . \tag{7.41}$$

It is clear by construction that this mock noise term has vanishing mean and the correct noise covariance matrix. Then, according to equation (7.26) the mock observation is given as:

$$d_i^* = \sum_k K_{ik} s_k^* + \epsilon_i^* . \tag{7.42}$$

The proof, that the fluctuation term y_i as generated by equation (7.38) truly generates the correct variance is given in Appendix A.3.

Note, that the application of the Wiener operator is a linear operation, and we can therefore rewrite equation (7.37) as:

$$s_i = s_i^* + \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} \left(d_l - d_l^* \right) , \tag{7.43}$$

where the Wiener operator is applied to the true data d_i and the mock observation d_i^* simultaneously. This greatly reduces the CPU time required for the generation of one signal sample.

7.5. Sampling the power-spectrum

As described above, the second step in the Gibbs sampling procedure (7.8) consists of drawing power-specrum samples from the conditional probability distribution $\mathcal{P}(S|\{s_i\},\{d_i\})$. The previous step of drawing a three dimensional signal sample yielded a noise-less full sky signal sample s_i consistent with the data. Since in this Gibbs sampling step the perfect sky signal amplitudes s_i are known, the power-spectrum is conditionally independent of the data amplitudes d_i. Hence, in this Gibbs sampling step, we can sample the power-spectrum from the probability distribution $\mathcal{P}(S|\{s_i\})$. In the following we will show that the power-spectrum can easily be drawn from an inverse gamma distribution.

7.5.1. Drawing power-spectrum samples

According to Bayes' theorem, we can rewrite the conditional probability $\mathcal{P}(S|\{s_i\})$ as:

$$\mathcal{P}(S|\{s_i\}) = \frac{\mathcal{P}(S)}{\mathcal{P}(\{s_i\})}\mathcal{P}(\{s_i\}|S)\,, \tag{7.44}$$

where $\mathcal{P}(S)$ is the prior for the signal covariance, $\mathcal{P}(\{s_i\}|S)$ is given by equation (7.11) and $\mathcal{P}(\{s_i\})$ is a normalization constant in this Gibbs sampling step.

More specifically, we are interested in the set of matrix coefficients $\{S_{ij}\}$ of the covariance matrix S. As already pointed out in section 7.4.1 the signal covariance matrix of an homogeneous and isotropic universe, has an especially appealing form in Fourier space, where it takes a diagonal form. In our application the real space covariance matrix coefficients $\{S_{ij}\}$ are related to their Fourier representation via the fast Fourier transform, as defined in Appendix A.1.6. We can therefore write:

$$\begin{aligned} S_{ij} &= C^2 \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} e^{2\pi i k \frac{\sqrt{-1}}{N}} \hat{\hat{S}}_{kl}\, e^{-2\pi j l \frac{\sqrt{-1}}{N}} \\ &= C^2 \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N}(ik-jl)} \hat{\hat{S}}_{kl}\,. \end{aligned} \tag{7.45}$$

Then we can express the conditional distribution for the Fourier signal covariance coefficients $\hat{\hat{S}}_{kl}$ as:

$$\mathcal{P}\left(\{\hat{\hat{S}}_{kl}\}|\{s_i\}\right) = \mathcal{P}\left(\{S_{ij}\}|\{s_i\}\right)\left|\frac{\partial\{S_{ij}\}}{\partial\{\hat{\hat{S}}_{kl}\}}\right|\,, \tag{7.46}$$

where

$$\left|\frac{\partial\{S_{ij}\}}{\partial\{\hat{\hat{S}}_{kl}\}}\right| = \left|\det(\mathcal{J}_{(ij)(kl)})\right|\,, \tag{7.47}$$

is the Jacobian determinant for this coordinate transformation. As the discrete Fourier transform is proportional to a unitary matrix, this Jacobian determinant only amounts to a normalization constant, as has been demonstrated in Appendix A.2.

With this definition we can rewrite the conditional probability in equation (7.44), by replacing all the real space covariance matrix coefficients S_{ij} by their Fourier representation $\hat{\hat{S}}_{kl}$, and normalizing it with the constant obtained from the coordinate transformation. We can therefore write:

$$\begin{aligned} \mathcal{P}(\{\hat{\hat{S}}_{kl}\}|\{s_i\}) &= \frac{\mathcal{P}(\{\hat{\hat{S}}_{kl}\})}{C^{N^2}\,\mathcal{P}(\{s_i\})}\mathcal{P}(\{s_i\}|\{\hat{\hat{S}}_{kl}\}) \\ &= \frac{\mathcal{P}(\{\hat{\hat{S}}_{kl}\})}{C^{N^2}\,\sqrt{\det\left(2\pi\hat{\hat{S}}\right)}\mathcal{P}(\{s_i\})}e^{-\frac{C^2}{2C^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\hat{s}_k^*\hat{\hat{S}}_{kl}^{-1}\hat{s}_l}\,, \end{aligned}$$

$$\tag{7.48}$$

where we used the discrete Fourier transform definition, given in Appendix A.1.6, to replace the real space signal amplitudes s_i by their Fourier counterparts $\hat{s}_k$. Introducing equation (7.12) then allows us to rewrite equation (7.48) in terms of the power-spectrum coefficients P_k as:

$$\mathcal{P}(\{P_k\}|\{s_i\}) = \frac{\mathcal{P}(\{P_k\})}{C^{N^2}\,\mathcal{P}(\{s_i\})} \prod_{k'=0}^{N-1} (2\pi\,P_{k'})^{-1/2}\, e^{-\frac{C^2}{2\hat{C}^2}\sum_{k=0}^{N-1}\frac{|\hat{s}_k|^2}{P_k}}\,,$$

where the determinant factorizes due to the diagonal form of the signal covariance matrix in Fourier space.

Note, that due to isotropy the power-spectrum is independent of direction in Fourier space, meaning the power-spectrum coefficients only depend on the modulus of the mode vector $\vec{k}_k$:

$$P_k = P(\vec{k}_k) = P(|\vec{k}_k|)\,. \tag{7.49}$$

For this reason, the angular dependence in Fourier space can be summed over.

To do so we remark that the mode vector $\vec{k}_k$, as a geometrical object, will not change if we express it in the basis of cartesian coordinates $\vec{k}_k = \vec{k}_k(k_k^1, k_k^2, k_k^3)$, or if we describe it in the basis of spherical coordinates $\vec{k}_k = \vec{k}_k(|\vec{k}_k|, \varphi_k, \vartheta_k)$. We can therefore split the multi index summation into the summation over the three spherical coordinates as:

$$
\begin{aligned}
\frac{C^2}{\hat{C}^2}\sum_{k=0}^{N-1}\frac{|\hat{s}_k|^2}{P_k} &= \sum_{|\vec{k}_k|}\frac{1}{P(|\vec{k}_k|)}\sum_{\varphi_k}\sum_{\vartheta_k}\frac{C^2}{\hat{C}^2}\left|\hat{s}\left(|\vec{k}_k|, \varphi_k, \vartheta_k\right)\right|^2 \\
&= \sum_{|\vec{k}_k|}\frac{\sigma(|\vec{k}_k|)}{P(|\vec{k}_k|)} \\
&= \sum_{m=0}^{M-1}\frac{\sigma_m}{P_m}\,,
\end{aligned}
\tag{7.50}
$$

where we introduced $\sigma(|\vec{k}_k|) = \sum_{\varphi_k}\sum_{\vartheta_k} C^2/\hat{C}^2 \left|\hat{s}\left(|\vec{k}_k|, \varphi_k, \vartheta_k\right)\right|^2$, which is the summed signal power on spherical shells around the origin in Fourier space, and the index m labels each of the M shells belonging to the different mode vector modulus $|\vec{k}_k|$ in the Fourier box.

Several different mode vectors $\vec{k}_k$ may have the same vector modulus $|\vec{k}_k|$, and therefore belong to the same shell. To account for this we introduce the number n_m, which counts the number of different mode vectors $\vec{k}_k$, belonging to the mth shell in Fourier space. This number n_m, therefore counts the degrees of freedom for each of the M modes. We can then express the product in equation (7.49) in terms of m as:

$$\prod_{k=0}^{N-1}(2\pi\,P_k)^{-1/2} = \prod_{m=0}^{M-1}(2\pi\,P_m)^{-n_m/2}\,. \tag{7.51}$$

With these definitions equation (7.49) turns to:

$$\mathcal{P}(\{P_k\}|\{s_i\}) = \frac{\mathcal{P}(\{P_k\})}{C^{N^2}\,\mathcal{P}(\{s_i\})}\prod_{m=0}^{M-1}(2\pi\,P_m)^{-n_m/2}\,e^{-\frac{1}{2}\frac{\sigma_m}{P_m}}\,, \tag{7.52}$$

When ignoring the power-spectrum prior $\mathcal{P}(\{P_k\})$ in the above equation (7.52), we see that the probability distribution factorizes in the different P_m, meaning they could be sampled independently.

If also the prior for the different P_m would factorize as:

$$\mathcal{P}(\{P_k\}) = \prod_{m=0}^{M-1}\mathcal{P}(P_m)\,, \tag{7.53}$$

then it is possible to sample each mode of the power-spectrum independently.

On large scales, or in the linear regime, the theory of gravitational structure formation tells us that the differ-

ent Fourier modes evolve independently. In these regimes the proposed power-spectrum prior would be the adequate choice. However, we also know that nonlinear structure formation couples the different Fourier modes, more strongly the deeper we reach into the nonlinear regime. In these regimes another prior would be more adequate, but also harder to sample.

Anyhow, as already described in section 7.3 the entire power-spectrum sampling method requires two steps. While the different power-spectrum modes are assumed to be independent in the power-spectrum sampling step, they are not in the signal sampling step. There the different modes are coupled via the observation mask and selection function, and furthermore, the physical coupling of the different modes is represented in the data.

Therefore, in the following we assume a power-spectrum prior, as proposed in equation (7.53), and defer a more thorough investigation of adequate prior choices in the nonlinear regime to future work.

With this prior choice each mode can be sampled independently from the following probability density distribution:

$$\mathcal{P}(P_m|\{s_i\}) = \frac{\mathcal{P}(P_m)}{\left(C^{N^2}\,\mathcal{P}(\{s_i\})\right)^{1/M}}\,(2\pi\,P_m)^{-n_m/2}\,e^{-\frac{1}{2}\frac{\sigma_m}{P_m}}\,. \tag{7.54}$$

Further, we will assume a power-law behavior for the individual mode prior $\mathcal{P}(P_m) \propto P_m^{-\alpha}$ where α is a power law index. Note, that a power-law index $\alpha = 0$ describes the flat prior, while $\alpha = 1$ amounts to the Jeffreys' prior. The Jeffreys' prior is a solution to a measure invariant scale transformation of the form $\mathcal{P}(P_m)dP_m = \mathcal{P}(\gamma\,P_m)\,\gamma dP_m$ (Wandelt 2004), and therefore is a scale independent prior, as different scales have the same probability.

Inserting this power law prior in equation (7.54) and imposing the correct normalization, reveals that the power-spectrum coefficients have to be sampled from an inverse gamma distribution given as:

$$\mathcal{P}(P_m|\{s_i\}) = \frac{\left(\frac{\sigma_m}{2}\right)^{(\alpha-1)+n_m/2}}{\Gamma\left((\alpha-1)+\frac{n_m}{2}\right)}\frac{1}{P_m^{(\alpha+n_m/2)}}e^{-\frac{1}{2}\frac{\sigma_m}{P_m}}\,. \tag{7.55}$$

By introducing the new variable $x_m = \sigma_m/P_m$ and performing the change of variables we yield the χ^2-distribution as:

$$\mathcal{P}(x_m|\{s_i\}) = \frac{x_m^{\beta_m/2-1}}{\Gamma\left(\beta_m/2\right)\,(2)^{\beta_m/2}}e^{-\frac{x_m}{2}}\,, \tag{7.56}$$

where $\beta_m = 2(\alpha + n_m/2 - 1)$. Sampling the power-spectrum coefficients is now an easy task, as it reduces to drawing random samples from the χ^2-distribution. A random sample from the χ^2-distribution for an integer β_m can be drawn as follows.

Let z_j be β_m independent, normally distributed random variates with zero mean and unit variance then:

$$x_m = \sum_{j=1}^{\beta_m} z_j^2 = |\vec{z}_m|^2 \tag{7.57}$$

is χ^2-distributed, and $\vec{z}_m$ is a β_m element vector, with each element being normally distributed. The power-spectrum coefficient sample is then obtained by:

$$P_m = \frac{\sigma_m}{|\vec{z}_m|^2}\,. \tag{7.58}$$

It is easy to see that each spectrum coefficient sample is a positive quantity, this ensures that the signal covariance matrix is positive definite as it has to be by definition.

To summarize, we provide an optimal inference procedure for the power-spectrum coefficients, and their uncertainties.

It is also worth mentioning, that the inverse gamma distribution is a highly non-Gaussian distribution, and that for this reason, the joint samples of signal amplitudes s_i and power-spectrum coefficients P_m are drawn from a non-Gaussian distribution.

7.5.2. Blackwell-Rao estimator

As described in the introduction, we seek to estimate the power-spectrum posterior $\mathcal{P}(\{P_m\}|\{d_i\})$, which we can now simply obtain by marginalizing over the signal samples:

$$
\begin{aligned}
\mathcal{P}(\{P_m\}|\{d_i\}) &= \int d\{s_i\}\, \mathcal{P}(\{P_m\}|\{s_i\}, \{d_i\})\, \mathcal{P}(\{s_i\}|\{d_i\}) \\
&= \int d\{s_i\}\, \mathcal{P}(\{P_m\}|\{s_i\})\, \mathcal{P}(\{s_i\}|\{d_i\}) \\
&\approx \frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \mathcal{P}(\{P_m\}|\{s_i\}^j)\,,
\end{aligned}
\tag{7.59}
$$

where $\{s_i\}^j$ are the signal Gibbs samples, and N_{Gibbs} is the total number of Gibbs samples.

This result is known as the Blackwell-Rao estimator of $\mathcal{P}(\{P_m\}|\{d_i\})$ which is guaranteed to have a lower variance than a binned estimator (Wandelt 2004).

It is worth noting, that $\mathcal{P}(\{P_m\}|\{s_i\})$ has a very simple analytic form, and therefore equation (7.63) provides an analytic approximation to $\mathcal{P}(\{P_m\}|\{d_i\})$ based on the Gibbs samples. All the information on $\mathcal{P}(\{P_m\}|\{d_i\})$ is therefore contained in the σ_m of the individual Gibbs steps, which generate a data set of size ($m_{max} N_{Gibbs}$), where m_{max} is the maximal number of independent modes. In addition, to being a faithful representation of $\mathcal{P}(\{P_m\}|\{d_i\})$ the Blackwell-Rao estimator is also a computationally efficient representation, which allows to calculate any moment of $\mathcal{P}(\{P_m\}|\{d_i\})$ as:

$$
\langle P_m\, P_{m'} ... P_{m''} \rangle|_{\mathcal{P}(P_m|\{d_i\})} \approx \frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \langle P_m\, P_{m'} ... P_{m''} \rangle|_{\mathcal{P}(P_m|\{s_i\}^j)}\,,
\tag{7.60}
$$

where each of the terms on the right handside can be calculated analytically.

For the inverse gamma distribution given in equation (7.55) we can then simply calculate the mean of the probability distribution $\mathcal{P}(P_m|\{d_i\})$ as:

$$
\langle P_m \rangle|_{\mathcal{P}(P_m|\{d_i\})} \approx \frac{\frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \sigma_m^j}{2(\alpha-2)+n_m}\,,
\tag{7.61}
$$

and in analogy the variance as:

$$
\left\langle [P_m - \langle P_m \rangle]^2 \right\rangle|_{\mathcal{P}(P_m|\{d_i\})} \approx \frac{\frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \left(\sigma_m^j\right)^2}{4\,((\alpha-2)+n_m/2)^2\,((\alpha-3)+n_m/2)}\,.
\tag{7.62}
$$

The Blackwell-Rao estimator also allows us to demonstrate another remarkable property of the Gibbs sampling approach. Although a specific power-spectrum prior has to be employed during the Gibbs analysis of the data, a post processing analysis of the power-spectrum can be performed with any desired power-spectrum prior. Lets assume one prefers to perform a post processing analysis with a power-spectrum prior $\mathcal{P}'(\{P_m\})$, rather than with the prior $\mathcal{P}(\{P_m\})$, which was employed during Gibbs sampling. We therefore want to infer the power-spectrum from the following posterior:

$$
\begin{aligned}
\mathcal{P}'(\{P_m\}|\{d_i\}) &= \mathcal{P}'(\{P_m\}) \frac{\mathcal{P}(\{d_i\}|\{P_m\})}{\mathcal{P}(\{d_i\})} \\
&= \frac{\mathcal{P}'(\{P_m\})}{\mathcal{P}(\{P_m\})} \mathcal{P}(\{P_m\}|\{d_i\}) \\
&= \frac{\mathcal{P}'(\{P_m\})}{\mathcal{P}(\{P_m\})} \int d\{s_i\}\, \mathcal{P}(\{P_m\}|\{s_i\})\, \mathcal{P}(\{s_i\}|\{d_i\}) \\
&= \int d\{s_i\}\, \mathcal{P}'(\{P_m\}) \frac{\mathcal{P}(\{s_i\}|\{P_m\})}{\mathcal{P}(\{s_i\})} \mathcal{P}(\{s_i\}|\{d_i\}) \\
&\approx \frac{1}{N_{Gibbs}} \sum_{j=1}^{N_{Gibbs}} \mathcal{P}'(\{P_m\}) \frac{\mathcal{P}(\{s_i\}^j|\{P_m\})}{\mathcal{P}(\{s_i\}^j)}\,,
\end{aligned}
\tag{7.63}
$$
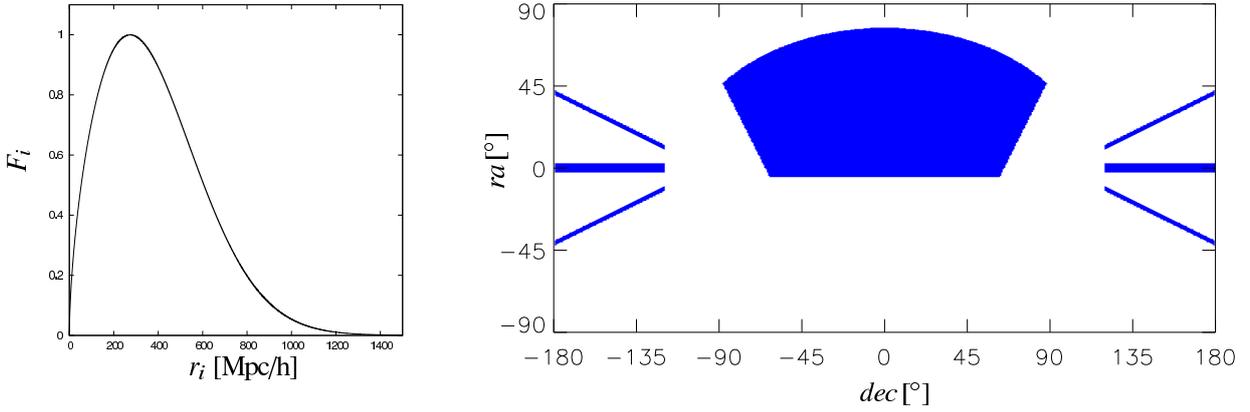
Figure 7.3.: Selection function and two dimensional sky mask. The observational mask was designed to represent some of the prominent features of the Sloan Digital Sky Survey geometry of the seventh data release (see e.g. Abazajian et al. 2009).

where we simply made use of the Bayes theorem. Since $\mathcal{P}(\{s_i\}|\{P_m\})$ is a simple Gaussian distribution, and therefore given analytically, the posterior $\mathcal{P}'(\{P_m\}|\{d_i\})$ can be calculated with any desired power-spectrum prior in a post-processing step.

7.6. The prior and the cosmic variance

The Gibbs sampling procedure consists of the two basic steps, of first sampling perfect noise-less full sky signal samples s_i and then sampling the power-spectrum coefficients P_m given s_i.

Therefore, the probability distribution $\mathcal{P}(P_m|\{s_i\})$, given in equation (7.55), encodes our knowledge on the power-spectrum coefficients P_m, if we had perfect knowledge of the true signal amplitudes s_i.

It is clear, that in the case of perfect observations the full posterior distribution for the power-spectrum coefficients $\mathcal{P}(P_m|\{d_i\})$ would reduce to that one given in equation (7.55).

This is, because in the case of perfect full-sky and noise-less observations, the signal posterior would collapse to a Dirac delta distribution, due to the vanishing noise covariance matrix. This means, that in this case the signal amplitudes s_i can be inferred with zero variance.

However, measuring the P_m to arbitrary precision will never be possible. The power-spectrum coefficients depend on the data through the σ_m, which measure the actual fluctuation power in the observed Universe. It is clear that the probability distribution function (7.55) for the P_m will not reduce to a Dirac delta distribution, even though the σ_m have been measured perfectly.

Owing to this fact, there will always remain some uncertainty in the power-spectrum inference, even in the case of perfect measurements. This residual uncertainty is well known as cosmic variance, which is the direct consequence of observing just one specific matter field realization.

The Gibbs sampling approach, as proposed here, takes this cosmic variance into account, by drawing samples from the probability distribution $\mathcal{P}(P_m|\{s_i\})$, which obey the correct statistical properties.

7.6.1. Flat versus Jeffreys' prior

The main characteristics of $\mathcal{P}(P_m|\{s_i\})$ can be summarized in terms of the mean, mode and variance, which allows us to discuss the influence of the actual power law prior choice. The mode of the inverse gamma distribution (7.55) is the most frequently used estimator for the power-spectrum coefficients when using fast Fourier transform techniques. The mode P_m^* is defined as $P_m^* \in \{P_m | \forall P_l : \ \mathcal{P}(P_l|\{s_i\}) \leq \mathcal{P}(P_m|\{s_i\})\}$, which is simply the value of P_m

which maximizes the distribution. For the inverse gamma distribution it is given as:

$$P_m^* = \frac{\sigma_m}{(2\alpha + n_m)} \ . \tag{7.64}$$

Assuming a flat prior $\alpha = 0$, this immediately returns the frequently used and simple power-spectrum estimator:

$$P_m^* = \frac{\sigma_m}{n_m} \ , \tag{7.65}$$

(see e.g. Cui et al. 2008).

However, note, that a flat prior is a questionable choice, when measuring a variance, which is a scale parameter, as it does not correspond to maximal ignorance but biases towards large excursions from zero (Wandelt 2004). Additionally the flat prior does not permit one to sample every mode in the three dimensional Fourier box with finite variance, which can be easily seen by looking at the variance of the inverse gamma distribution given as:

$$\left\langle [P_m - \langle P_m \rangle]^2 \right\rangle = \frac{\sigma_m^2}{4\left((\alpha - 2) + n_m/2\right)^2\left((\alpha - 3) + n_m/2\right)} \tag{7.66}$$

which is only finite for $2\alpha + n_m > 6$.

In a three dimensional cubic Fourier box the minimal number n_m for a mode is $min(n_m) = 6$ (except for the zero mode). This corresponds to the six mode vectors $\vec{k}_k$ with same vector modulus $|\vec{k}_k|$ along the three axes in Fourier space. Nevertheless, a flat prior ($\alpha = 0$) requires $n_m > 6$ in order to sample the modes with finite variance, which cannot be fulfilled for these modes.

Therefore, we favor the Jeffreys' prior with $\alpha = 1$, which requires only $n_m > 4$ to sample each mode, except for the zero mode, with finite variance. Jeffreys' prior is also scale invariant, and therefore does not introduce any bias on a log-scale.

7.6.2. Informative prior

The prior discussed in the previous section is a maximal ignorance prior in the sense, that every scale has the same probability. This prior therefore allows for large excursions around the true value of the power-spectrum. This is especially important when sampling the largest scales in a galaxy survey, which are poorly constrained by measurements. A maximum ignorance prior will therefore require sampling a huge space of possible power-spectrum configurations.

However, one can argue, that knowledge about the largest scales exists, either through theory, or CMB measurements, which provide detailed information on the largest scales.

For this reason, it might be interesting to incorporate this a priori knowledge on the power-spectrum into our sampling scheme, and therefore allow for a more efficient strategy to sample the space of possible power-spectra.

The most simple informative prior can be obtained by limiting the range of the Jeffreys' prior, by setting the Jeffreys' prior equal to zero for power-spectrum excursion of more than a certain factor:

$$\mathcal{P}(P_m) \quad \propto \quad \begin{cases} P_m^{-\alpha} & \text{for } P_m^{Prior}/\tau \leq P_m \leq P_m^{Prior}\,\tau \\ 0 & \text{otherwise} \end{cases} \tag{7.67}$$

where P_m^{Prior} is our best guess power-spectrum prior, and τ is a factor, which permits a certain range around the prior power-spectrum. The sampling scheme, which arises by implementing this prior, is basically the same as described in section 7.5.1, with the only exception that power specrum coefficients P_m, which do not fulfill the requirement described in equation (7.67), are rejected and have to be resampled. This prior would still be a maximum ignorance prior over the allowed range.

Another possible informative prior, which allows to sample the entire range, but favoring the region, in which we expect the true power-spectrum to exist, can be very easily found by assuming some a priori knowledge on the σ_m. For example, this can be achieved by incorporating an independent observation to the sampling scheme. In this case

we can again assume an inverse gamma distribution for the power-spectrum prior:

$$
\mathcal{P}(P_m) = \frac{\left(\frac{\sigma_m^{Prior}}{2}\right)^{(\alpha^{Prior}-1)+n_m^{Prior}/2}}{\Gamma\left((\alpha^{Prior}-1)+\frac{n_m^{Prior}}{2}\right)} \frac{1}{P_m^{(\alpha^{Prior}+n_m^{Prior}/2)}} e^{-\frac{1}{2}\frac{\sigma_m^{Prior}}{P_m}} , \tag{7.68}
$$

where σ_m^{Prior} describes our a priori knowledge on the σ_m, α^{Prior} is the spectral index of of our power law prior, which we choose $\alpha^{Prior} = 1$ to be the Jeffreys' prior, and n_m^{Prior} is the number of mode counts for our prior guess. Note, that the combination of α^{Prior} and n_m^{Prior} defines how sharp this prior would be. As we want our prior to contain as little information as possible, we choose $n_m^{Prior} = 5$ as this is the minimal number of modes, which lead to a finite variance with the Jeffreys' prior.

Introducing an inverse gamma prior will then yield again a inverse gamma distribution for the power-spectrum sampling procedure:

$$
\mathcal{P}(P_m|\{s_i\}) = \frac{\left(\frac{\sigma_m^{Prior}+\sigma_m}{2}\right)^{(\alpha^{Prior}-1)+(n_m^{Prior}+n_m)/2}}{\Gamma\left((\alpha^{Prior}-1)+\frac{(n_m^{Prior}+n_m)}{2}\right)} \frac{e^{-\frac{1}{2}\frac{(\sigma_m^{Prior}+\sigma_m)}{P_m}}}{P_m^{(\alpha^{Prior}+(n_m^{Prior}+n_m)/2)}} . \tag{7.69}
$$

By introducing $x_m = (\sigma_m^{Prior}+\sigma_m)/P_m$, this can again be rewritten as a χ^2-distribution:

$$
\mathcal{P}(x_m|\{s_i\}) = \frac{x_m^{\beta_m/2-1}}{\Gamma(\beta_m/2)\,(2)^{\beta_m/2}} e^{-\frac{x_m}{2}} , \tag{7.70}
$$

where $\beta_m = 2(\alpha^{Prior}+(n_m^{Prior}+n_m)/2-1)$.

A power-spectrum sample is then obtained in the same fashion as described in section 7.5.1, by

$$
P_m = \frac{\sigma_m^{Prior}+\sigma_m}{|\vec{z}_m|^2} , \tag{7.71}
$$

with $\vec{z}_m$ being β_m element vector, with each element being normally distributed with zero mean and unit variance.

As an example of incorporating theoretical information one can generate the σ_m^{Prior} by:

$$
\sigma_m^{Prior} = n_m^{Prior} P_m , \tag{7.72}
$$

which will yield on average the prior power-spectrum. In this fashion the Gibbs sampler will sample around our prior guess for the power-spectrum.

Note, that this method provides a possible interface for joint power-spectrum inference from a joint CMB and large scale structure analysis, where the $\sigma_m^{Prior} = \sigma_m^{CMB}$ will be obtained from a CMB analysis step.

7.6.3. Hidden prior

In the previous section, we described how to sample each mode $|\vec{k}_k|$ of the Fourier box individually, which yields extremely fine frequency resolution in the power-spectrum sample. As in practice such high-frequency resolution is not required, or desired, one allows each shell m' to have a finite thickness $\Delta k'_m$, rather than treating it as infinitely thin.

As the newly designated shells m' have a finite thickness, different infinitely thin shells m now belong to the same shell m'.

With the shells having a finite thickness, different close by Fourier modes $|\vec{k}_k|$ fall into the same bin, and therefore an assumption about the functional shape $f_m(|\vec{k}_k|)$, which yields the correct weighting, over this shell m' around the central mode $|\vec{k}|_{m'}$ must be assumed:

$$
P_m = A_{m'} f_m^{m'} \text{ for } \vec{k}_k \in \left[|\vec{k}|_m - \Delta k_m/2, |\vec{k}|_m \Delta k_m/2\right] , \tag{7.73}
$$

where $A_{m'}$ is the constant amplitude for the $m'th$ shell. Usually, the shape of the power-spectrum $f_m^{m'}$ is assumed to be constant over the shell width $\Delta k'_m$, which amounts to "binning" the power-spectrum. However, in general $f_m^{m'}$
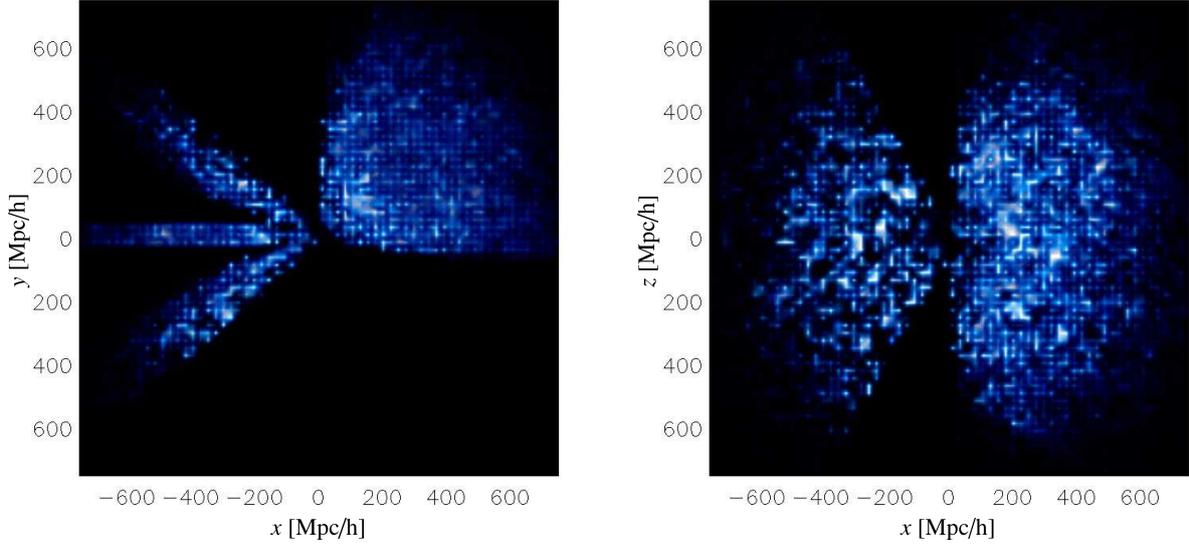
Figure 7.4.: Volume rendering of observed galaxy densities in two different projections. This mock galaxy observation has been generated according to the Gaussian model described in section 7.4.3.

could assume any desired functional shape.

With this assumption, the power-spectrum sample is further constrained by the assumed functional shape $f_m^{m'}$, and we seek to infer the set of power-spectrum coefficients $\{P_m\}$ from a probability distribution given as:

$$
\begin{aligned}
\mathcal{P}(\{P_m\}|\{s_i\},\{f_m^{m'}\}) &= \frac{\mathcal{P}(\{f_m^{m'}\}|\{P_m\})}{\mathcal{P}(\{f_m^{m'}\}|\{s_i\})} \frac{\mathcal{P}(\{P_m\})\mathcal{P}(\{s_i\}|\{P_m\},\{f_m^{m'}\})}{\mathcal{P}(\{s_i\})} \\
&= \frac{\mathcal{P}(\{f_m^{m'}\}|\{P_m\})}{\mathcal{P}(\{f_m^{m'}\}|\{s_i\})} \frac{\mathcal{P}(\{P_m\})\mathcal{P}(\{s_i\}|\{P_m\})}{\mathcal{P}(\{s_i\})},
\end{aligned}
\tag{7.74}
$$

where we assumed conditional independence $\mathcal{P}(\{s_i\}|\{P_m\},\{f_m^{m'}\}) = \mathcal{P}(\{s_i\}|\{P_m\})$ of the functional shape, once all power-spectrum coefficients are given.

The usual implicit assumption when introducing this kind of power-spectrum binning is:

$$
\mathcal{P}(\{f_m^{m'}\}|\{P_m\})_s \propto \prod_m \delta^D(A_m' f_m^{m'} - P_m),
\tag{7.75}
$$

meaning, we claim to know the exact functional shape of the power-spectrum within the shells m'.

This reduces the amount of free parameters to be sampled, from the set of N_m power-spectrum coefficients P_m to the set of $N_{m'}$ amplitudes $A_{m'}$.

If, for instance, we knew the exact shape of the entire power-spectrum, sampling the power-spectrum could be reduced to the task of just sampling the overall amplitude. Such an approach is pursued, when trying to sample the cosmological parameters.

However, with the above definition, sampling the power-spectrum reduces to sample the amplitudes $A_{m'}$, and we

can write:

$$
\begin{aligned}
\mathcal{P}(\{A_{m'}\}|\{s_i\},\{f_m^{m'}\}) &= \mathcal{P}(\{P_m\}|\{s_i\},\{f_m^{m'}\}) \left| \frac{\partial \{P_m\}}{\partial \{A_{m'}\}} \right| \\
&= \prod_m \frac{\left(\frac{\sigma_m}{2\,f_m^{m'}}\right)^{(\alpha-1)+n_m/2}}{\Gamma\left((\alpha-1)+\frac{n_m}{2}\right)} \frac{e^{-\frac{1}{2}\frac{\sigma_m}{A_{m'}\,f_m^{m'}}}}{(A_{m'})^{(\alpha+n_m/2)}} \\
&= \prod_{m'} \prod_{m\in m'} \frac{\left(\frac{\sigma_m}{2\,f_m^{m'}}\right)^{(\alpha-1)+n_m/2}}{\Gamma\left((\alpha-1)+\frac{n_m}{2}\right)} \frac{e^{-\frac{1}{2}\frac{\sigma_m}{A_{m'}\,f_m^{m'}}}}{(A_{m'})^{(\alpha+n_m/2)}} \\
&= \prod_{m'} \frac{\left(\frac{\sigma_{m'}}{2}\right)^{(\alpha-1)+n_{m'}/2}}{\Gamma\left((\alpha-1)+\frac{n_{m'}}{2}\right)} \frac{e^{-\frac{1}{2A_{m'}}\sum_{m\in m'}\frac{\sigma_m}{f_m^{m'}}}}{(A_{m'})^{(\alpha+n_{m'}/2)}} \\
&= \prod_{m'} \frac{\left(\frac{\sigma_{m'}}{2}\right)^{(\alpha-1)+n_{m'}/2}}{\Gamma\left((\alpha-1)+\frac{n_{m'}}{2}\right)} \frac{e^{-\frac{1}{2}\frac{\sigma_{m'}}{A_{m'}}}}{(A_{m'})^{(\alpha+n_{m'}/2)}} \,,
\end{aligned}
$$
(7.76)

where $\sigma_{m'} = \sum_{m\in m'} \sigma_m/f_m^{m'}$ and $n_{m'} = \sum_{m\in m'} n_m$. As this probability distribution factorizes in the amplitudes $A_{m'}$, each of these amplitudes can be independently sampled from the inverse gamma distribution, with the method described in section 7.5.1.

A power-spectrum coefficient P_m is therefore obtained as:

$$
P_m = \frac{\sigma_{m'}\,f_m^{m'}}{|\vec{z}_{m'}|^2} \,,
$$
(7.77)

where $\vec{z}_{m'}$ is a $n_{m'}$ component vector, with the elements being Gaussian distributed with zero mean and unit variance.

Note, since $n_{m'} \geq n_m$ the variance for the power-spectrum coefficients P_m is reduced. This is the result of reducing the amount of free parameters, by introducing "binning", as now each amplitude estimate for the $A_{m'}$ is based on more supporting points than the individual P_m.

Due to the finite shell width $\Delta k_{m'}$ neighboring modes are coupled. This fact could be exploited to circumvent the problem of missing mode coupling in the nonlinear regime. For example, if the different shells would be logarithmically spaced, it is possible to sample the largest scales independently, while towards the non-linear regime, the modes get more and more coupled. From a physical point of view, the logarithmic spacing would therefore be best suited for this problem. On the other hand, introducing "binning" to the power-spectrum sampling procedure, makes the method insensitive to fluctuations on scales smaller than the shell width $\Delta k_{m'}$, and a $\Delta k_{m'}$ should therefore be chosen carefully in order not to miss features we intend to recover.

It is also important to note, that the variance in the power-spectrum sampling step defines the stepsize for the random walk, to sample the joint probability distribution of signal and power-spectrum. If the "binning" is unreasonable large, and therefore variance is dramatically reduced, it takes much longer to explore the joint space of signal amplitudes s_i and power-spectrum coefficients P_m. For this reason, we prefer to sample with rather high spectral resolution for the power-spectrum.

7.7. Numerical Implementation

Our numerical implementation of the large scale structure Gibbs sampler is called ARES (Algorithm for REconstruction and Sampling). It can be separated into the described two Gibbs sampling steps of inferring a signal sample, which involves solving large systems of equations, and sampling the power-spectrum, by drawing random samples from the inverse gamma distribution.
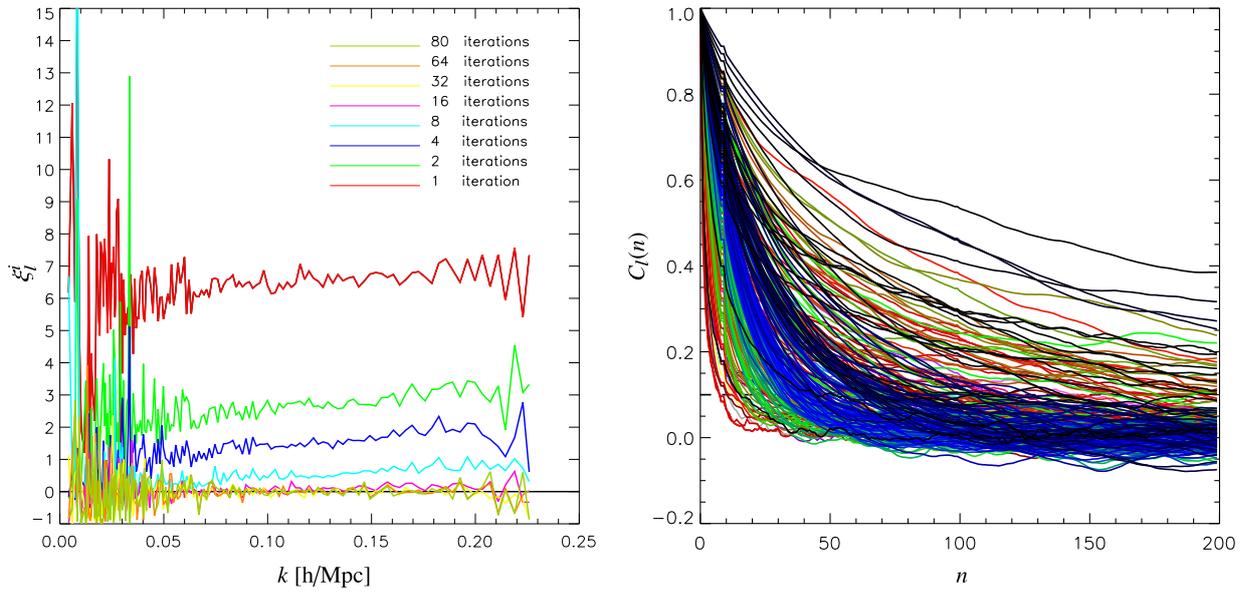
Figure 7.5.: Test results from low resolution test run. The left panel shows the results of the Burn-in test for eight sequential ξ_l^i as a function of Fourier modes k (see equation (8.27)). It clearly depicts the systematic drift and convergence of the successive power-spectra during the Burn-in period towards the true underlying power-spectrum. The right panel displays results for the correlation coefficients $C_l(n)$ of the correlation length test for all Fourier modes. As one can see, the vast majority of the different Fourier modes have a correlation length of about $n_C \sim 100$ Gibbs iterations.
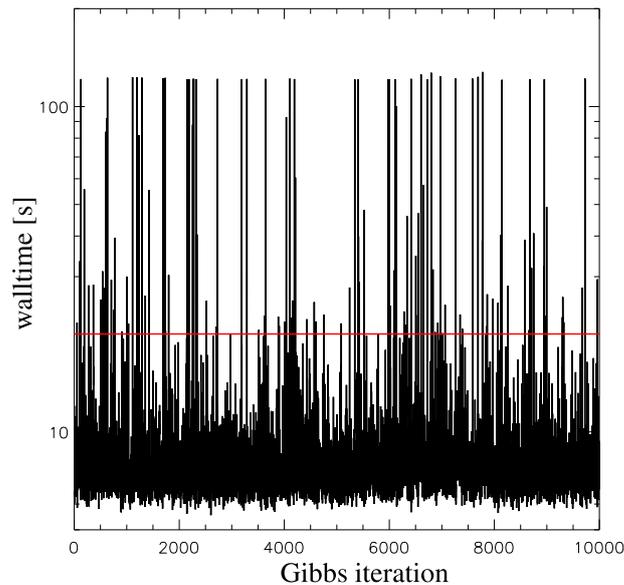
Figure 7.6.: Wall clock times for a single map making step over 10000 Gibbs iterations. The red line denotes the average wall clock time for one matrix inversion step.

7.7.1. Inversion of matrices

The signal sampling step is by far the most numerically demanding step as it requires fast and efficient inversions of large matrices. ARES utilizes the fast operator based conjugate gradients inversion technique as presented in the ARGO code (Kitaura & Enßlin 2008), which has recently been applied to Sloan Digital Sky Data, to obtain matter field reconstructions (Kitaura et al. 2009). Operator based inversion techniques have been previously developed for CMB data analysis (Wandelt et al. 2004, Wandelt 2004, Eriksen et al. 2004, Jewell et al. 2004).

Rather than requiring to store the matrices under consideration explicitly in computer memory, which would be numerically prohibitive, this approach only requires to know how these matrices would act on a vector, and therefore reduces the required amount of computer memory to numerically feasible amounts. Further, it is possible to reduce the scaling of the most expensive matrix operation to that of a fast Fourier transform. ARES uses the FFTW3 library, which therefore reduces the scaling of the most expensive operation to $\mathcal{O}(N\,log(N))$ (Frigo & Johnson 2005).

The FFTW3 library also incorporates the feature of parallel Fourier transforms, which allows for straightforward parallelization of our code.

7.7.2. Random number generation

Our random number generation relies on a pseudo random number generator as provided by the GNU scientific library (gsl) (Galassi et al. 2003). In particular, we use the Mersenne Twister MT19937, with 32-bit word length, as provided by the gsl_rng_mt19937 routine.

The Mersenne Twister algorithm was designed for Monte Carlo simulations, where primarily good quality numbers and speed are decisive (Matsumoto & Nishimura 1998). It has been proven to have a period of $2^{19937} - 1$, where for our application in practice we usually require only about 2^{35} unique random numbers. Also note, that the very high order of dimensional equidistribution guarantees negligible serial correlation in the output sequence. The Mersenne Twister algorithm passed several tests for statistical randomness, including the Diehard tests.

7.7.3. Parallelization

Parallelization of the code is a crucial issue, since CPU time is the main limiting factor of our method. Even though the conjugate gradient method allows for very efficient matrix inversions, for a complete data analysis one has to perform many of these matrix inversions, and therefore the total prefactor of the algorithm increases.
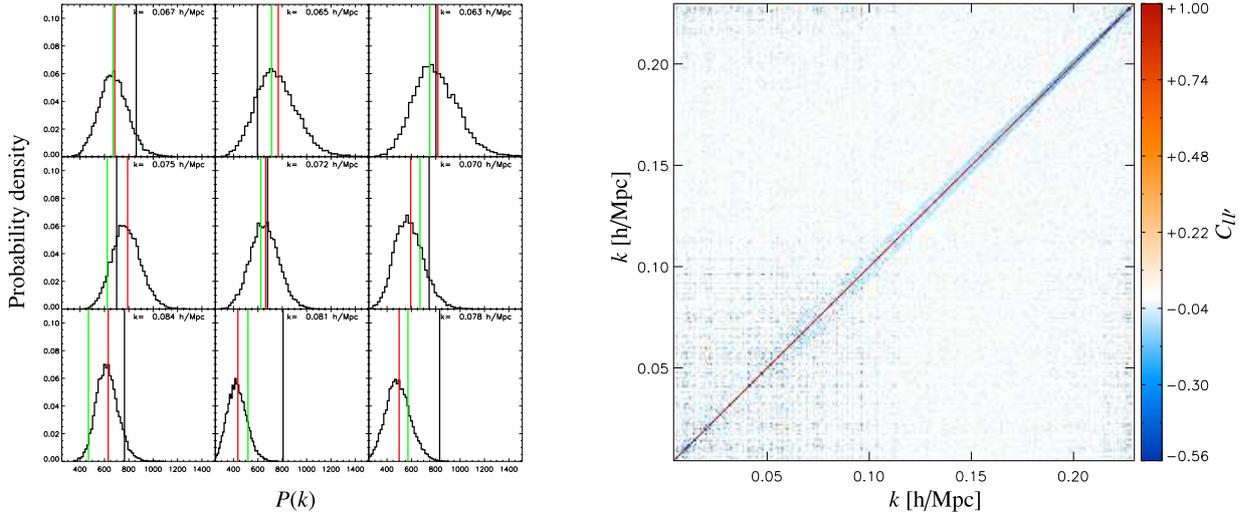
Figure 7.7.: The left panel shows the mode statistics for some selected Fourier modes. The green vertical line denotes the true underlying power-spectrum. The red line denotes the averaged mean obtained from the Gibbs chain, and the black vertical line denotes the power-spectrum of the specific matter field realization. The right panel shows the power-spectrum covariance matrix estimated from the 40,000 Gibbs samples.

In Fig. 7.6 we show a typical progression for the wall clock times of the map making algorithm during the Gibbs sampling chain for a low resolution simulation with $N_{VOX} = 64^3$ and a setup as described in section 7.8.

As can be seen, the wall clock times may vary from Gibbs iteration to Gibbs iteration.

This variation is due to complexity of the actual problem in the Gibbs iteration, which may require more conjugate gradients steps to reach the desired numerical accuracy.

The average creation time for one Gibbs sample in this low resolution simulation is about 20 seconds. Where this test was run on a single Desktop CPU. Doubling the resolution will require roughly eight times longer to create a sample. Hence, a $N_{VOX} = 128^3$ simulation will require roughly 5 minutes and a $N_{VOX} = 512^3$ about 8 hours to create a single Gibbs sample. Resolution also defines the range of wavelength $k = 0 \dots \sqrt{3} \pi N/L$ which can be probed. A larger range in wavelength also requires higher resolution. This immediately clarifies the need to parallelize the code, since usually tens of thousands of Gibbs samples are required.

There are in principle two different approaches to parallelize the code.

The most demanding step in the Gibbs sampling chain is the map making process. One could therefore parallelize the map making algorithm, which in principle requires to parallelize the fast Fourier transform. The FFTW3 library provides parallelized fast Fourier transform procedures, and implementation of those is straightforward (Frigo & Johnson 2005). However, optimal speed up cannot be achieved. The other approach relies on the fact that our method is a Monte Carlo process, and each CPU can therefore calculate its own Markov chain. In this fashion we gain optimal speed up and the possibility to initialize each chain with different initial guesses.

The major difference between these two parallelization approaches is, that with the first method one tries to calculate a rather long sampling chain, while the latter one produces many shorter chains.

As we will see in the next section, successive Gibbs samples are highly correlated in the low signal to noise regime and producing a larger number of independent samples requires longer chains. This problem, however, can be partially overcome by initializing each Markov chain with an independent power-spectrum guess. Nevertheless, the Gibbs sampling procedure is a numerically demanding process and the spatial resolution as well as the wavelength range of the recovered density fields and power-spectra will be limited by the amount of computational time available for the procedure. However, our method aims at recovering the largest scales rather than inferring the statistical properties of the density field deep into the nonliner regime. Therefore, the spatial resolution for the recovered density fields as well as the wavelength ranges for the power-spectra which can be covered by the Gibbs sampling procedure with realistic CPU availability will be sufficient for this purpose.

7.8. Testing ARES

In this section, we apply ARES to simulated mock observations, where the underlying dark matter signal is perfectly known. With these mock observations we will be able to demonstrate that the code produces results consistent with the theoretical expectation. In addition, we will gain insight in how the code may perform in real-world applications, when CPU time is limited. Therefore, we will set up Gaussian mock cases, designed to highlight some specific feature of the code.

7.8.1. Setting up a Gaussian Mock observation

For the cases studied in this section we set up a set of low resolution mock observations based on Gaussian random fields, which will allow us to calculate a large number of samples in reasonable computational times.

These mock observations are generated on a three dimensional cartesian box with $N_{side} = 64$, corresponding to $N_{vox} = 262144$ volume elements, and a box length of $L = 1500$ Mpc, with the observer positioned at the center.

The mock observations are generated according to the procedure described in section 7.4.3, with the underlying cosmological power-spectrum being calculated, with baryonic wiggles, following the prescription described in Eisenstein & Hu (1998) and Eisenstein & Hu (1999). For these simulations we assumed a standard ΛCDM cosmology with the set of cosmological parameters ($\Omega_m = 0.24$, $\Omega_\Lambda = 0.76$, $\Omega_b = 0.04$, $h = 0.73$, $\sigma_8 = 0.74$, $n_s = 1$).

To generate the uncorrelated noise component, we assume a noise structure function $n_i^{SF} = M_i F_i / \bar{\lambda}$, with $\bar{\lambda} = 8.0 \times 10^{-3} L^3/N_{vox}$ in voxel space. Then we draw random Poission samples via the procedure described in section 7.4.3.

Note, that this is equivalent to drawing a galaxy distribution from the ensemble mean dark matter density, which has been obtained by averaging over all possible matter field realizations.

The survey properties are described by the galaxy selection function F_i and the observation Mask M_i. The selection function is given by:

$$F_i = \left(\frac{r_i}{r_0}\right)^b \left(\frac{b}{\gamma}\right)^{-b/\gamma} e^{\frac{b}{\gamma} - \left(\frac{r_i}{r_0}\right)^\gamma} , \tag{7.78}$$

where r_i is the co-moving distance from the observer to the center of the ith voxel. For our simulation we chose parameters $b = 0.6$, $r_0 = 500$ Mpc and $\gamma = 2$.

In Fig. 7.4 we show the selection function together with the sky mask, which defines the observed regions in the sky. The two dimensional sky mask is given in sky coordinates of right ascension and declination. The projection of this two dimensional mask into the three dimensional volume yields the three dimensional mask M_i.

Two different projections of this generated mock galaxy survey are presented in Fig. 7.4 to give a visual impression of the artificial low resolution observation.

7.8.2. Testing convergence and correlations

The theory of Gibbs sampling states that the individual Gibbs samples converge to being samples from the joint probability distribution. However, the theory itself does not provide any criterion to detect when the samples start being samples from the joint probability distribution. Therefore, this initial burn-in behavior has to be tested by experiments. We will follow a similar precedure to Eriksen et al. (2004), to estimate the convergence behavior and the correlation length of the Gibbs samples.

This analysis is important, as it allows for the understanding of the code performance in real world applications, which is particularly relevant for estimating how many Gibbs samples are required for an accurate power-spectrum inference.

We study the initial burn-in behavior by setting up a simple experiment, in which we set the initial guess for the power-spectrum to be ten times larger in amplitude than the true underlying power-spectrum. Then ARES is applied to the mock observation, as described in the previous section 7.8.1, to calculate a number of Gibbs iterations. The obtained power-spectrum samples P_l^i of the ith iteration are then compared to the true power-spectrum via:

$$\xi_l^i = \frac{P_l^i - P_l^{true}}{P_l^{true}} , \tag{7.79}$$
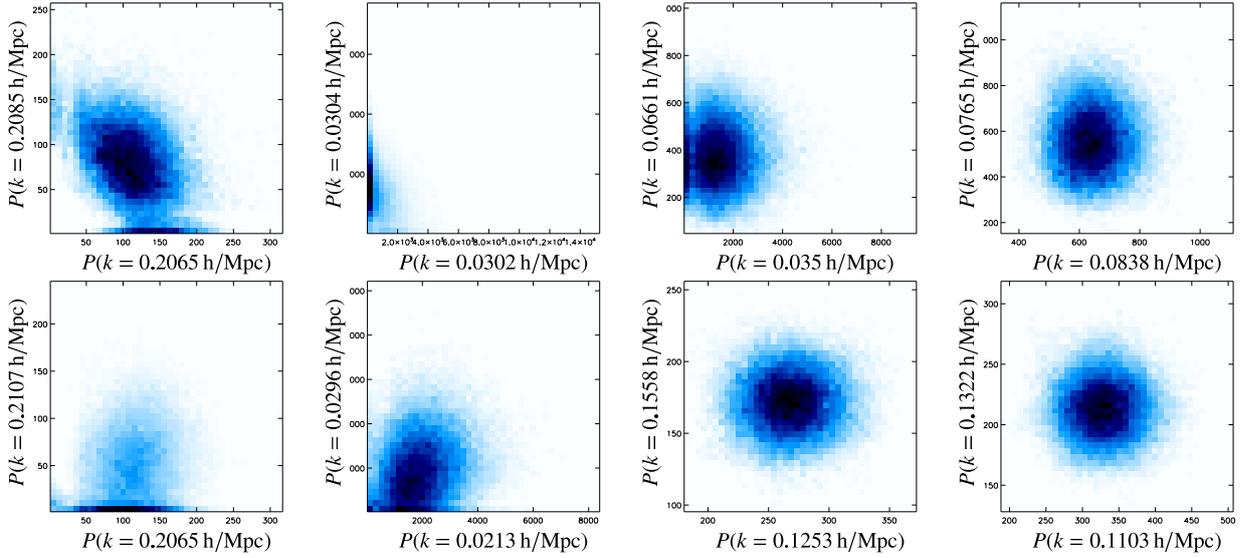
Figure 7.8.: 2D marginalized posterior densities. Each plot shows the full joint posterior of the data, integrated over all dimensions except for the two shown. The four panels to the left correspond to the maximally anti-correlated (upper panels) and maximally correlated (lower panels) Fourier modes. The four panels to the right show some random examples of the marginalized joint posterior distribution.

where the P_l^{true} are the true power-spectrum coefficients, with which the mock dark matter signal was simulated.

The results for the ξ_l^i are shown in Fig. 7.5. Here, we observe a systematic drift of the successive power-spectra towards the true underlying power-spectrum. Further we see that the initial burn-in phase, for the kind of setup as demonstrated here, is rather short. The algorithm requires about 30 Gibbs iteration after which we can assume the samples to be samples from the joint probability distribution. Also note, that in this test, we do not observe any particular hysteresis for the poorly constrained large scale modes, meaning they do not remain at their initially set values. This demonstrates the ability of the code to handle correctly the mode coupling introduced by the sky cut.

However, it is clear that a poor initial guess invalidates a certain number of samples, especially at large scales, where the uncertainty due to the sky mask dominates. For this reason, it is advantageous to initialize the Gibbs sampling chain with an initial guess, which is close to the true power-spectrum, to ensure short burn-in times. As can be seen in Fig. 7.5, any bad initial guess would reveal itself by a systematic drift in the Gibbs chain, and can therefore be detected easily.

Next, we want to analyze the correlation of the individual Gibbs samples in the sequence. This is a crucial point, as it permits us to estimate the number of independent samples, which can be obtained from a Gibbs chain of given length.

The correlation between sequential Gibbs samples can be best understood by reviewing the sampling algorithm.

A Gibbs sample of the joint probability distribution of signal and power-spectrum is obtained in two steps. In the first step a Wiener reconstruction is performed, based on the assumption of a given power-spectrum, and the power lost due to noise filtering, masks and selection effects is replaced by a fluctuation term. The signal obtained in this first sampling step mimics a full sky noise-less signal. It is clear, that the power-spectrum of this signal is determined by the data in the high signal-to-noise region and by the assumed power-spectrum in the low signal-to-noise region.

In the second step the power-spectrum sample is generated, based on this full sky noise-less signal sample, obtained in the previous signal sampling step. The obtained power-spectrum then works as input power-spectrum to the next Gibbs step, and the iteration starts again.

In this fashion the Gibbs sampler performs a random walk in the multi-dimensional space of signal map and power-spectrum. The stepsize of the power-spectrum sampling step is solely determined by cosmic variance, and does not take into account the noise variance, as described in section 7.5.1.

However, we want to probe the full probability distribution, which includes both noise and cosmic variance. This
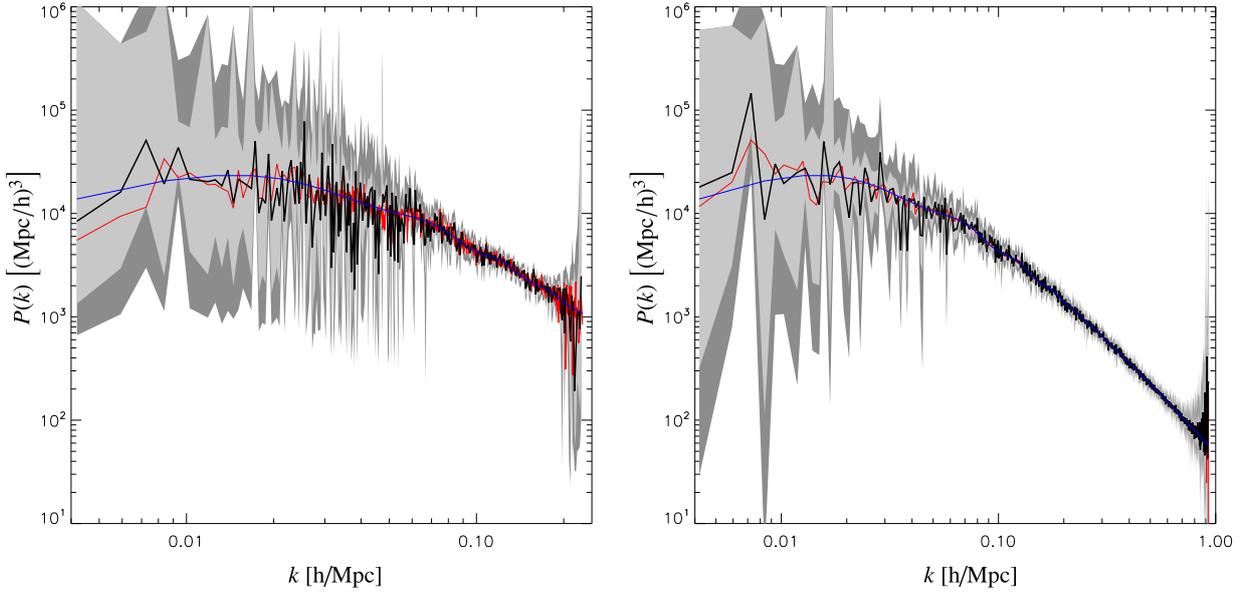
Figure 7.9.: Power-spectrum estimates obtained from the Gibbs sampling chain for a low resolution (left panel) and a high resolution run (right panel). The black lines represent the ensemble mean of the sample set and the light gray and dark gray shaded regions denote the one and two sigma confidence regions respectively. Additionally we show the according input power-spectra. The blue line shows the cosmological power-spectrum from which the matter field realization was drawn, and the red line is the power-spectrum of this specific matter field realization.

difference does not matter in the high signal to noise regime, since there cosmic variance will dominate the total variance, and for any practical purposes all Gibbs samples will be uncorrelated in this regime. This, however, is not true in the low signal to noise regime. Since the random stepsize between two subsequent samples is determined only by the cosmic variance, and not by the much larger noise variance, two sequential samples will be strongly correlated. In this case a longer Gibbs sequence is required to produce uncorrelated samples.

Reducing the variance by introducing binning to the power-spectrum, as described in section 7.6.3, will lead to even longer correlation length in the Gibbs chain. This simply means, the joint probability distribution will be sampled with a finer resolution in power-spectrum space.

We study this correlation effect by assuming the power-spectrum coefficients P_l of different modes l in the Gibbs chain to be independent and estimate their correlation in the chain by calculating the autocorrelation function:

$$C_l(n) = \left\langle \frac{P_l^i - \langle P_l \rangle}{\sqrt{VarP_l}} \frac{P_l^{i+n} - \langle P_l \rangle}{\sqrt{VarP_l}} \right\rangle, \tag{7.80}$$

where n is the distance in the chain measured in iterations (for a similar discussion in case of the CMB see Eriksen et al. 2004).

We can then define the correlation length of the Gibbs sampler as the distance in the chain n_C beyond which the correlation coefficient $C_l(n)$ has dropped below 0.1.

The results for the different modes l are presented in Fig. 7.5. As one can see, the vast majority of the different Fourier modes have a correlation length of about $n_C \sim 100$ Gibbs iterations. The rest of the modes show increasingly longer correlation length, which increases towards the highest frequencies contained in the box. For this reason, the Nyquist frequency has the longest correlation length.

Especially for the highest frequencies towards the Nyquist frequencies, noise domination is only a partial explanation for the long correlation length. The far more important fact is, that the Nyquist frequency can in general not be resolved properly by the data, for example the Nyquist frequency is contained only once in the Fourier box.
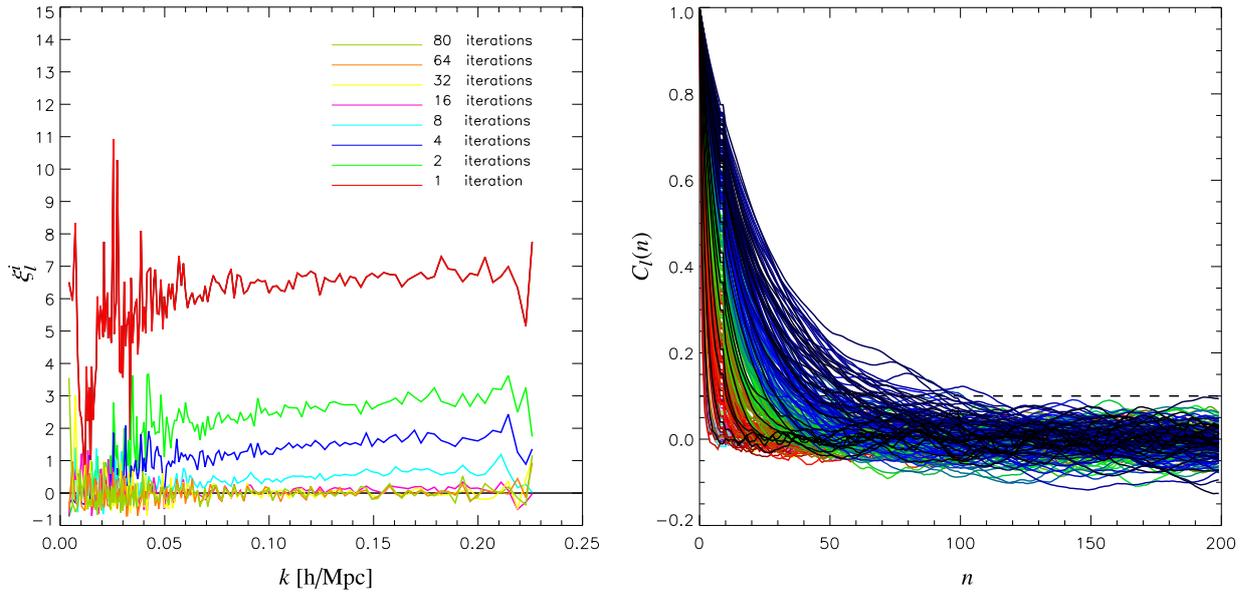
Figure 7.10.: Same as figure 7.5 but for low resolution runs with the inverse gamma prior. In comparison to figure 7.5 here one observes on average smaller convergence length.

For this reason, the variance increases towards the Nyquist frequency. Note, that this effect arises from the technical implementation of the fast Fourier transform, which operates on a finite grid, and that we are in principle able to account for these technical effects of the analysis scheme itself. However, the long correlation length for the Nyquist frequencies will in general only provide a rare amount of independent estimates at the Nyquist frequency and therefore this must be taken into account in further analysis.

It is clear, that this effect shifts to higher frequencies as soon as the resolution of our analysis scheme is increased. This can be observed in the high resolution analysis discussed in the next section 7.8.3. Also note, that this effect can be cured by introducing an informative prior which will greatly reduce the correlation length at these frequencies, as shown later in section 7.8.4.

To study the marginalized posterior P_l distributions in more detail we plot their histograms in Fig. 7.7. It is worth mentioning that none of them is even approximately Gaussian.

Another crucial point to address is the question how well we were able to account for effects of the survey geometry. This information is contained in the correlation structure of the samples. Therefore, we can examine this effect by calculating the correlation matrix of the P_l samples:

$$C_{ll'} = \left\langle \frac{P_l - \langle P_l \rangle}{\sqrt{Var P_l}} \frac{P_{l'} - \langle P_{l'} \rangle}{\sqrt{Var P_l}} \right\rangle, \tag{7.81}$$

where the ensemble averages are taken over $40,000$ Gibbs samples.

We present the result in Fig. 7.7. It can be clearly seen that this correlation matrix has a very well defined diagonal structure, as expected from theory. The highest off-diagonal correlations have been measured to be less than 20%, and are found at the highest frequencies close to the Nyquist frequency. Figure 7.7 also shows a blue band of anti-correlation around the diagonal. This anti-correlation indicates that the power-spectrum frequency resolution is higher than supported by the data. Since the data is fixed, and the mask couples neighbored Fourier modes, an increase in the power-spectrum amplitude P_l has to be compensated with a decrease in the neighboring power-spectrum amplitude P_{l+1} to have a good fit to the data. It is therefore possible, in a post-processing step, to reduce the frequency resolution of the estimated power-spectra until the anti-correlation vanishes. This is the idea behind the running median estimator, which will be presented in section 7.9.

However, since the posterior distributions for the P_l are non-Gaussian, the two point correlations do not contain
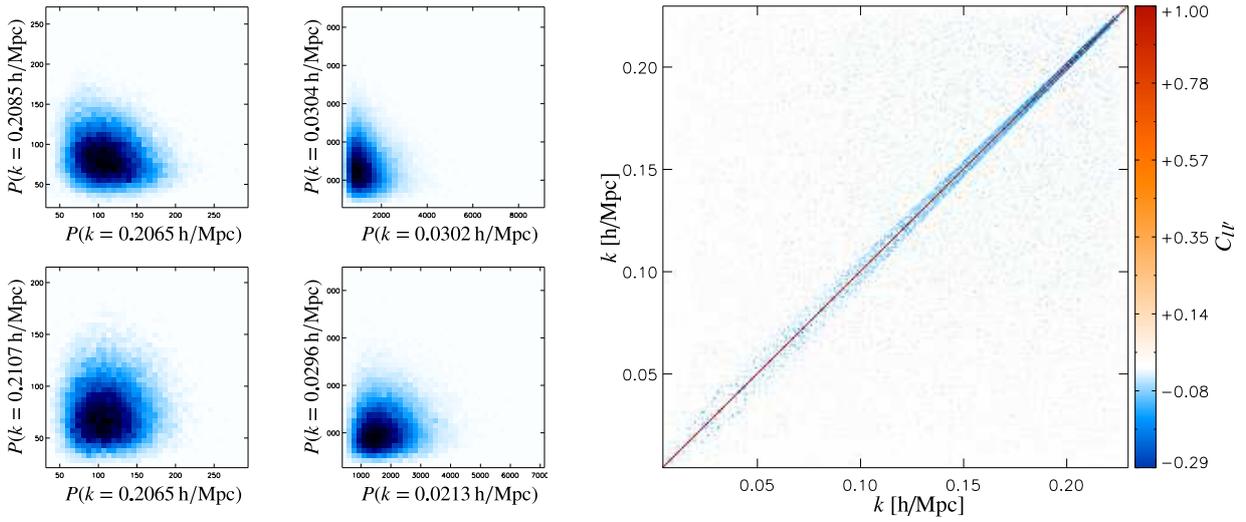
Figure 7.11.: The left panel shows four 2D marginalized posterior densities, which correspond to the maximally correlated and anti-correlated cases of the previous run. The right panel shows the power-spectrum covariance matrix estimated from the 40,000 Gibbs samples calculated with the informative prior.

all information. For this reason, we also demonstrate the marginalized posterior distribution for pairs of P_ls in Fig. 7.8, where we also show examples of maximally correlated and maximally anti-correlated modes.

Finally, we have plotted the full spectrum computed from our 40,000 sample run in Fig. 7.9. As can be seen, we chose a very high frequency resolution to reduce the correlation length of the Gibbs sampler in the low signal to noise regime. The estimated ensemble mean power-spectrum follows the true underlying power-spectrum, in particular the baryonic features. Towards the large scales, the uncertainty increases, as expected, since due to survey geometry and selection effects these scales are only poorly constrained by the data.

7.8.3. High resolution Simulation

In the previous section, we performed a low resolution analysis to compute a sufficiently large set of samples to estimate the correlation behavior of our algorithm. However, such a large amount of samples is not necessarily required and computational time may be better invested in performing higher resolution analysis. Therefore, in this section we describe the results obtained from such a high resolution analysis.

The setup for this test is the same as described in section 7.8.1, with the exception that here we use 256^3 voxels on an equidistant grid.

The main limitation for this test is CPU time, as a single sampling step takes about one hour. For this reason, extremely long chains are not feasible. Hence, we run many, rather short, parallel Gibbs chains, as described in section 8.6.3. In this fashion we obtain an optimal speed up for this parallelization scheme. The Gibbs sampler was run over 24 independently initialized chains and provided a total of 4230 samples.

We present the power-spectra obtained from the multiple-chain analysis in Fig. 7.9. As can be seen there is overall good agreement with the realization specific power-spectrum and the ensemble mean estimate found by the Gibbs sampler. We also do not observe a detectable bias in any parts of the spectrum.

Towards the Nyquist frequency the uncertainty increases. This is expected, as towards the edges of the box, the number of modes decreases, and variance increases. Note, that in this fashion the method takes into account the uncertainty introduced by the analysis scheme itself, for example the Fourier space discretization of the FFT.

7.8.4. Testing an informative Prior

When trying to analyze real galaxy surveys, one is faced with the situation that due to sky cuts and selection effects, usually less than 30% of the volume is available for analysis. This affects mainly the inference of the largest scales in the survey, as they are sparsely sampled, and therefore poorly constrained by the data.

In this situation the Jeffreys' prior, as a maximal ignorance prior, allows for large excursions from the expected true underlying power-spectrum. Since the Jeffreys' prior provides equal probability for all scales between zero and plus infinity, it also allows for power-spectrum values, which can be excluded on theoretical grounds, or by complementary experiments, which are more sensitive at the largest scales, such as CMB experiments.

From a Bayesian point of view, one could argue, that in the presence of a priori knowledge, a maximal ignorance prior is not the optimal choice.

Rather than sampling the entire space of possible power-spectrum coefficients P_m with equal probability, it would be beneficial to preferably sample the region in which we expect the true power-spectrum to exist, allowing for larger excursions with smaller probability.

This would have the effect, that the region of interest would be sampled more densely, thus allowing for better power-spectrum inferences for the same number of Gibbs samples. Also remember that, according to the discussion in section 7.5.2, the prior can be changed for any final post-processing analysis.

As an informative prior can lead to a more efficient sampling strategy in the presence of a priori knowledge, in the following we test ARES when employing the inverse gamma prior as described in section 7.6.2.

We base the inverse gamma prior on a flat power-spectrum guess, which was calculated according to Eisenstein & Hu (1998) and Eisenstein & Hu (1999), without the baryonic wiggles. We explicitly do not incorporate the information on the baryonic features, to demonstrate that solely the data drive their inference. Further, we choose $n_m^{Prior} = 5$, in order to make the prior sufficiently broad, while at the same time ensuring that it has finite variance.

With this prior choice we repeat the standard testing procedure as described in section 7.8.1.

At first we test the initial burn-in time, by starting with a power-spectrum which is a factor 10 higher in amplitude than the underlying true power-spectrum.

The results for the corresponding ξ_l^i, described in equation (8.27), are presented in Fig. 7.10. It can be seen that the burn-in time is much shorter for the large scale modes, which are poorly constrained by the data. Also note, that the overall burn-in times for the power-spectra are comparable to those of the maximal ignorance prior case (see Fig. 7.5), indicating that these modes are not influenced by the informative inverse gamma prior.

The real advantage of the informative prior becomes obvious when analyzing the correlation length for the individual power-spectrum coefficients P_m in the Gibbs chain. Again we used a low resolution 64^3 Voxel simulation in order to estimate the correlation coefficients, as given by equation (7.81). The results for this test are presented in Fig. 7.10.

In comparison to Fig. 7.5 it is clear, that the informative prior has a positive influence on the correlation length, which in this test are maximally on the order of a hundred Gibbs iterations.

As discussed previously the long correlation length at the highest frequencies are mainly of technical nature, as the Nyquist frequencies cannot be properly represented in the finite Fourier box required for the FFT. This fact introduced artificial variance, which, however, our method can take into account. The informative prior helps in this situation, as it stabilizes these artificial fluctuations by prior information.

In addition, we observed a much better numerical behavior of our method when employing the informative prior, as the code does not run as frequently into numerically extreme regimes as with the maximum ignorance prior. This leads to a faster convergence of the conjugate gradient algorithm towards the desired accuracy.

Further, we also observe a better correction of survey geometry effects. The correlation function for the P_l is plotted in Fig. 7.11. The maximal correlation between different P_l in this test was less than 10%, which is a clear improvement.

For comparison we also plot the 2D marginalized posterior densities, for the maximally correlated and anticorrelated modes in the maximum ignorance case.

Finally, we present the low and high resolution power-spectra for the informative prior in Fig. 7.12. Note, that our prior did not contain any information on the baryonic oscillations. As can be clearly seen, the baryonic features have nicely been recovered, demonstrating, that our informative prior provided much less information than contained in the data.
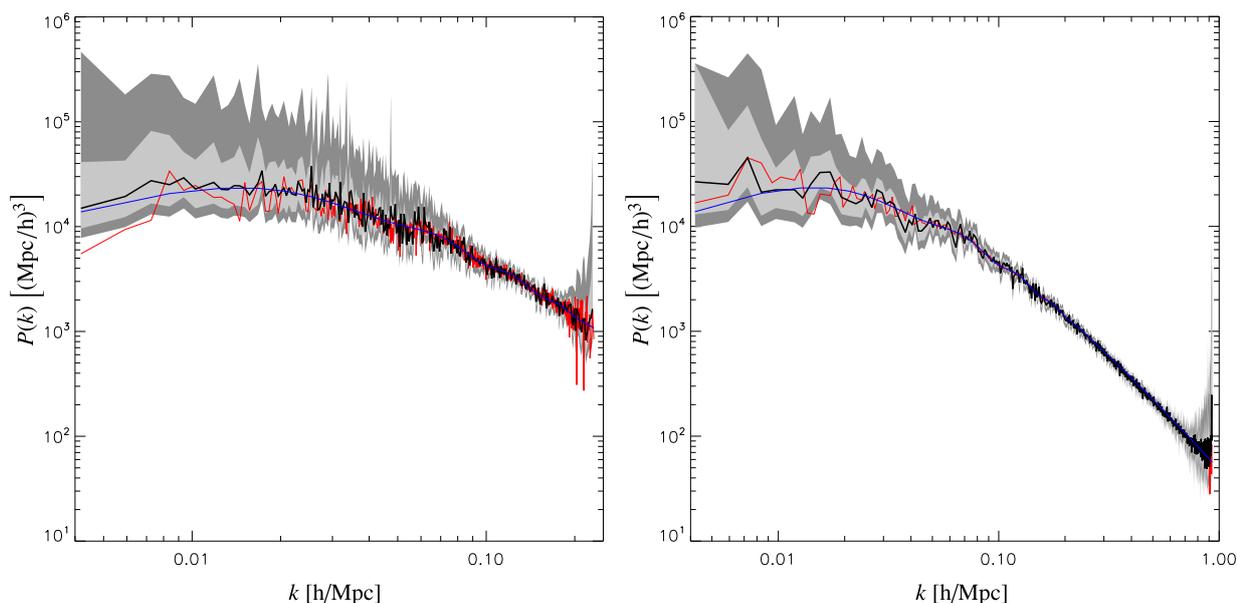
Figure 7.12.: Same as figure 7.9 but for the inverse gamma informative prior.

7.8.5. Testing with galaxy mock catalogs

In this section, we describe the application of ARES to a mock galaxy survey based on the Millennium run (De Lucia & Blaizot 2007).

The intention of this exercise is two-fold. First we want to test ARES in a more realistic setup, where the intrinsic shot noise of the galaxy distribution is correlated with the underlying signal, which could not be tested with the Gaussian tests before. And second, we want to demonstrate that ARES is able to reconstruct the fully evolved non-linear matter distribution of the N-body simulation.

This mock galaxy survey consists of a set of co-moving galaxy coordinates distributed in a 500 Mpc box. To obtain a realistic sky observation from this full sky galaxy sample, we virtually observe these galaxies through the sky mask and according to the selection function presented in Fig. 7.4. The discrete galaxy distribution resulting from this mock observation is then sampled to a 128^3 equidistant grid.

To reduce gridding artifacts, such as aliasing power, we employ the supersampling technique presented in chapter 5. This allows us to accurately treat the mode coupling, and will yield precision inference of the power-spectra up to the highest frequencies contained in the box.

Similarly to the method described in 7.8.3, here we will run 4 independently initialized chains. Further, we employ the maximum ignorance Jeffreys' prior. The galaxy distribution of this mock galaxy catalog follows the fully evolved non-linear matter distribution. Nevertheless, we initialize the Gibbs sampler with the linear power-spectrum. Then the initial burn-in period, of about 50 samples, is required to reach the non-linear power-spectrum. The systematic drift towards the correct power-spectrum is represented in Fig. 7.13. This experiment nicely demonstrates that the Gibbs sampling approach is able to recover the non-linearities of the fully evolved matter density field. At this point it is important to note that the Wiener filter is a linear operation on the data, and as such leaves the statistics of the data intact. This has been demonstrated by Kitaura et al. (2009), where they show, that the statistics of the reconstructed density field is consistent with a log-normal distribution, as expected for a non-linearly evolved matter distribution. This discussion clarifies, that the Wiener filter, or the Gibbs sampling approach as presented in this thesis, is very well able to capture the non-Gaussian characteristics of the density field.

However, in case one would like to perform a higher resolution analysis, it would be advantageous to initialize the Gibbs chain with a nonlinear power-spectrum guess, to yield even shorter burn-in times.

The ensemble averaged power-spectrum obtained from this run, together with the one and two sigma confidence regions, is also presented in Fig. 7.13. Here it can clearly be seen, that the recovered power-spectrum is consistent
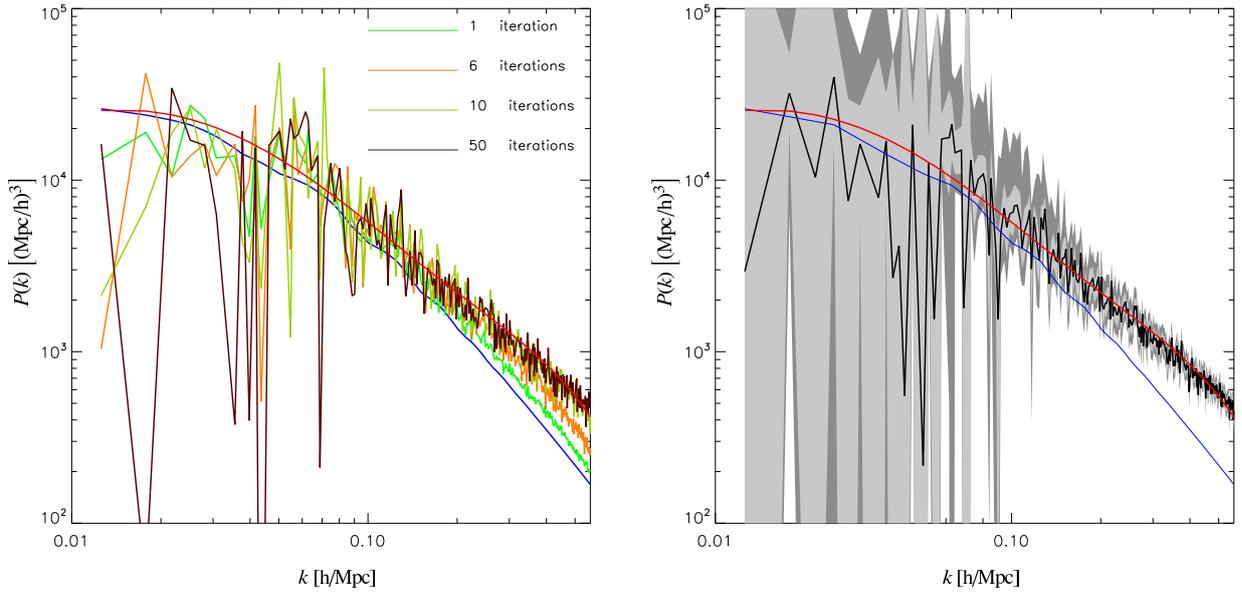
Figure 7.13.: The left panel shows successive power-spectrum samples during the burn-in period together with the initial guess (blue line).The plot clearly demonstrates the initial drift towards the non-linear power-spectrum. For comparison we also plotted the empirical fitting formula of Smith et al. (2003) for the non-linear power-spectrum (red line). The right panel shows the estimated power-spectrum of the Gibbs sampling analysis. Here, the initial power-spectrum and the non-linear fitting function are represented by the blue and red curve respectively. The black curve is the ensemble mean power-spectrum and the light gray and dark gray shaded regions represent the one sigma two sigma confidence regions respectively. It can be clearly seen that the Gibbs sampler is able to recover the non-linear power-spectrum.

with the fully non-linearly evolved matter field. Towards the larger scales the uncertainty increases, which is due to the imposed survey geometry.

So far, in all tests, we have focussed only on the recovery of the power-spectrum, and ignored the sample of reconstructed density fields. Since the Gibbs sampler also provides samples from the matter density field posterior, we are able to calculate any desired statistical summary for the matter field reconstructions. The ability to provide uncertainty information for the recovered density fields will in general be valuable for further science based on the matter field samples. For this reason, in Fig. 7.14, we present the estimated mean and variance maps obtained from the 4000 Gibbs samples. As can be seen, the variance map clearly captures the features of the survey geometry and selection effects.With the set of Gibbs samples being available, all joint uncertainties can easily be propagated to the finally inferred quantities, such as the gravitational potential or large scale cosmic flows, by applying the corresponding operation to the individual matter field samples. The result of such a procedure then again yields a probability distribution in the final quantity, enabling us to provide uncertainty information for these quantities.

7.9. Operations on the set of Gibbs samples

In this section, we present an example application operating on the set of Gibbs samples.

The outcome of our Bayesian method is not a single estimate but a Gibbs sample representation of the full posterior probability distribution for the power-spectrum coefficients. We are therefore able to propagate all uncertainties to any final result, simply by applying a post processing step to all Gibbs samples. The result of such an operation would again yield a probability distribution of the finally inferred quantity.

As a simple demonstration, we apply a running median filter to the set of power-spectra, which will reduce the spectral resolution.

It is known, that the median is a better estimator of the typical value of a sample than the mean when there are large extraneous outliers in the sample (Stuart & Ord 1994). For this reason, we choose the median to estimate the mode power in a given frequency bin Δk_m. Such a bin can be chosen to be large enough to smooth out any fluctuation below a certain scale. In our specific case we vary the bin width Δk_m with frequency, to allow for a logarithmic binning.

The median P_m^ν of a set of power-spectrum amplitudes $\{P_m\}$ contained within the frequency bin of width Δk_m around the mode k_m then satisfies the inequalities

$$\mathcal{P}\left(P_m \le P_m^\nu\right) \ge \frac{1}{2} \tag{7.82}$$

and

$$\mathcal{P}\left(P_m \ge P_m^\nu\right) \ge \frac{1}{2} . \tag{7.83}$$

The running median is then evaluated for every frequency in the power-spectrum sample.

We apply the running median to the set of power-spectrum samples obtained from the two Gaussian mock cases with the Jeffreys' and the inverse gamma prior. In doing so, we are able to calculate the mean and according uncertainty regions for the running median estimates. This effect has already been discussed in section 7.6.3. Since the reduction of frequency resolution also decreases the amount of free parameters, the total variance decreases as well.

The results of this experiment are demonstrated in Fig. 7.15. As one can easily see, the running median estimates are much smoother than the according Gibbs estimates. Also the reduction of frequency resolution by the running median estimator yields smaller confidence regions.

Finally we are interested in the recovery of the baryonic features in the power-spectrum. We therefore employ the common procedure of dividing the measured power-spectrum by one without baryonic wiggles $P_m^{nowiggles}$. We then obtain the wiggle function as:

$$f_m^{wiggle} = \frac{P_m}{P_m^{nowiggles}} . \tag{7.84}$$

Calculating the wiggle function for all Gibbs samples and applying the running median estimator to the set of wiggle functions will yield the distribution of wiggle functions. We then calculate the mean and the according one and two sigma confidence regions. The result of this calculation is presented in Fig. 7.15 for the two Gaussian test cases. As expected, the variance towards the largest scales increases. Nevertheless, Fig. 7.15 clearly demonstrates that the baryonic features have been recovered precisely by the Gibbs sampling approach.

This example nicely demonstrates that the uncertainty estimation can easily be transported to any final quantity estimated from the set of Gibbs samples.

7.10. Conclusion

This chapter provided a description of ARES, a new Bayesian computer algorithm, designed to extract the full information on the two point statistics from any given probe of the three dimensional large scale structure.

The scientific aim of this algorithm is to provide an estimate of the power-spectrum posterior $\mathcal{P}\left(\{P(k_i)\}|\{d_i\}\right)$, conditional on a data set, while accounting and correcting for all possible sources of uncertainties, such as survey geometry, selection effects and biases. This is achieved by exploring the power-spectrum posterior $\mathcal{P}\left(\{P(k_i)\}|\{d_i\}\right)$ via a Gibbs sampling approach.

While direct sampling from the power-spectrum posterior is not possible, it is possible to draw samples from the full joint posterior distribution $\mathcal{P}\left(\{P(k_i)\}, \{s_i\}|\{d_i\}\right)$ of the power-spectrum coefficients $P(k_i)$ and the three dimensional matter density contrast amplitudes s_i conditional on a given set of data points $\{d_i\}$.

The entire Gibbs sampling algorithm therefore consists of two basic sampling steps, in which first a full three dimensional Wiener reconstruction algorithm is applied to the data and then a power-spectrum is drawn from the inverse gamma distribution. In this fashion we obtain a set of power-spectrum and signal amplitude samples, which provide a full representation of the full joint posterior distribution $\mathcal{P}\left(\{P(k_i)\}, \{s_i\}|\{d_i\}\right)$. The scientific output of this Bayesian method therefore is not a single estimate but a complete probability distribution, enabling us to calculate any desired statistical summary such as the mean, mode or variance.

We also demonstrated, that given a set of Gibbs samples, it is possible to provide an analytic approximation
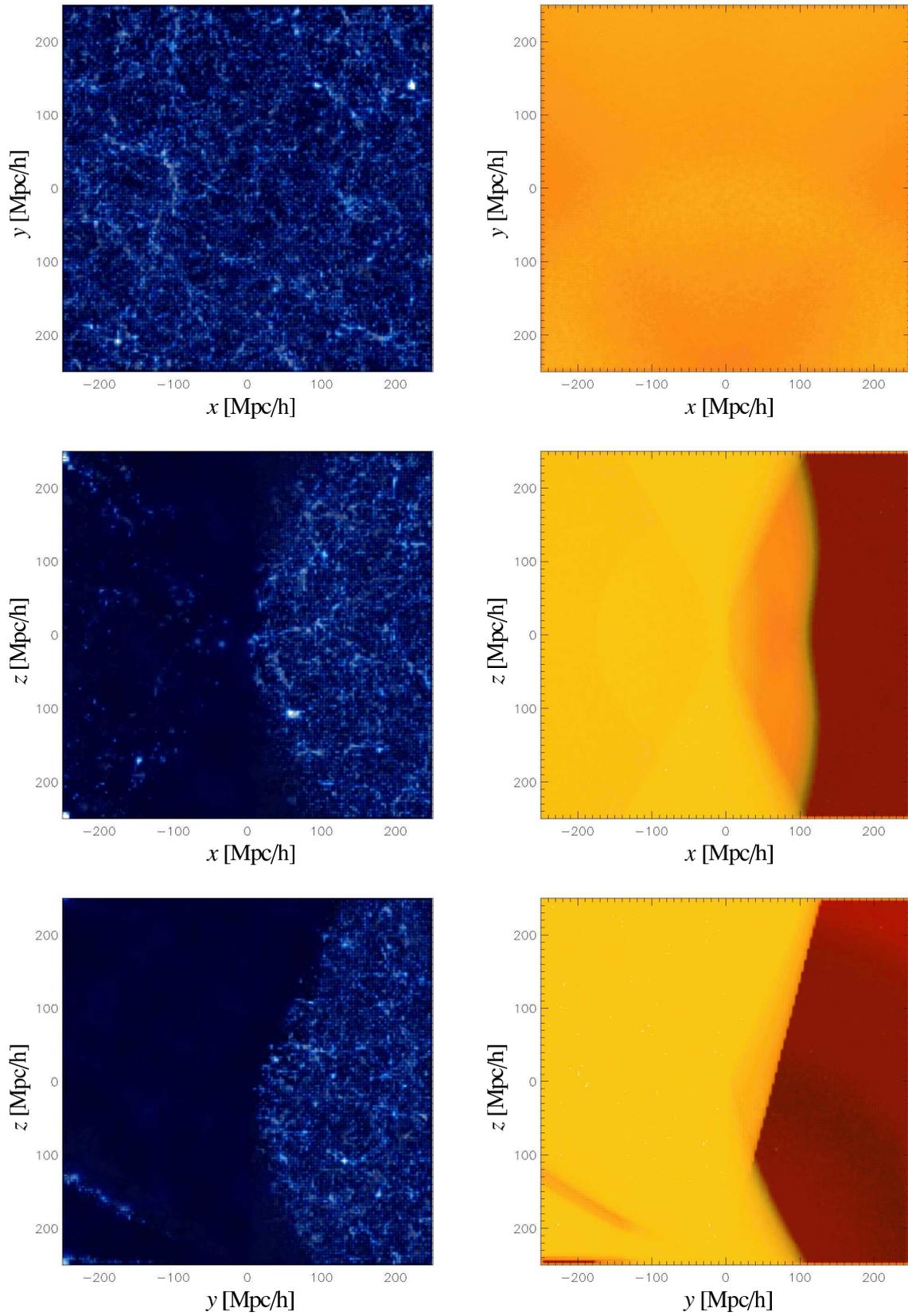
Figure 7.14.: Volume rendering from three different perspectives of the ensemble variance (right panels) and the ensemble mean (left panels) obtained from the mock galaxy catalog analysis.
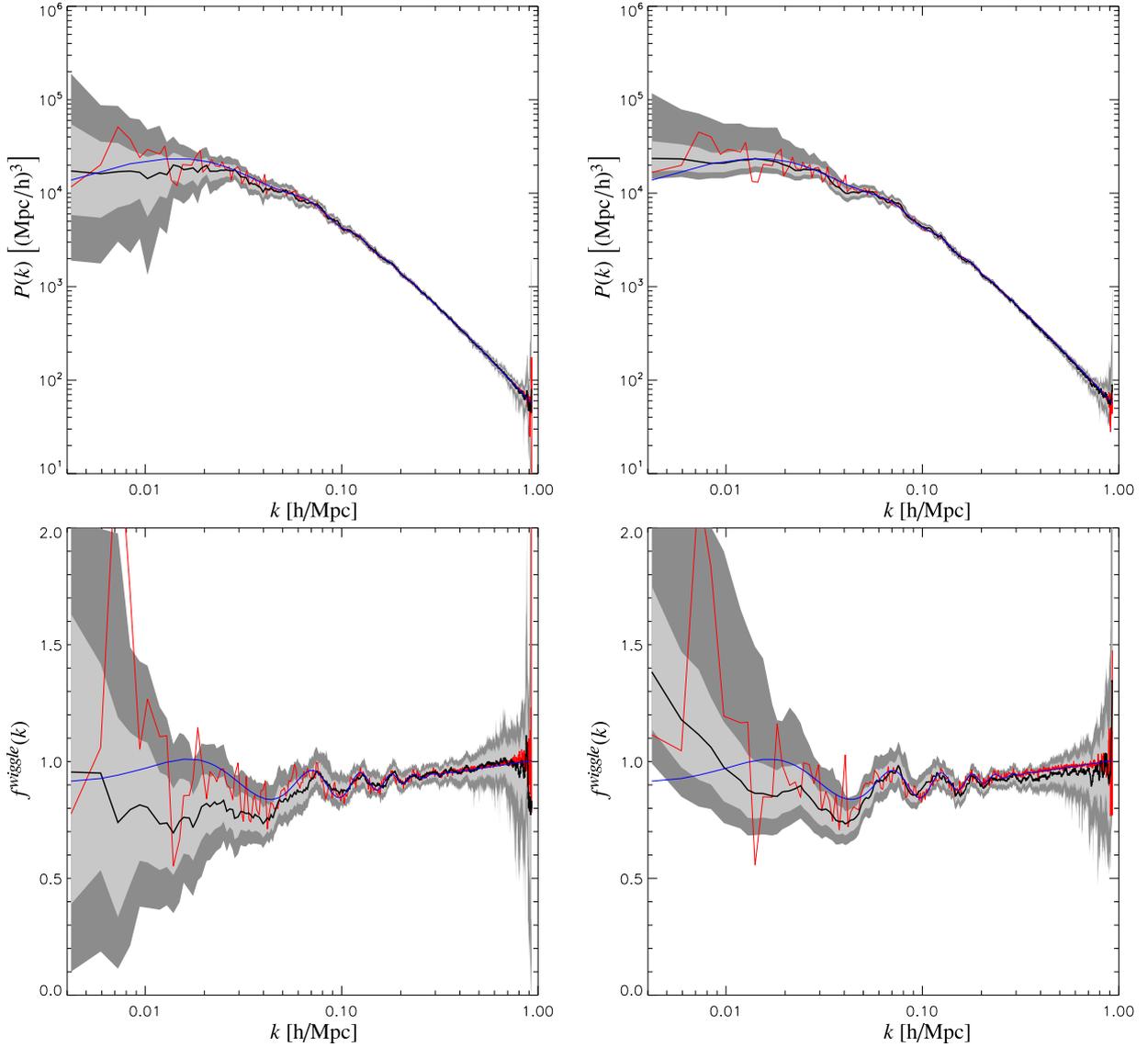
Figure 7.15.: Running median estimates of the power-spectra (upper panels) and the according wiggle functions (lower panels) for the set of Gibbs samples with the Jeffreys' prior (left panels), and the inverse gamma prior (right panels). The black lines represent the ensemble mean of the sample set and the light gray and dark gray shaded regions denote the one and two sigma confidence regions respectively. Additionally we show the according input power-spectra. The blue line shows the cosmological power-spectrum from which the matter field realization was drawn, and the red line is the power-spectrum of this specific matter field realization.

to the power-spectrum posterior $\mathcal{P}(\{P(k_i)\}|\{d_i\})$. This Blackwell-Rao estimator has an analytically appealing form enabling us to calculate any desired moment of the probability distribution in a simple analytic way.

In addition, since the full joint probability distribution is available, it is easy to carefully propagate all uncertainties through to the result of further post-processing analysis steps, such as parameter estimation.

In this chapter, we focused on thoroughly testing the performance and behavior of ARES by applying it to simulated data with controlled properties. These tests were designed to highlight the problematic of survey geometry and selection effects, for the two cases of Gaussian random fields and a mock galaxy catalog based on the Millennium run.

One of the main goals of these tests was to build up intuition on the phenomenological behavior of the Gibbs sampling algorithm, estimating particularly issues, such as the correlation length of the Gibbs chain, burn-in and convergence times. The result of these tests is of special relevance, as it shows how long the Gibbs sampling chain has to run in order to produce a sufficient amount of independent samples.

Through these experiments we found that the longest correlation lengths are dominated by the poorly constrained Nyquist modes of the box, which can be easily alleviated by imposing some prior knowledge on these modes. In doing so we found that the maximal correlation length for the Gibbs chain was on the order of hundred Gibbs samples. Thus, creating a large number of independent samples in a full scale data analysis is numerically very well feasible.

However, the most important result of these tests is, that our method is able to correct for artificial mode coupling due to the survey geometry and selection effects. This was tested by examining the correlation structure of the Gibbs samples, which showed that the maximal residual correlation can be reduced to less than 10%, demonstrating that this method correctly accounts for geometry effects.

The application of ARES to a galaxy mock catalog, based on the Millennium run, demonstrated the ability of our method to capture the characteristics of the fully nonlinearly evolved matter field. This is owed to the fact, that the Wiener filter is a linear operation on the data, and as such does not destroy the intrinsic statistical characteristics of the data set.

Nevertheless, a full real data analysis of existing redshift surveys requires the treatment of additional systematic effects such as scale or luminosity dependent biases or redshift space distortion corrections, which we defer to future works. However, the Bayesian framework, as presented here, can take all these effects naturally into account and treats them statistically fully consistent. Beside the possibility to include various kinds of uncertainties, the Gibbs sampling approach also allows for a natural joint analysis of different data sets, taking into account the systematics of each individual data set.

In summary, we showed that ARES is a highly flexible and adaptive machinery for large scale structure analysis, which is able to account for a large variety of systematic effects and uncertainties. For this reason, ARES has the potential to contribute to the era of precision cosmology.

8. HADES - A fast Hamiltonian sampler for large scale structure inference

Astronomy compels the soul to look upwards and leads us from this world to another.

Plato (The Republic)

Abstract

In this chapter, a new and efficient Bayesian method for nonlinear three dimensional large scale structure inference is presented. I employ a Hamiltonian Monte Carlo (HMC) sampler to obtain samples from a multivariate highly non-Gaussian lognormal Poissonian density posterior given a set of observations. The HMC allows to take into account the nonlinear relations between the observations and the underlying density field which we seek to recover. As already described in the previous chapter, also the HMC provides a sampled representation of the density posterior, and thus any desired statistical summary, such as the mean, mode or variance, can be calculated from the set of samples. Further, it allows us to seamlessly propagate non-Gaussian uncertainty information to any final quantity inferred from the set of samples. The developed method is extensively tested in a variety of test scenarios, taking into account a highly structured survey geometry and selection effects. Tests with a mock galaxy catalog based on the Millennium Run show that the method is able to recover the filamentary structure of the nonlinear density field. The results further demonstrate the feasibility of non-Gaussian sampling in high dimensional spaces, as required for precision nonlinear large scale structure inference. The HMC is a flexible and efficient method, which allows for simple extension and incorporation of additional observational constraints.

8.1. Introduction: non-linear sampling

Modern large galaxy surveys allow us to probe cosmic large scale structure to very high accuracy if the enormous amount of data can be processed and analyzed efficiently. Especially, precision reconstruction of the three dimensional density field from observations poses complex numerical challenges. For this reason, several reconstruction methods and attempts to recover the underlying density field from galaxy observations have been presented in the literature (see e.g. Bertschinger & Dekel 1989, 1991, Lahav et al. 1994, Hoffman 1994, Fisher et al. 1995, Bistolas & Hoffman 1998, Webster et al. 1997, Schmoldt et al. 1999, Zaroubi 2002, Erdoğdu et al. 2004, Kitaura et al. 2009, Jasche et al. 2009). In particular, Kitaura et al. (2009) presented a high resolution Wiener reconstruction of the Sloan Digital Sky Survey (SDSS) matter density field, and demonstrated the feasibility of precision large scale structure analysis. The Wiener filtering approach is based on a linear data model, which takes into account several observational effects, such as survey geometry, selection effects and noise (Kitaura & Enßlin 2008, Kitaura et al. 2009, Jasche et al. 2009). Although, the Wiener filter, as described in chapter 7, has proven to be extremely efficient for three dimensional matter field reconstruction, it still relies on a Gaussian approximation of the density posterior. While this is an adequate approximation for the largest scales, precision recovery of nonlinear density structures may require non-Gaussian posteriors. Especially, the detailed treatment of the non-Gaussian behavior and the structure of the Poissonian shot noise contribution may allow for more precise recovery of poorly sampled objects. In addition, for a long time it has been suggested that the fully evolved nonlinear matter field can be well described by lognormal statistics (see e.g. Hubble 1934, Peebles 1980, Coles & Jones 1991, Gaztanaga & Yokoyama 1993, Kayo et al. 2001). However, it should be noted that a multivariate lognormal distribution can only represent one and two point statistics of the density field correctly. In the absence of any better prior model, these discussions suggest the use of a lognormal Poissonian posterior for large scale structure inference. Several methods have been proposed to take into account non-Gaussian density posteriors (see e.g. Saunders & Ballinger 2000, Kitaura & Enßlin 2008, Enßlin et al. 2009).

However, if the recovered nonlinear density field is to be used for scientific purposes, the method not only has to provide a single estimate, such as a mean or maximum a postiori reconstruction, but it should also provide uncertainty information, and the means to nonlinearly propagate this uncertainty to any final quantity inferred from the recovered density field.

For this reason, in this chapter a new Bayesian method for nonlinear large scale structure inference is proposed. The developed computer algorithm HADES (HAmiltonian Density Estimation and Sampling) explores the posterior distribution via an Hamiltonian Monte Carlo (HMC) sampling scheme. Unlike conventional Metropolis Hastings algorithms, which move through the parameter space by a random walk, and therefore require prohibitive numbers of steps to explore high dimensional spaces, the HMC sampler suppresses random walk behavior by introducing a persistent motion of the Markov chain through the parameter space (Duane et al. 1987, Neal 1993, 1996). In this fashion, the HMC sampler maintains reasonable efficiency even for high dimensional problems (Hanson 2001). The HMC sampler has been widely used in Bayesian computation (see e.g. Neal 1993). In cosmology it has been employed for cosmological parameter estimation and CMB data analysis (Hajian 2007, Taylor et al. 2008).

In this chapter we demonstrate that the HMC can efficiently be used to sample the lognormal Poissonian posterior even in high dimensional spaces. In this fashion, the method is able to take into account the nonlinear relationship between the observation and the underlying density which we seek to recover. The scientific output of the HMC is a sampled representation of the density posterior. For this reason, any desired statistical summary such as the mean, mode or variance can easily be calculated from the HMC samples. Further, the full non-Gaussian uncertainty information can seamlessly be propagated to any finally estimated quantity by simply applying the according estimation procedure to all samples. This allows us to estimate the robustness of conclusions drawn from the analyzed data.

This chapter begins by presenting a short justification for the use of the lognormal distribution as a prior for nonlinear density inference in section 8.2, followed by a discussion of the lognormal Poissonian posterior in section 8.3. Section 8.4 outlines the HMC method. In section 8.5 the Hamiltonian equations of motion for the lognormal Poissonian posterior are presented. Details of the numerical implementation are described in section 8.6. The method is extensively tested in section 8.7 by applying HADES to generated mock observations, taking into account a highly structured survey geometry and selection effects. In section 8.8 we summarize and conclude.

8.2. The lognormal distribution of density

In standard cosmological pictures, it is assumed that the initial seed perturbations in the primordial density field originated from an inflationary phase in the early stages of the Big Bang. This inflationary phase enhances microscopic quantum fluctuations to macroscopic scales yielding the initial density fluctuations required for gravitational collapse. These theories predict the initial density field amplitudes to be Gaussian distributed. However, it is obvious that Gaussianity of the density field can only be true in the limit $|\delta| \ll 1$, where δ is the density contrast. In fully evolved density fields with amplitudes of $\sigma_8 > 1$, as observed in the sky on the scale of galaxies, clusters and super clusters, a Gaussian density distribution would allow for negative densities, and therefore would violate weak and strong energy conditions. In particular, it would give rise to negative mass ($\delta < -1$). Therefore, in the course of gravitational structure formation the density field must have changed its statistical properties. Coles & Jones (1991) argue that assuming Gaussian initial conditions in the density and velocity distributions will lead to a log-normally distributed density field.

Although, the exact probability distribution for the density field in nonlinear regimes is not known, the lognormal distribution seems to be a good phenomenological guess with a long history. Already Hubble noticed that galaxy counts in two dimensional cells on the sky can be well approximated by a lognormal distribution (Hubble 1934). Subsequently, the lognormal distribution has been extensively discussed and agreements with galaxy observations have been found (e.g. Hubble 1934, Peebles 1980, Coles & Jones 1991, Gaztanaga & Yokoyama 1993, Kayo et al. 2001). Kayo et al. (2001) studied the probability distribution of smoothed cosmological nonlinear density fluctuations from N-body simulations with Gaussian initial conditions. They found that the lognormal distribution accurately describes the one point statistics of the nonlinear density field even up to values of the density contrast of $\delta \sim 100$. In addition, Kitaura et al. (2009) analyzed the statistical properties of the SDSS DR6 Wiener reconstructed density field, and confirmed a lognormal behavior.

Therefore, according to observations and theoretical considerations, we will model the statistical behavior of the

nonlinear density field by a multivariate lognormal distribution, as given by:

$$\mathcal{P}(\{s_k\}|Q) = \frac{1}{\sqrt{2\pi det(Q)}} e^{-\frac{1}{2}\sum_{ij}\left(ln(1+s_i)+\mu_i\right)Q_{ij}^{-1}\left(ln(1+s_j)+\mu_j\right)} \prod_k \frac{1}{1+s_k}, \tag{8.1}$$

where s_i is the density signal at the three dimensional cartesian position $\vec{x}_i$, Q is the covariance matrix of the lognormal distribution and μ_i describes a constant mean field given by:

$$\mu_i = \frac{1}{2}\sum_{l,m} Q_{lm}. \tag{8.2}$$

This probability distribution, seems to be an adequate prior choice for reconstructing the present density field, at least up to two point statistics. However, using such a prior requires highly nonlinear reconstruction methods, as will be presented in the following.

8.3. Lognormal Poissonian posterior

Studying the actual matter distribution of the Universe requires drawing inferences from some observable tracer particle. The most obvious tracer particles for the cosmic density field are galaxies, which tend to follow the gravitational potential of matter. Assuming galaxies to be discrete particles, their galaxy distribution can be described as a specific realization drawn from an inhomogeneous Poisson process (see e.g. Layzer 1956, Peebles 1980, Martínez & Saar 2002). The corresponding probability distribution is:

$$\mathcal{P}(\{N_k^g\}|\{\lambda_k\}) = \prod_k \frac{(\lambda_k)^{N_k^g} e^{-\lambda_k}}{N_k^g!}, \tag{8.3}$$

where N_k^g is the observed galaxy number at position $\vec{x}_k$ in the sky and λ_k is the expected number of galaxies at this position. The mean galaxy number is related to the signal s_k via:

$$\lambda_k = R_k \bar{N}(1 + B(s)_k), \tag{8.4}$$

where R_k is a linear response operator, incorporating survey geometries and selection effects, $\bar{N}$ is the mean number of galaxies in the volume and $B(x)_k$ is a nonlinear, non local, bias operator at position $\vec{x}_k$. The lognormal prior given in equation (8.1) together with the Poissonian likelihood given in equation (8.3) yields the lognormal Poissonian posterior, for the density contrast s_k given some galaxy observations N_k^g:

$$\mathcal{P}(\{s_k\}|\{N_k^g\}) \quad = \quad \frac{e^{-\frac{1}{2}\sum_{ij}\left(ln(1+s_i)+\mu_i\right)Q_{ij}^{-1}\left(ln(1+s_j)+\mu_j\right)}}{\sqrt{2\pi det(Q)}} \prod_l \frac{1}{1+s_l} \prod_k \frac{\left(R_k\bar{N}(1+B(s)_k)\right)^{N_k^g} e^{-R_k\bar{N}(1+B(s)_k)}}{N_k^g!} \tag{8.5}$$

However, this posterior greatly simplifies if we perform the change of variables by introducing $r_k = ln(1 + s_k)$. Note, that this change of variables is also numerically advantageous, as it prevents numerical instabilities at values $\delta \sim -1$. Hence, we yield the Posterior

$$\mathcal{P}(\{r_k\}|\{N_k^g\}) \quad = \quad \frac{e^{-\frac{1}{2}\sum_{ij}(r_i+\mu_i)Q_{ij}^{-1}(r_j+\mu_j)}}{\sqrt{2\pi det(Q)}} \prod_k \frac{\left(R_k\bar{N}(1+B(e^r-1)_k)\right)^{N_k^g} e^{-R_k\bar{N}(1+B(e^r-1)_k)}}{N_k^g!}. \tag{8.6}$$

It is important to note, that this is a highly non-Gaussian distribution, and non-linear reconstruction methods are required in order to perform accurate matter field reconstructions in the non-linear regime. In example, estimating the maximum a posteriori values from the lognormal Poissonian distribution involves the solution of implicit equations. Several attempts to use a lognormal Poissonian posterior for density inference have been presented in literature. These attempts date back to Sheth (1995) who proposed to use a variable transformation in order to derive

a generalized Wiener filter for the lognormal distribution. This approach, however, yielded a very complex form for the noise covariance matrix making applications to real data sets impractical. The first successful application of the lognormal Poissonian distribution for density inference was presented by Saunders et al. (2000). Their method is based on the expansion of the density logarithm into spherical harmonics (Saunders & Ballinger 2000). More accurate schemes based on maximum and mean posteriori principles were derived by (Enßlin et al. 2009). Recently, an implementation of the maximum a posteriori scheme was presented and thoroughly tested by (Kitaura et al. 2009). They found that, assuming a linear bias, the lognormal Poissonian posterior permits estimation of the density field deep in the nonlinear regime up to values $\delta \geq 1000$ of the density contrast. However, I am not solely interested in a single estimate of the density distribution, I rather prefer to draw samples from the lognormal Poissonian posterior. In the following, a numerically efficient method to sample from this highly non-Gaussian distribution is described.

8.4. Hamiltonian sampling

As already described in the previous section the lognormal Poissonian posterior will involve highly nonlinear reconstruction methods and will therefore be numerically demanding. Nevertheless, since we propose a Bayesian method, we are not interested in only providing a single estimate of the density field, but would rather be able to sample from the full non-Gaussian posterior. Unlike, in the Gibbs sampling approach to density field sampling, as proposed in chapter 7, there unfortunately exists no known way to directly draw samples from the lognormal Poissonian distribution. For this reason, a Metropolis-Hastings sampling mechanism has to be employed.

However, the Metropolis-Hastings has the numerical disadvantage that not every sample will be accepted. A low acceptance rate can therefore result in a prohibitive numerical scaling for the method, especially since we are interested in estimating full three dimensional matter fields which usually have about 10^6 or more free parameters s_k. This high rejection rate is due to the fact, that conventional Markov Chain Monte Carlo (MCMC) methods move through the parameter space by a random walk and therefore require a prohibitive amount of samples to explore high-dimensional spaces. Given this situation, we propose to use a Hybrid Monte Carlo method, which in the absence of numerical errors, would yield an acceptance rate of unity.

The so called Hamiltonian Monte Carlo (HMC) method exploits techniques developed to follow classical dynamical particle motion in potentials (Duane et al. 1987, Neal 1993, 1996). In this fashion the Markov sampler follows a persistent motion through the parameter space, supressing the random walk behavior. This enables us to sample with reasonable efficiency in high dimensional spaces (Hanson 2001).

The idea of the Hamiltonian sampling can be easily explained. Suppose, that we wish to draw samples from the probability distribution $\mathcal{P}(\{x_i\})$, where $\{x_i\}$ is a set consisting of the N elements x_i. If we interpret the negative logarithm of this posterior distribution as a potential:

$$\psi(x) = -ln(\mathcal{P}(x)), \tag{8.7}$$

and by introducing a 'momentum' variable p_i and a 'mass matrix' M, as nuisance parameters, we can formulate a Hamiltonian describing the dynamics in the multi dimensional phase space. Such a Hamiltonian is then given as:

$$H = \sum_i \sum_j \frac{1}{2} p_i M_{ij}^{-1} p_j + \psi(x), \tag{8.8}$$

As can be seen in equation (8.8), the form of the Hamiltonian is such, that this distribution is separable into a Gaussian distribution in the momenta $\{p_i\}$ and the target distribution $\mathcal{P}(\{x_i\})$ as:

$$e^{-H} = \mathcal{P}(\{x_i\}) \, e^{-\frac{1}{2} \sum_i \sum_j p_i M_{ij}^{-1} p_j} \,. \tag{8.9}$$

It is therefore obvious that, marginalizing over all momenta will yield again our original target distribution $\mathcal{P}(\{x_i\})$.

Our task now is to draw samples from the joint distribution, which is proportional to $exp(-H)$. To find a new sample of the joint distribution we first draw a set of momenta from the distribution defined by the kinetic energy term, that is an N dimensional Gaussian with a covariance matrix M. We then allow our system to evolve deterministically, from our starting point $(\{x_i\}, \{p_i\})$ in the phase space for some fixed pseudo time τ according to Hamilton's

equations:

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i} \, . \tag{8.10}$$

$$\frac{dp_i}{dt} = \frac{\partial H}{\partial x_i} = -\frac{\partial \psi(x)}{\partial x_i} \, . \tag{8.11}$$

The integration of this equations of motion yields the new position $(\{x'_i\}, \{p'_i\})$ in phase space. This new point is accepted according to the usual acceptance rule:

$$\mathcal{P}_A = min\left[1, exp(-\left(H(\{x'_i\}, \{p'_i\}) - H(\{x_i\}, \{p_i\})\right))\right] \, . \tag{8.12}$$

Since the equations of motion provide a solution to a Hamiltonian system, energy or the Hamiltonian given in equation (8.8) is conserved, and therefore the solution to this system provides an acceptance rate of unity. In practice however, numerical errors can lead to a somewhat lower acceptance rate. Once a new sample has been accepted the momentum variable is discarded and the process restarts by randomly drawing a new set of momenta. The individual momenta $\{p_i\}$ will not be stored, and therefore discarding them amounts to marginalizing over this auxiliary quantity. Hence, the Hamiltonian sampling procedure basically consists of two steps. The first step is a Gibbs sampling step, which yields a new set of Gaussian distributed momenta. The second step, on the other hand amounts to solving a dynamical trajectory on the posterior surface.

8.5. Equations of motion for a log-normal Poissonian system

In the framework of Hamiltonian sampling the task of sampling from the lognormal Poissonian posterior reduces to solving the corresponding Hamiltonian system. Given the posterior distribution defined in equation (8.6) we can write the potential $\psi(\{r_k\})$ as:

$$
\begin{aligned}
\psi(\{r_k\}) \ = \ & \frac{1}{2}ln(2\pi det(Q)) + \frac{1}{2}\sum_{ij}(r_i + \mu_i)\,Q_{ij}^{-1}\left(r_j + \mu_j\right) \\
& - \sum_k\left[ln\left(\frac{(R_k\bar{N})^{N_k^g}}{N_k^g!}\right) + N_k^g\,ln(1 + B(e^r - 1)_k) - R_k\bar{N}(1 + B(e^r - 1)_k)\right] \, .
\end{aligned}
\tag{8.13}
$$

The gradient of this potential with respect to r_l then yields the forces, given as:

$$
\frac{\partial\psi(\{r_k\})}{\partial r_l} \ = \ \sum_j Q_{lj}^{-1}\left(r_j + \mu_j\right) - \left(\frac{N_l^g}{(1 + B(e^r - 1)_l)} - R_l\bar{N}\right)\frac{\partial B(e^r - 1)}{\partial(e^r - 1)}\bigg|_l\, e^{r_l} \, .
\tag{8.14}
$$

Equation (8.14) obviously is a very general formulation of the reconstruction problem, and it demonstrates that the Hamiltonian sampler can in principle deal with all kinds of nonlinearities, especially in the case of the bias operator $B(x)$. However, for the sake of this theis, but without loss of generality, in the following we will assume a linear bias model:

$$B(x)_k = b\, x_k \, , \tag{8.15}$$

where b is a constant linear bias factor. We then obtain the potential:

$$
\begin{aligned}
\psi(\{r_k\}) \ = \ & \frac{1}{2}ln(2\pi det(Q)) + \frac{1}{2}\sum_{ij}(r_i + \mu_i)\,Q_{ij}^{-1}\left(r_j + \mu_j\right) \\
& - \sum_k\left[ln\left(\frac{(R_k\bar{N})^{N_k^g}}{N_k^g!}\right) + N_k^g\,ln(1 + b\,(e^r - 1)) - R_k\bar{N}(1 + b\,(e^r - 1))\right] \, ,
\end{aligned}
\tag{8.16}
$$
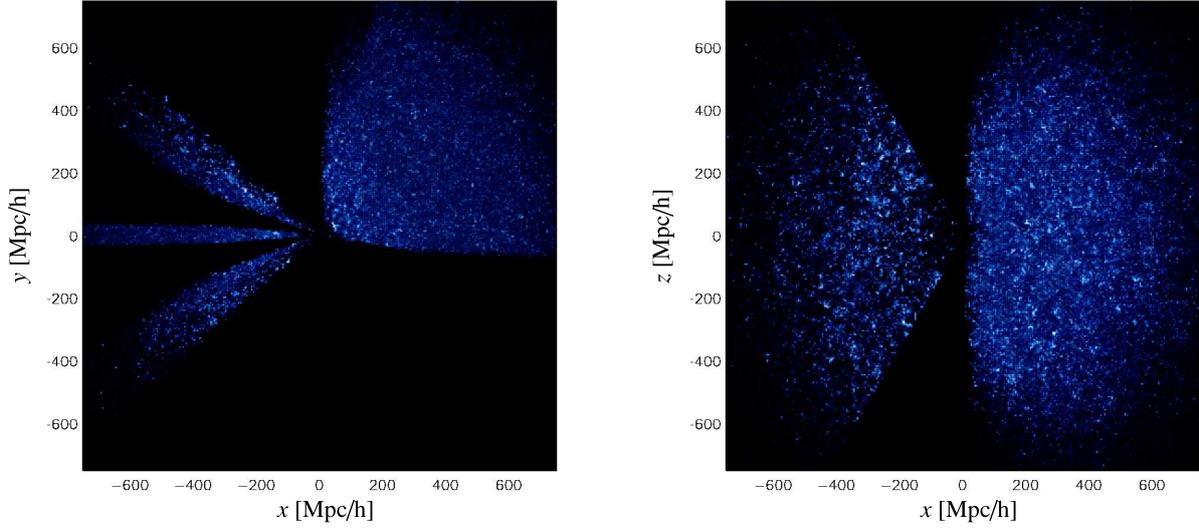
Figure 8.1.: Volume rendering of artificial galaxy counts, generated as described in section 8.7.1. The two pannels show different projections. Effects of survey geometry and selection function are clearly visible. The observer is centered at $(0, 0, 0)$.

and the corresponding gradient reads:

$$\frac{\partial \psi(\{r_k\})}{\partial r_l} = \sum_j Q_{lj}^{-1} \left(r_j + \mu_j \right) - \left(\frac{N_l^g}{(1 + b\,(e^r - 1))} - R_l \bar{N} \right) b\, e^{r_l}.$$

(8.17)

Inserting these results in equations (8.10) and (8.11) then yields the equations of motion:

$$\frac{dr_i}{dt} = \sum_j M_{ij}^{-1}\, p_j,$$

(8.18)

and

$$\frac{dp_i}{dt} = -\sum_j Q_{ij}^{-1} \left(r_j + \mu_j \right) - \left(\frac{N_i^g}{(1 + b\,(e^{r_i} - 1))} - R_i \bar{N} \right) b\, e^{r_i}.$$

(8.19)

New points on the lognormal Poissonian posterior surface can then easily be obtained by solving for the trajectory governed by the dynamical equations (8.18) and (8.19).

8.6. Numerical Implementation

Our numerical implementation of the lognormal Poissonian Sampler is named HADES (Hamiltonian Density Estimation and Sampling). It utilizes the FFTW3 library for Fast Fourier Transforms and the GNU scientific library (gsl) for random number generation (Frigo & Johnson 2005, Galassi et al. 2003). In particular, we use the Mersenne Twister MT19937, with 32-bit word length, as provided by the gsl_rng_mt19937 routine, which was designed for Monte Carlo simulations (Matsumoto & Nishimura 1998).

8.6.1. The leapfrog scheme

As described above, a new sample can be obtained by calculating a point on the trajectory governed by equations (8.18) and (8.19). This means that if we are able to integrate the Hamiltonian system exactly energy will be conserved along such a trajectory, yielding a high probability of acceptance. However, the method is more general due to the Metropolis acceptance criterion given in equation (8.12). In fact, it is allowed to follow any trajectory to generate a new sample. This would enable us to use approximate Hamiltonians, which may be evaluated computationally more efficiently. Note, however, that only trajectories that approximately conserve the Hamiltonian given in equation (8.8) will result in high acceptance rates.

In order to achieve an optimal acceptance rate, we seek to solve the equations of motion exactly. For this reason, we employ a leapfrog scheme for the numerical integration. Since the leapfrog is a symplectic integrator, it is exactly reversible, a property required to ensure the chain satisfies detailed balance (Duane et al. 1987). It is also numerically robust, and allows for simple propagation of errors. Here, we will implement the Kick-Drift-Kick scheme. The equations of motions are integrated by making n steps with a finite stepsize ϵ, such that $\tau = n\epsilon$:

$$p_i\left(t + \frac{\epsilon}{2}\right) = p_i(t) - \frac{\epsilon}{2} \left.\frac{\partial \psi(\{r_k\})}{\partial r_l}\right|_{r_i(t)}, \tag{8.20}$$

$$r_i(t + \epsilon) = r_i(t) - \frac{\epsilon}{m_i} p_i\left(t + \frac{\epsilon}{2}\right), \tag{8.21}$$

$$p_i(t + \epsilon) = p_i\left(t + \frac{\epsilon}{2}\right) - \frac{\epsilon}{2} \left.\frac{\partial \psi(\{r_k\})}{\partial r_l}\right|_{r_i(t+\epsilon)}. \tag{8.22}$$

We iterate these equations until $t = \tau$. Also note, that it is important to vary the pseudo time interval τ, to avoid resonant trajectories. We do so by drawing n and ϵ randomly from a uniform distribution. For the time being we will employ the simple leapfrog scheme. However, it is possible to use higher order integration schemes, provided that exact reversibility is maintained.

8.6.2. Hamiltonian mass

The Hamiltonian sampler has a large number of adjustable parameters, namely the Hamiltonian 'mass matrix, M, which can greately influence the sampling efficiency. If the individual r_k were Gaussian distributed, a good choice for HMC masses would be to set them inversely proportional to the variance of that specific r_k (Taylor et al. 2008). However, for non-Gaussian distributions, such as the lognormal Poissonian posterior, it is reasonable to use some measure of the width of the distribution (Taylor et al. 2008). Neal (1996) proposes to use the curvature at the peak.

In our case, we expanded the Hamiltonian given in equation (8.16) in a Taylor series up to quadratic order for $|r_i| << 1$. This Taylor expansion yields a Gaussian approximation of the lognormal Poissonian posterior. Given this approximation and according to the discussion in Appendix A.4, the Hamiltonian mass should be set as:

$$M_{ij} = Q_{ij}^{-1} - \left(\left(N_i^g - R_i \bar{N}\right) b - N_i^g b^2\right) \delta_{ij}^K. \tag{8.23}$$

However, calculation of the leapfrog scheme requires inversions of M. Considering the high dimensionality of the problem, inverting and storing M^{-1} is computationally impractical. For this reason we construct a diagonal 'mass matrix' from equation (8.23). We found, that choosing the diagonal of M, as given in equation (8.23), in its Fourier basis yields faster convergence for the sampler than a real space representation, since it accounts for the correlation structure of the underlying density field.

8.6.3. Parallelization

For any three dimensional sampling method, such as the lognormal Poisson sampler or the Gibbs sampler presented in chapter 7, CPU time is the main limiting factor. For this reason parallelization of the code is a crucial issue. Since our method is a true Monte Carlo method, there exist in principle two different approaches to parallelize our code.

The numerically most demanding step in the sampling chain is the leapfrog integration with the evaluation of the potential. One could therefore parallelize the leapfrog integration scheme, which requires parallelizing the fast Fourier transform. The FFTW3 library provides parallelized fast Fourier transform procedures, and implementation
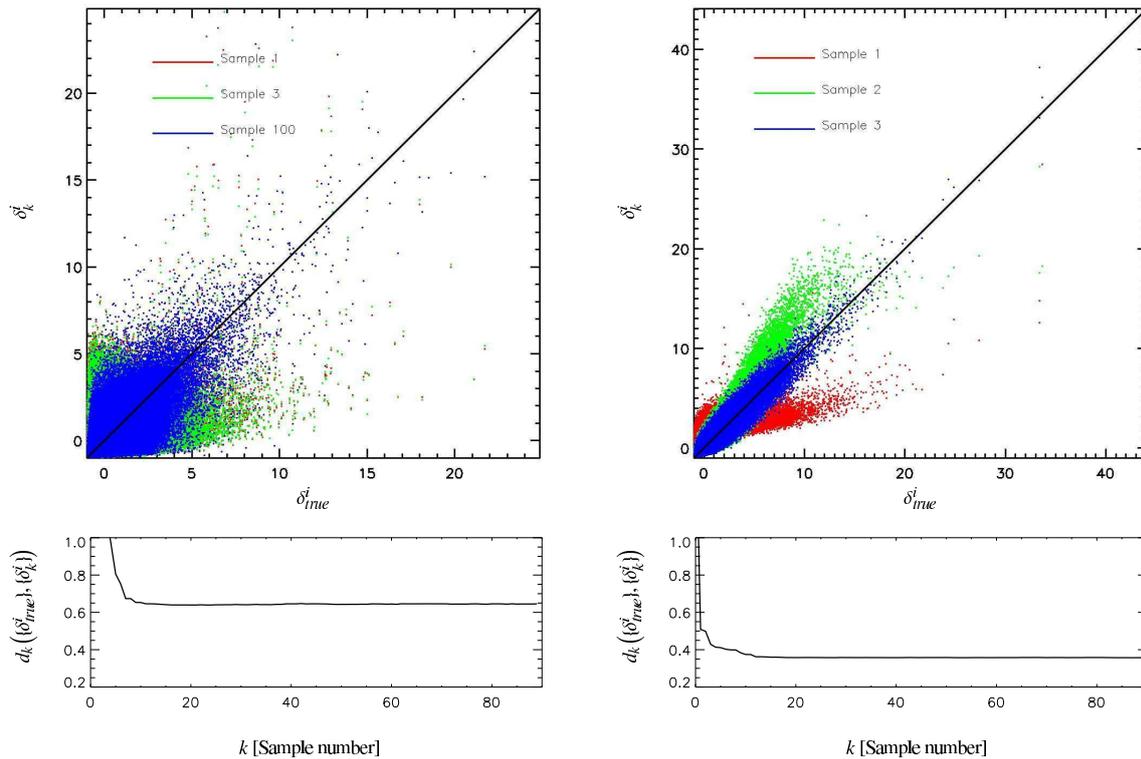
Figure 8.2.: Tests of the initial burn-in behavior for the two test cases of the fiducial calculation (right panels) and the full test taking into account the observational uncertainties of survey geometry and selection effects (left panels). The upper panels show successive point to point statistics between the individual samples and the true underlying mock signal. It can be seen that the successive Hamiltonian samples show increasing correlation with the true underlying signal. The lower panels show the successive Euclidean distances between samples and the true underlying signal during burn-in.

of those is straightforward (Frigo & Johnson 2005). However, optimal speed up cannot be achieved. The other approach relies on the fact that our method is a true Monte Carlo process, and each CPU can therefore calculate its own Markov chain. In this fashion, we gain optimal speed up and the possibility to initialize each chain with different initial guesses.

The major difference between these two parallelization approaches is, that with the first method one tries to calculate a rather long sampling chain, while the latter one produces many shorter chains.

8.7. Testing HADES

In this section, we apply HADES to simulated mock observations, where the underlying matter signal is perfectly known. With these tests we will be able to demonstrate that the code produces results consistent with the theoretical expectation. Further more, we wish to gain insight into how the code performs in real world applications, when CPU time is limited.

8.7.1. Setting up Mock observations

In this section we will describe a similar testing setup as described in chapter 7. For the purpose of this thesis we generate lognormal random fields according to the probability distribution given in equation (8.1). These lognormal fields are generated based on cosmological power-spectra for the density contrast δ. We generate these power-
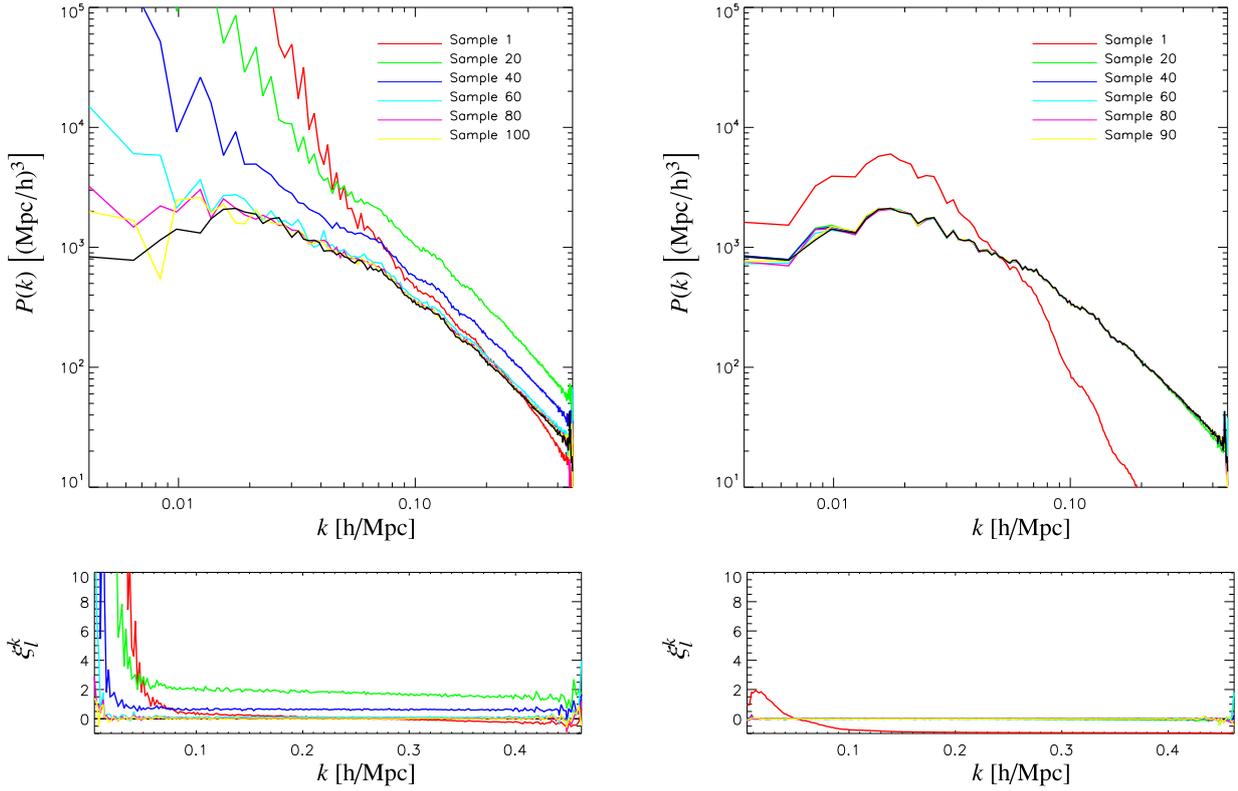
Figure 8.3.: Successive power-spectra measured from the Hamiltonian samples during burn-in. The right panels correspond to the fiducial calculation, while the left panels display the burn-in behavior of the complete observational problem. The upper panels show the convergence of the individual sample spectra towards the spectrum of the true underlying matter field realization (black curve). The lower panels display the deviation from the true underlying spectrum ξ_l^k, demonstrating good convergence at the end of the burn-in period.

spectra, with baryonic wiggles, following the prescription described in Eisenstein & Hu (1998) and Eisenstein & Hu (1999) and assuming a standard ΛCDM cosmology with the set of cosmological parameters ($\Omega_m = 0.24$, $\Omega_\Lambda = 0.76$, $\Omega_b = 0.04$, $h = 0.73$, $\sigma_8 = 0.74$, $n_s = 1$). Given these generated density fields we draw random Poissonian samples according to the Poissonian process described in equation (8.3).

The survey properties are described by the galaxy selection function F_i and the observation Mask M_i where the product:

$$R_i = F_i M_i \tag{8.24}$$

yields the linear response operator. The selection function is given by:

$$F_i = \left(\frac{r_i}{r_0}\right)^b \left(\frac{b}{\gamma}\right)^{-b/\gamma} e^{\frac{b}{\gamma} - \left(\frac{r_i}{r_0}\right)^\gamma} , \tag{8.25}$$

where r_i is the co-moving distance from the observer to the center of the ith voxel. For our simulation we chose parameters $b = 0.6$, $r_0 = 500$ Mpc and $\gamma = 2$.

The selection function and the survey geometry have been chosen to be the same as described in chapter 7. Both, the selection function and the sky mask, which defines the observed regions in the sky, have been plotted in figure 7.4. The two dimensional sky mask is given in sky coordinates of right ascension and declination. We designed the observational mask to represent some of the prominent features of the Sloan Digital Sky Survey (SDSS) mask (see Abazajian et al. 2009,for a description of the SDSS data release 7). The projection of this mask into the three dimensional volume yields the three dimensional mask M_i.

Two different projections of this generated mock galaxy survey are presented in Fig. 8.1 to give a visual impression of the artificial galaxy observation.

8.7.2. Burn in behavior

The theory described above demonstrates that the Hamiltonian sampler will provide samples from the correct probability distribution function as long as the initial conditions for the leapfrog integration are part of the posterior surface. However, in practice the sampler is not initialized with a point on the posterior surface, and therefore an initial burn-in phase is required until a point on the correct posterior surface is identified. As there exists no theoretical criterion, which tells us when the initial burn-in period is completed, we have to test this initial sampling phase through experiments. These experiments are of practical relevance for realworld applications, as they allow us to estimate how many samples are required before the sampler starts sampling from the correct posterior distribution. To gain intuition we set up a simple experiment, in which we set the initial guess for the lognormal field constant to unity ($r_k^0 = 1$). Therefore, the initial samples in the chain will be required to recover structures contained in the observation. In order to gain intuition for the behavior of our nonlinear Hamiltonian sampler, we compare two cases. The first case consists of an artificial observation including selection effects and observational mask generated as described above. The second case is a comparison calculation, where we set the observation response operator $R_i = 1$. In this latter fiducial case, only shot noise remains as observational uncertainty. It is important to note, that the individual Markov samples are unbiased in the sense that they possess the correct power information. Unlike a filter, which suppresses power in the low signal to noise regions, the Hamiltonian sampler draws true samples from the lognormal Poissonian posterior, given in equation (8.5), once burn-in is completed. Therefore, a lognormal Poissonian sample has to be understood as consisting of a true signal part, which can be extracted from the observation and a fluctuating component, which restores power lost due to the observation. This interpretation is similar to the behavior of the Gibbs sampler, as discussed in chapter 7, with the exception that there is no obvious way to separate the true signal part from the variance contribution for the nonlinear Hamiltonian sampler. Hence, the lower the signal to noise ratio of the data, the higher will be the fluctuating component.

This effect can be observed in Fig. 8.2 where we compare three successive Markov samples to the true mock signal via a point to point statistics. It can be nicely seen, that the correlation with the true underlying mock signal improves as burn-in progresses. As expected, the fiducial calculation, shown in the right panels of 8.2, has a much better correlation with the underlying true mock signal than the full observation. This is clearly owing to the fact, that the full observation introduces much more variance than in the fiducial case. To visualize this fact further, we
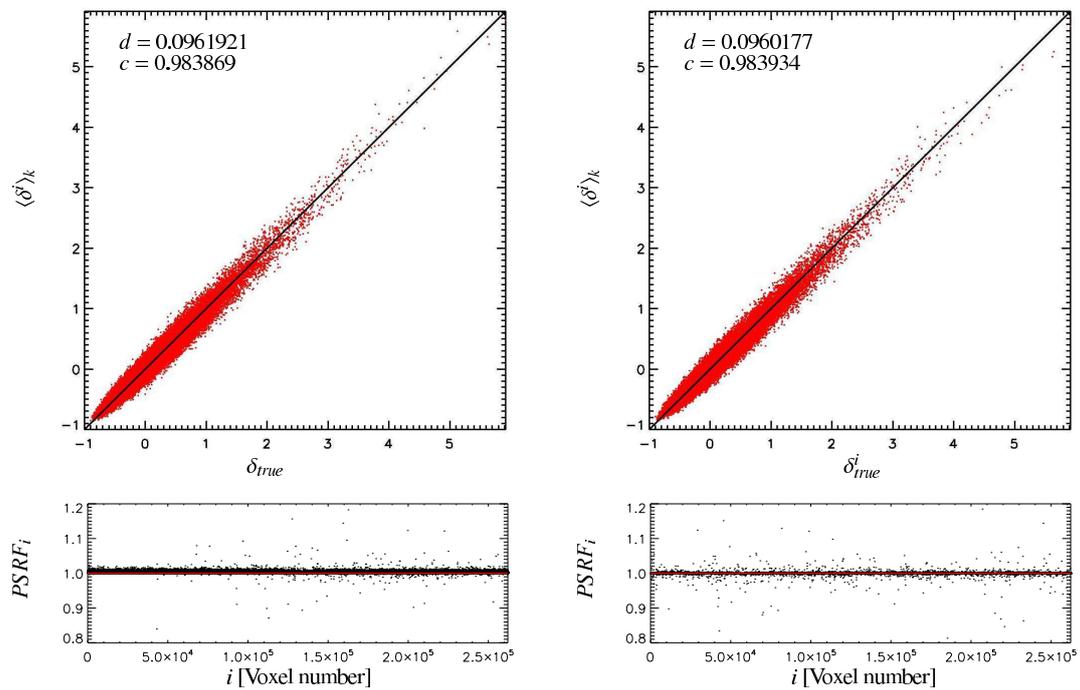
Figure 8.4.: The upper panels show the point to point statistic of the ensemble mean field to the true underlying density field in the observed region for the fiducial calculation (right panel) and the full observational problem (left panel). The numbers in the upper left part of the plots correspond to the Euclidean distance d and the correlation factor c. In the lower panels we plotted the results of the Gelman&Rubin convergence diagnostic for the according tests. The PSRF indicate good convergence.

calculate the Euclidean distance between Markov samples and the true mock signal:

$$d_k\left(\{\delta^i_{true}\}, \{\delta^i_k\}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\delta^i_{true} - \delta^i_k\right)^2}, \qquad (8.26)$$

over the progression of the Markov chain. In the lower panels of figure 8.2, it can be observed that the Eucledian distance drops initially and then saturates at a constant minimal d_k. This minimal d_k is related to the intrinsic variance contribution in the individual samples. While the variance is lower for the fiducial observation, it is higher for the full observation.

As HADES produces unbiased samples, we can gain more detailed insight into the initial burn-in phase of the Markov chain, by following the evolution of successive power-spectra measured from the samples. In addition, we measure the deviation ξ^k_l of the sample power-spectra P^k_l to the power-spectrum of the true mock matter field realization P^{true}_l via:

$$\xi^k_l = \frac{P^k_l - P^{true}_l}{P^{true}_l}. \qquad (8.27)$$

Figure 8.3 demonstrates that HADES completes burn-in after ~ 20 samples in the case of the fiducial calculation (right panels). However, the burn-in history for the full observation (left panels) reveals a more interesting behavior.

Initially, the power-spectra show huge excursions at large scales. This is due to the observational mask and the fact, that initially these regions are dominated by the constant initial guess ($r^0_k = 1$). It is interesting to note, that the first sample seems to be much closer to the true underlying power specrum at the smaller scales, while the 20th samples is much further away. This clearly demonstrates the nonlinear behavior of the lognormal Poissonian sampler. We observe, that with iterative correction of the large scale power, the entire power-spectrum progressively approaches the true mock power-spectrum. This can be seen nicely in the lower left panel of figure 8.3. After one hundred samples have been calculated the true mock power-spectrum is recovered for all following samples. Thus, the initial burn-in period for a realistic observational setting can be expected to be on the order of 100 samples. Such a burn-in period is numerically not very demanding, and can easily be achieved in even higher resolution calculations.

Further, we ran a full Markov analysis for both test cases, by calculating 20000 samples with a resolution of 64^3 voxels. We then estimate the ensemble mean and compared the recovered density field in the observed region via a point to point statistic to the true underlying mock signal. The results are presented in the upper panels of figure 8.4. It can be seen that both results are strongly correlated with the true underlying signal. To emphasize this fact, we also calculate the correlation factor given as:

$$c = \frac{\sum_{i=0}^{N-1} \delta^{true}_i \delta^{mean}_i}{\sqrt{\sum_{i=0}^{N-1}\left(\delta^{true}_i\right)^2}\sqrt{\sum_{i=0}^{N-1}\left(\delta^{mean}_i\right)^2}}. \qquad (8.28)$$

The correlation factors for the two test scenarios are also given in Fig. 8.4. They clearly demonstrate, that the Hamiltonian sampler was able to recover the underlying density field to high accuracy in both cases.

8.7.3. Convergence

Testing the convergence of Markov chains is subject of many discussions in literature (see e.g. Heidelberger & Welch 1981, Gelman & Rubin 1992, Geweke 1992, Raftery & Lewis 1995, Cowles & Carlin 1996, Hanson 2001, Dunkley et al. 2005). In principle, there exist two categories of possible diagnostics. The methods of the first category rely on comparing inter chain quantities between several chains while others try to estimate the convergence behavior from inter chain quantities within a single chain. In this thesis the widely used Gelman&Rubin diagnostic will be employed. It is based on multiple simulated chains by comparing the variances within each chain and the variance between chains (Gelman & Rubin 1992). In particular, we calculate the potential scale reduction factor (PSRF) (see Appendix A.5 for details). A large PSRF indicates that the inter chain variance is substantially greater than the intra chain variance and longer chains are required. Once the PSRF approaches unity, one can conclude that each chain has reached the target distribution.

We calculated the PSRF for each voxel of our test cases for chains with length $N_{samp} = 20000$. The results for

the two tests, as discussed above, are presented in Fig. 8.4. They clearly indicate the convergence of the Markov chains.

For the time being we use the Gelman&Rubin statistic to test convergence because of technical simplicity, although for the expense of having to calculate at least two chains. In the future we plan to explore other convergence diagnostics. In particular we are aiming at including intra chain methods as proposed in Hanson (2001) or Dunkley et al. (2005). This would allow us to detect convergence behavior within the chain during burn-in. Such a convergence criterion could then be used to adjust the Hamiltonian masses for optimal sampling efficiency, as was proposed in Taylor et al. (2008).

8.7.4. Testing with simulated galaxy surveys

In this section, we describe the application of HADES to a mock galaxy survey based on the millennium run (Croton et al. 2006). The intention of this exercise is to test HADES in a realistic observational scenario. In particular, we want to demonstrate that HADES is able to reconstruct the fully evolved nonlinear density field of the N-body simulation. The mock galaxy survey consists of a set of co-moving galaxy positions distributed in a 500 Mpc box. To introduce survey geometry and selection effects, we virtually observe these galaxies through the sky mask and according to the selection function described in section 8.7.1. The resulting galaxy distribution is then sampled to a 128^3 grid. This mock observation is then processed by HADES, which generates 20000 lognormal Poissonian samples.

In Fig. 8.5 we present successive slices through density samples of the initial burn-in period. As can be seen, the first Hamiltonian sample (upper panels in Fig. 8.5) is largely corrupted by the false density information in the masked regions. This is due to the fact, that the Hamiltonian sampler cannot be initialized with a point on the posterior surface. The initial samples are therefore required to identify a point on the according posterior surface. As can be seen, the power in the unobserved and observed regions equalizes in the following samples. Also note, that the first density sample depicts only very coarse structures. However, subsequent samples resolve finer and finer details. With the hundredth sample burn-in is completed. The lower panels of Fig. 8.5 demonstrate, that the Hamiltonian sampler nicely recovers the filamentary structure of the density field.

Being a fully Bayesian method, the Hamiltonian sampler does not aim at calculating only a single estimate, such as a mean or maximum a postiori value, it rather produces samples from the full lognormal Poissonian posterior. Given these samples we are able to calculate any desired statistical summary. In particular, we are able to calculate the mean and the corresponding variance of the Hamiltonian samples.

In Fig. 8.6 we show three different volume renderings of the ensemble mean density and the according ensemble variance fields. It can be seen that the variance projections nicely reflect the Poissonian noise structure. Comparing high density regions in the ensemble mean projections to the corresponding positions in the variance projections, reveals a higher variance contribution for these regions, as expected for Poissonian noise. This demonstrates, that our method allows us to provide uncertainty information for any resulting final estimate.

8.8. Summary and Conclusion

In this chapter the Hamiltonian Monte Carlo sampler for nonlinear large scale structure inference has been introduced and its performance in a variety of tests has been demonstrated. As already described above, according to observational evidence and theoretical considerations, the posterior for nonlinear density field inference can be approximately represented by a lognormal Poissonian distribution, up to overdensities of $\delta \sim 100$. Hence, any method aiming at precision estimation of the fully evolved large scale structure in the Universe needs to handle the nonlinear relation between observations and the signal we seek to recover. The Hamiltonian Monte Carlo sampler, presented in this thesis, is a fully Bayesian method, and as such tries to evaluate the lognormal Poissonian posterior, given in equation 8.5, via sampling. In this fashion, the scientific output of the method is not a single estimate, but a sampled representation of the multidimensional posterior distribution. Given this representation of the posterior any desired statistical summary, such as mean, mode or variances can easily be calculated. Further, any uncertainty can seamlessly be propagated to the finally estimated quantities, by simply applying the corresponding estimation procedure to all Hamiltonian samples.

Unlike conventional Metropolis Hastings algorithms, which move through the parameter space by random walk, the Hamiltonian Monte Carlo sampler suppresses random walk behavior by following a persistent motion. The HMC
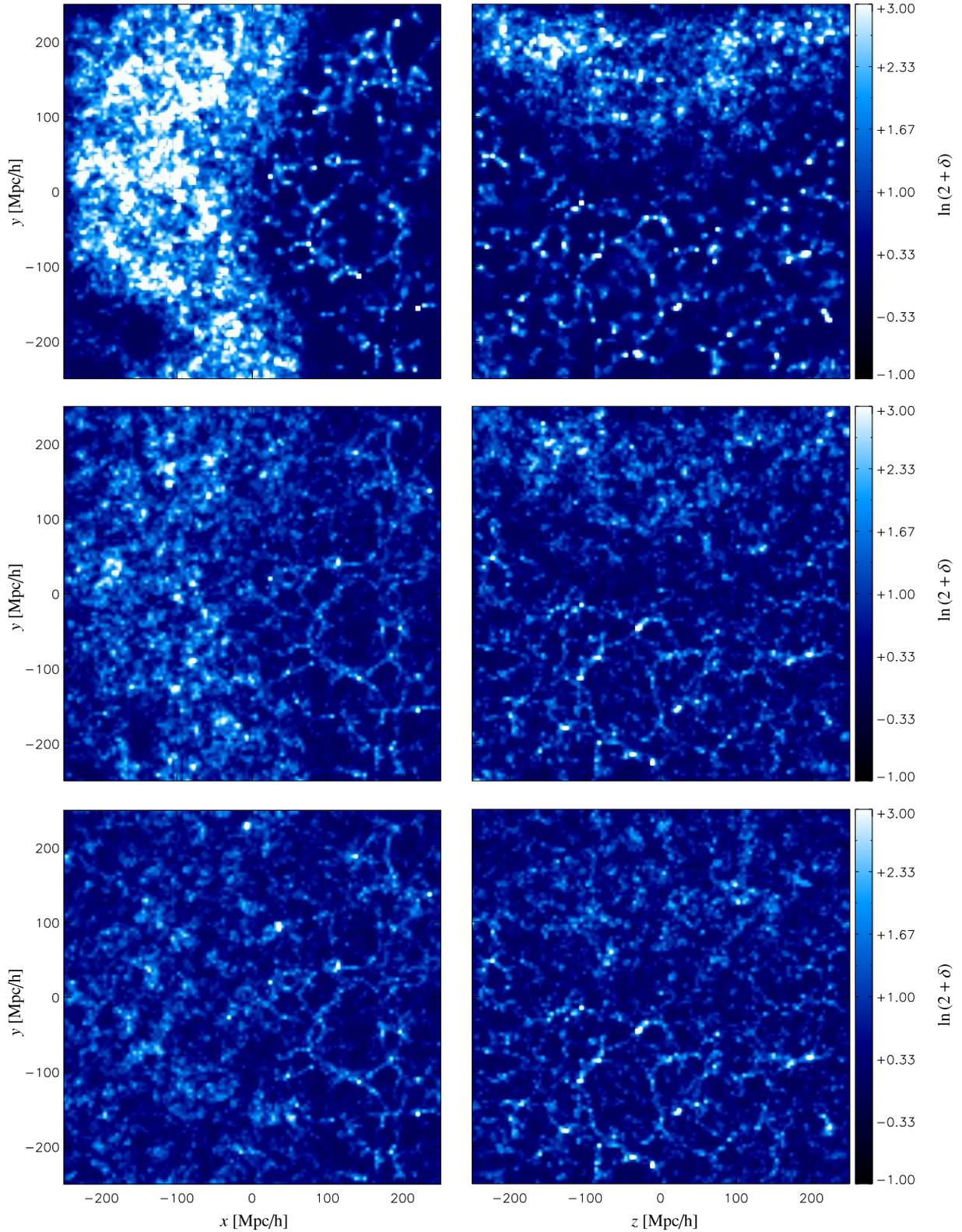
Figure 8.5.: Slices through density samples during the initial burn-in phase. The upper panels correspond to the first sample, middle panels show the tenth sample and the lower panels present the hundredth sample. Left and right panels show two different slices through the corresponding sample. It can be seen that during the initial burn-in phase power equalizes between the observed and unobserved regions. Successive samples recover finer and finer details.
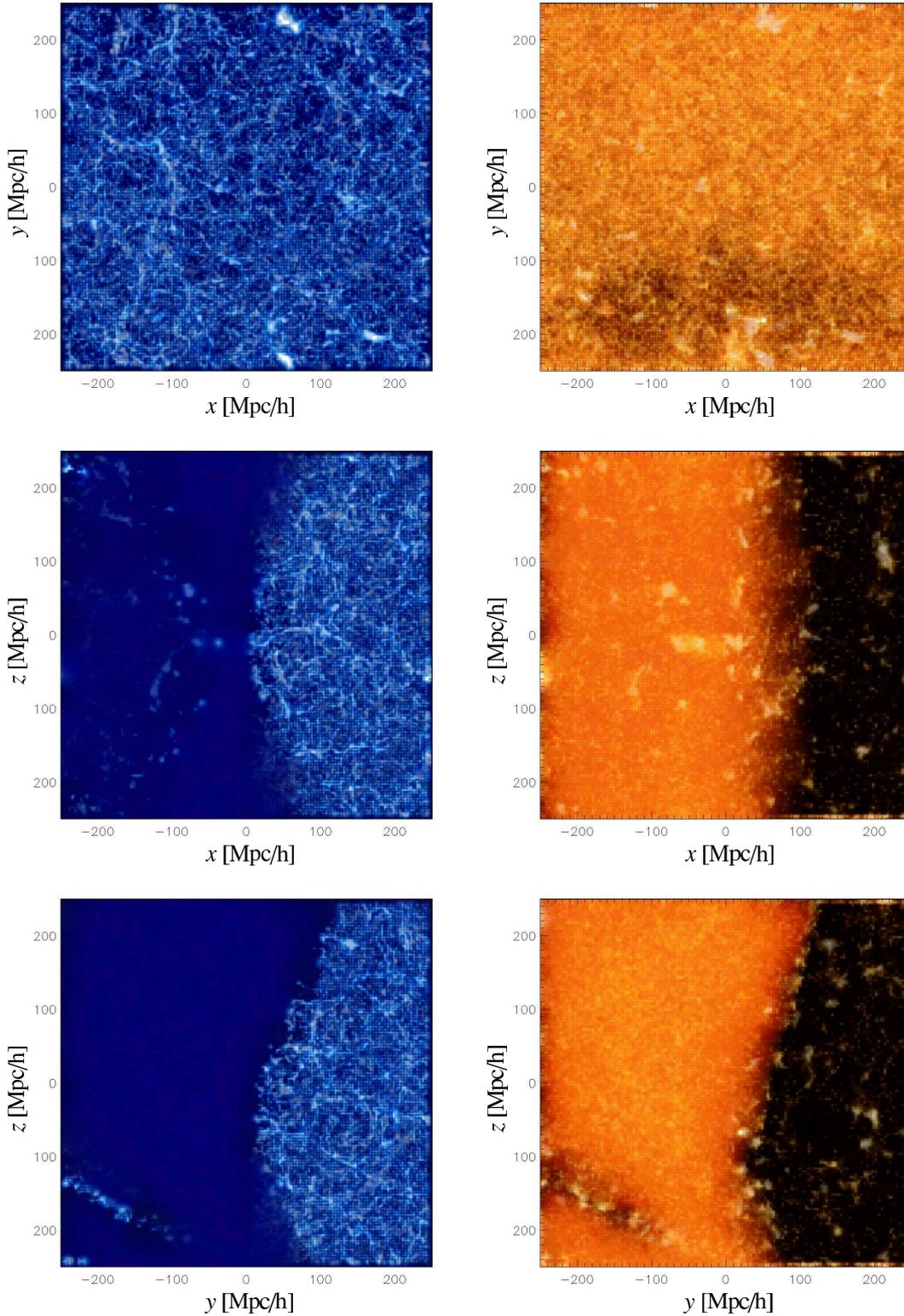
Figure 8.6.: Volume rendering of the ensemble variance (right panels) and the ensemble mean (left panels) obtained from the mock galaxy catalog analysis for three different perspectives. The mean shows filigree structures which have been recovered. It can also be seen that the ensemble variance reflects the Poissonian behavior of the noise. High density regions in the ensemble mean field correspond to regions with high variance as is expected for a Poissonian shot noise contribution. Comparison with Fig. 7.14 demonstrates that the non-linear sampler takes into account the detailed Poissonian noise structure, while a standard Wiener filtering approach does not.

113

exploits techniques developed to follow classical dynamical particle motion in potentials, which, in the absence of numerical errors, yield an acceptance probability of unity. Although, in this thesis we focused on the use of the lognormal Poissonian posterior, the method is more general. The discussion of the Hamiltonian sampler in section 8.4, demonstrates that the method can in principle take into account a broad class of posterior distributions.

In section 8.7, we demonstrated applications of the method to mock test cases, taking into account observational uncertainties such as selection effects, survey geometries and noise. These tests were designed to study the performance of the method in real world applications.

In particular, it was of interest to establish intuition for the behavior of the Hamiltonian sampler during the initial burn-in phase. Especially, the required amount of samples before the sampler starts drawing samples from the correct posterior distribution was of practical relevance. The tests demonstrated, that for a realistic setup, the initial burn-in period is on the order of ∼ 100 samples.

Further, the tests demonstrated that the Hamiltonian sampler produces unbiased samples, in the sense that each sample possesses correct power. Unlike a filter, which suppresses the signal in low signal to noise regions, the Hamiltonian sampler nonlinearly augments the poorly or not observed regions with correct statistical information. In this fashion, each sample represents a complete matter field realization consistent with the observations.

The convergence of the Markov chain was tested via a Gelman&Rubin diagnostic. We compared the intra and inter chain variances of two Markov chains each of length 20000 samples. The estimated PSRF indicated good convergence of the chain. This result demonstrates, that it is possible to efficiently sample from non-Gaussian distributions in very high dimensional spaces.

In a final test the method was applied to a realistic galaxy mock observation based on the millennium run (Croton et al. 2006). Here we introduced again survey geometry and selection effects and generated 20000 samples of the lognormal Poissonian posterior. The results nicely demonstrate that the Hamiltonian sampler recovers the filamentary structure of the underlying matter field realization. For this test we also calculated the ensemble mean and the corresponding ensemble variance of the Hamiltonian samples, demonstrating that the Hamiltonian sampler also provides error information for a final estimate.

To conclude, this chapter presents a new and numerically efficient Bayesian method for large scale structure inference, and its numerical implementation HADES. HADES provides a sampled representation of the very high dimensional non-Gaussian large scale structure posterior, conditional on galaxy observations. This permits us to easily calculate any desired statistical summary, such as mean, mode and variance. In this fashion HADES is able to provide uncertainty information to any final quantity estimated from the Hamiltonian samples. The method, as presented here, is very flexible and can easily be extended to take into account additional nonlinear observational constraints and joint uncertainties.

In summary, HADES, in its present form, provides the basis for future nonlinear high precision large scale structure analysis.

9. Nonlinear density field inference from SDSS data

Somewhere, something incredible is waiting to be known.

Carl Sagan

Abstract

This chapter describes the first non-linear, non-Gaussian full Bayesian large scale structure analysis of the cosmic density field conducted so far. The density inference is based on the latest Sloan Digital Sky Survey data release 7, which covers the northern galactic cap. For this analysis the previously described novel Bayesian sampling algorithm HADES is employed, enabling us to explore the extremely high dimensional non-Gaussian, non-linear log-normal Poissonian posterior of the three dimensional density field conditional on the data. The HADES computer algorithm permits a precise recovery of poorly sampled objects and non-linear density fields. The non-linear density inference is performed on a 750 Mpc cube with roughly 3 Mpc grid-resolution, while accounting for systematic effects, introduced by survey geometry and selection function of the SDSS, and the correct treatment of a Poissonian shot noise contribution. The high resolution results represent remarkably well the cosmic web structure of the cosmic density field. Filaments, voids and clusters are clearly visible. Further, a dynamical web classification is conducted to estimate the web type posterior distribution conditional on the SDSS data.

9.1. Introduction

As already discussed in the preceding chapters, probes of the large scale structure, such as large galaxy surveys, enable us to test current physical and cosmological theories and will generally further our understanding of the Universe. Especially a cosmographical description of the matter distribution will permit us to study details of structure formation mechanisms and the clustering behavior of galaxies as well as providing information on the initial fluctuations and large scale cosmic flows. For this reason, several different methods to recover the three dimensional density or velocity field from galaxy observations have been developed and applied to existing galaxy surveys (Ebeling & Wiedenmann 1993, Hoffman 1994, Lahav 1994, Lahav et al. 1994, Zaninetti 1995, Fisher et al. 1995, Zaroubi et al. 1995, Webster et al. 1997, Zaroubi et al. 1999, van de Weygaert & Schaap 2001, Erdoğdu et al. 2006, 2004, Kitaura et al. 2009). In particular, recently Kitaura et al. (2009) presented a high resolution three dimensional Wiener reconstruction of the Sloan Digital Sky Survey data release 6 data, which demonstrated the feasibility of high precision density field inference from galaxy redshift surveys. These three dimensional density maps are interesting for a variety of different scientific applications, such as studying the dependence of galaxy properties on their cosmic environment, increasing the detectability of the integrated Sachs-Wolfe effect in the CMB or performing constrained simulations (see e.g. Bistolas & Hoffman 1998, Lee & Lee 2008, Lee & Li 2008, Frommert et al. 2008, Klypin et al. 2003, Libeskind et al. 2010, Martinez-Vaquero et al. 2009).

However, modern precision cosmology demands an increasing control of observational systematic and statistical uncertainties, and the means to propagate them to any finally inferred quantity in order not to draw wrong conclusion about the theoretical model to be tested. For this reason, in this chapter the first application of the previously described Bayesian large scale structure inference computer algorithm HADES (HAmiltonian Density Estimation and Sampling) to data is presented. As described in chapter 8, HADES performs a full scale non-linear, non-Gaussian Markov Chain Monte Carlo analysis by drawing samples from the lognormal Poissonian posterior of the three dimensional density field conditional on the data. This extremely high dimensional posterior distribution, consisting of $\sim 10^6$ or more free parameters, is explored via a numerically efficient Hamiltonian sampling scheme

Structure type	rule
Void	$\lambda_1, \lambda_2, \lambda_3 < \lambda_{th}$
Sheet	$\lambda_1 > \lambda_{th}$ and $\lambda_2, \lambda_3 < \lambda_{th}$
Filament	$\lambda_1, \lambda_2 > \lambda_{th}$ and $\lambda_3 < \lambda_{th}$
Halo	$\lambda_1, \lambda_2, \lambda_3 > \lambda_{th}$

Table 9.1.: Rules for the dynamic classification of web types.

which suppresses the random walk behavior of conventional Metropolis Hastings algorithms by following persistent trajectories through the parameter space (Duane et al. 1987, Neal 1993, 1996). The advantages of this method are manyfold. Beside correcting observational systematics introduced by survey geometry and selection effects, the exact treatment of the non-Gaussian behavior and structure of the Poissonian shot noise contribution of discrete galaxy distributions, permits very accurate recovery of poorly sampled objects, such as voids and filaments. In addition, the lognormal prior has been demonstrated to be a good approximate statistical description for the present density field and hence enables us to infer the cosmic density field deep into the non-linear regime (see e.g. Hubble 1934, Peebles 1980, Coles & Jones 1991, Gaztanaga & Yokoyama 1993, Kayo et al. 2001). The important thing to remark about HADES is, that it does not only yield a single estimate, such as a mean, mode or variance, in fact it provides a sampled representation of the full non-Gaussian density posterior. This posterior encodes the full non-linear and non-Gaussian observational uncertainties, which can easily be propagated to any finally inferred quantity.

The application of HADES to Sloan Digital Sky Survey (SDSS) data therefore is the first non-linear, non-Gaussian full Bayesian large scale structure analysis conducted so far (SDSS; York et al. 2000). In particular, we will apply the method to the recent SDSS data release 7 (DR7) data (DR7; Abazajian et al. 2009), and will produce about 3TB of valuable scientific information in the form of 40000 high resolution non-linear density samples. The density inference is conducted on a cubic grid with side length 750 Mpc consisting of 256^3 volume elements. The recovered density field clearly reveals the cosmic web structure, consisting of voids, filaments and clusters, of the large scale structure surrounding us.

These results provide the basis for forthcoming investigations on the clustering behavior of galaxies in relation to their large-scale environment. Such analyses yield valuable information about the formation and evolution of galaxies. For example, it has long been known that physical properties such as morphological type, color, luminosity, spin parameter, star formation rate, concentration parameter, etc., are functions of the cosmic environment (see e.g. Dressler 1980, Postman & Geller 1984, Whitmore et al. 1993, Lewis et al. 2002, Gómez et al. 2003, Goto et al. 2003, Rojas et al. 2005, Kuehn & Ryden 2005, Blanton et al. 2005, Bernardi et al. 2006, Choi et al. 2007, Park et al. 2007, Lee & Lee 2008, Lee & Li 2008).

In this thesis, a preliminary examination of the dependence of stellar mass $M_\star$ and $g - r$ color of galaxies on their large-scale environment is already conducted. However, more thorough investigations will be presented in following work. Analyzing galaxy properties in the large-scale environment also requires classifying the large scale structure into different cosmic web types. We do so by following the dynamic cosmic web type classification procedure as proposed by Hahn et al. (2007) with the extension of Forero-Romero et al. (2009). The application of this procedure to our results yields the cosmic web type posterior, which provides the probability of finding a certain web type (void, sheet, filament, halo) at a given position in the volume conditional on the SDSS data. This permits simple propagation of all observational uncertainties to the final analysis of galaxy properties. This chapter is structured as follows. In section 9.2, we start by briefly reviewing the dynamic web classification procedure, as mentioned above. A description of the SDSS DR7 data and the presentation of necessary data preparation steps required to apply the analysis procedure are given in section 9.3. Specifically, we describe the preparation of the linear observation response operator and the creation of the three dimensional data cube. In the following section 9.4 we present the results obtained from the non-linear, non-Gaussian sampling procedure. We start by analyzing the convergence behavior of the Markov chain via a Gelman & Rubin diagnostic, followed by a discussion of the properties of individual Hamiltonian samples. Here we also provide estimates for the ensemble mean density field and corresponding variance. These fields depict remarkably well the properties of the cosmic web consisting of voids, filaments and halos. Based on these results we perform a simple cosmic web classification in section 9.5. In section 9.6, a preliminary examination on the correlation between the large-scale environment of galaxies and their

116
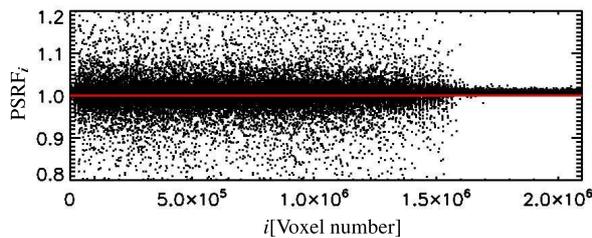
Figure 9.1.: Results of the Gelman & Rubin convergence diagnostic. The PSRF indicates convergence. As can be seen the Gelman & Rubin test converges faster in regions with good data.

physical properties is conducted. In particular, here we study the stellar mass and $g - r$ color of galaxies in relation with the density contrast δ. The chapter is concluded in section 9.7 by summarizing and discussing the results.

9.2. Classification of the cosmic web

The results generated by HADES will permit a variety of scientific analyses of the large scale structure in the observed Universe. An interesting example is to classify the cosmic web, in particular identifying different types of structures in the density field. Such an analysis is valuable, for example for studying the environmental dependence of galaxy formation and evolution (see e.g. Lee & Lee 2008, Lee & Li 2008). Since the structure classification is not always unique, we provide the full Bayesian posterior distribution of the structure type at a given position conditional on the observations.

However, to do so we first need a means to identify different structure types from the density field. Numerous methods for structure analysis have been presented in literature (see e.g. Lemson & Kauffmann 1999, Colberg et al. 2005, Novikov et al. 2006, Hahn et al. 2007, Aragón-Calvo et al. 2007, Colberg et al. 2008, Forero-Romero et al. 2009). In principle, all of these methods can be applied for the analysis of the Hamiltonian samples, however for the purpose of this thesis we follow the dynamical cosmic web classification procedure as proposed by Hahn et al. (2007). They propose to classify the large scale structure environment into four web types (voids, sheets, filaments and halos) based on a local-stability criterion for the orbits of test particles. The basic idea of this dynamical classification approach is that the eigenvalues of the deformation tensor characterize the geometrical properties of each point in space. The deformation tensor T_{ij} is given by the Hessian of the gravitational potential Φ:

$$T_{ij} = \frac{\partial^2 \Phi}{\partial x_i \, \partial x_j} \,, \tag{9.1}$$

with Φ being the rescaled gravitational potential given as (see Forero-Romero et al. 2009):

$$\nabla^2 \Phi = \delta \,. \tag{9.2}$$

It is important to note, that the deformation tensor, and the rescaled gravitational potential are both physical quantities, and hence their calculation requires the availability of a full physical density field in contrast to a smoothed mean reconstruction of the density field. As was already mentioned above, and will be clarified in section 9.4.2, the Hamiltonian samples provide such required full physical density fields. The deformation tensor can therefore easily be calculated for each Hamiltonian sample from the Fourier space representation of equation (9.1). Each spatial point can then be classified as a specific web type by considering the three eigenvalues, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, of the deformation tensor. Namely, a void point corresponds to no positive eigenvalue, a sheet to one, a filament to two and a halo to three positive eigenvalues (Forero-Romero et al. 2009). The interpretation of this rule is straight forward, as the sign of a given eigenvalue at a given position defines, whether the density field in the direction of the principal direction of the corresponding eigenvector is contracting (positive eigenvalues) or expanding (negative eigenvalues). However, Forero-Romero et al. (2009) found that rather than using a threshold value λ_{th} of zero, different positive values can yield better web classifications. For this reason, in this thesis, we use the extended

117
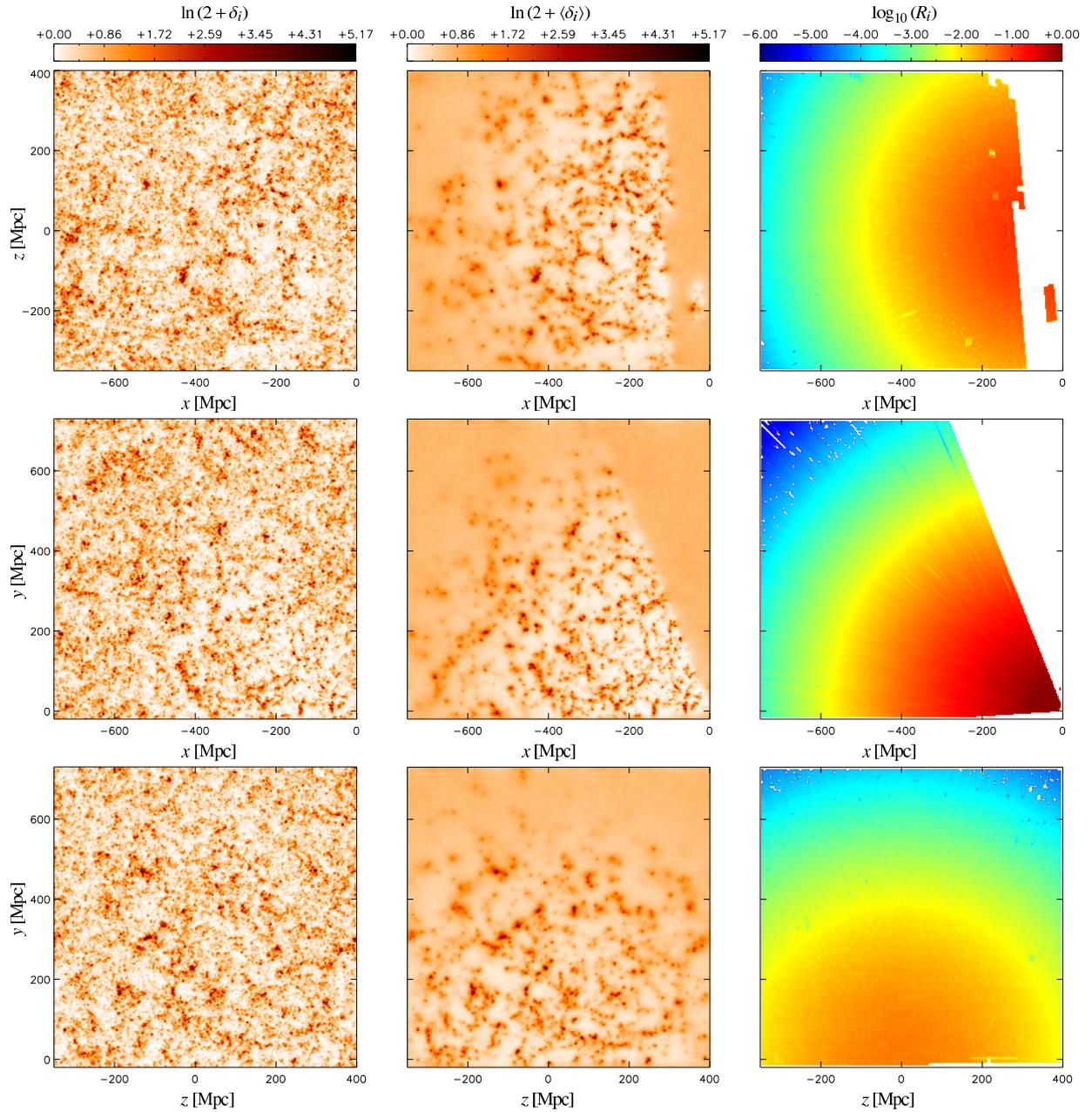
**Figure 9.2.:** Three different slices from different sides through density fields. Left panels show slices through one of the 40000 density sample, middle panels depict the estimated ensemble mean and right panels demonstrate the according slices through the three dimensional response operator R_i. It can be seen that the density sample (left panels) possesses equal power throughout the entire domain, even in the unobserved regions.
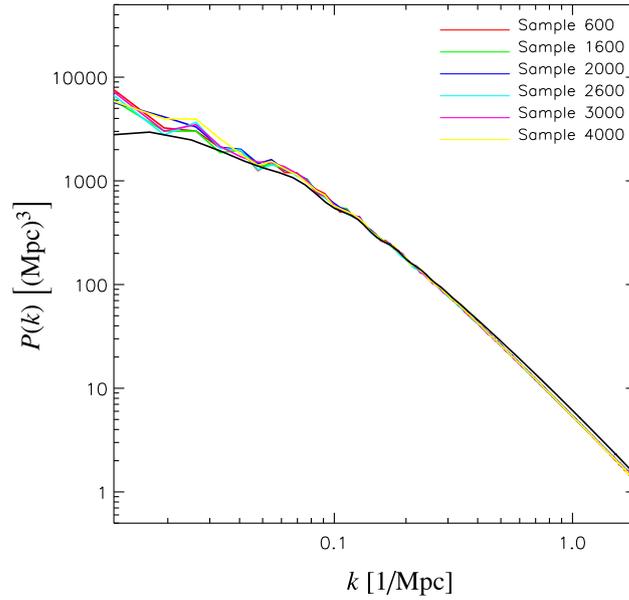
Figure 9.3.: Power-spectra of some Hamiltonian samples. The black curve corresponds to a linear ΛCDM power-spectrum.

classification procedure as proposed by Forero-Romero et al. (2009). The structures are then classified according to the rules given in table 9.1. By applying this classification procedure to all Hamiltonian samples we are able to estimate the web type posterior $\mathcal{P}(\{T_i(\vec{x}_k)\}|\{N_k^g\}, \lambda_{th})$ of four different web types ($T_1(\vec{x}_k)$ = void, $T_2(\vec{x}_k)$ = sheet, $T_3(\vec{x}_k)$ = filament, $T_4(\vec{x}_k)$ = halo) conditional on the observations and the threshold criterion λ_{th}. Details of the web type analysis of the SDSS data are described in section 9.5.

9.3. DATA

In this section we describe the SDSS galaxy sample used for the analysis. Additionally, we discuss the data preparation steps required to perform the three dimensional density inference procedure.

9.3.1. The SDSS galaxy sample

We use data from `Sample dr72` of the New York University Value Added Catalogue (NYU-VAGC) [1]. This is an update of the catalogue constructed by Blanton et al. (2005) and is based on the final data release (DR7; Abazajian et al. 2009) of the Sloan Digital Sky Survey (SDSS; York et al. 2000). Starting from `Sample dr72`, we construct a magnitude-limited sample of galaxies with spectroscopically measured redshifts in the range $0.001 < z < 0.4$, r-band Petrosian apparent magnitude $r \leq 17.6$ after correction for Galactic extinction, and r-band absolute magnitude $-23 < M_{0.1_r} < -17$. Here $M_{0.1_r}$ is corrected to its $z = 0.1$ value using the K-correction code of Blanton et al. (2003) and Blanton & Roweis (2007) and the luminosity evolution model of Blanton et al. (2003). The apparent magnitude limit is chosen in order to get a sample that is uniform and complete over the entire area of the survey. We also restrict ourselves to galaxies located in the main contiguous area of the survey in the northern Galactic cap, excluding the three survey strips in the southern cap (about 10 per cent of the full survey area). In addition, we consider only galaxies which are inside a co-moving cube of side length 750 Mpc. These restrictions result in a final sample of 463,230 galaxies.

The NYU-VAGC also provides the necessary information to correct for incompleteness in our spectroscopic sample. This includes two parts: a mask which shows which areas of the sky have been targeted and which have

---

[1] http://sdss.physics.nyu.edu/vagc/

not, and a radial selection function which gives the fraction of galaxies in the absolute magnitude range being considered that are within the apparent magnitude range of the sample at a given redshift. The mask defines the effective area of the survey on the sky, which is 6437 deg^2 for the sample we use here. This survey area is divided into a large number of smaller subareas, called *polygons*, for each of which the NYU-VAGC lists a spectroscopic completeness, defined as the fraction of photometrically identified target galaxies in the polygon for which usable spectra were obtained. Over our sample the average completeness is 0.92.

9.3.2. Completeness and selection function

Three dimensional density field inference requires the definition of the linear observation response operator R_k, as introduced in chapter 8. This response operator describes to what percentage each volume element of the three dimensional domain has been observed. It is hence a projection of the product of radial and angular selection function into the three dimensional voxelized space. In particular, we have to solve the convolution integral:

$$R_k = R(\vec{x}_k) = \int d\vec{y}\, W\left(\vec{x}_k - \vec{y}\right)\, f\left(r(\vec{y})\right)\, M\left(\alpha(\vec{y}), \delta(\vec{y})\right)\,, \tag{9.3}$$

where $W(\vec{x})$ is the voxel kernel, $f(r)$ is the radial selection function, with r being the distance from the observer and $M(\alpha, \delta)$ is the angular selection function, where α and δ are right ascension and declination respectively. We evaluate this integral numerically for the nearest grid point kernel by following different line of sights and calculating the contribution of the product of angular and radial selection function to each voxel.

As mentioned above, in this thesis we used the two dimensional sky mask and the radial selection function provided by the NYU-VAGC.

9.3.3. Creating the three dimensional data cube

The large scale structure sampler operates on a three dimensional equidistant grid. In particular, in this thesis we set up a cubic grid with side length 750Mpc and 256^3 voxels. This amounts to a resolution of $\sim$ 3Mpc voxel side length. Since our algorithm relies on the correlation function in co-moving space, all calculations are performed with co-moving length units rather than with redshift distances. For this reason, we transform all galaxy redshifts z to co-moving distances via the relation:

$$r = \int_0^{z_i} dz \frac{1}{c\,H(z)} \tag{9.4}$$

where z_i is the estimated galaxy redshift, c is the speed of light and $H(z)$ is the Hubble parameter given as:

$$H(z) = H_0 \sqrt{\Omega_m\,(1+z)^3 + \Omega_c\,(1+z)^2 + \Omega_\Lambda}\,. \tag{9.5}$$

Further, we choose a concordance ΛCDM model with a set of cosmological parameters ($\Omega_m = 0.24$, $\Omega_c = 0.00$, $\Omega_\Lambda = 0.76$, $h = 0.73$, $H_0 = h\,100\,\text{km/s/Mpc}$) (Spergel et al. 2007). With these definitions we can calculate the three dimensional cartesian coordinates for each galaxy as:

$$
\begin{aligned}
x &= r\,\cos(\delta)\,\cos(\alpha) \\
y &= r\,\cos(\delta)\,\sin(\alpha) \\
z &= r\,\sin(\delta)
\end{aligned}
\tag{9.6}
$$

where α and δ are the right ascension and declination respectively. We then sort the galaxy distribution into the three dimensional equidistant grid via a nearest grid point procedure (see e.g. Hockney & Eastwood 1988). An estimate for the expected number of galaxies $\bar{N}$ can then be calculated as:

$$\bar{N} = \frac{\sum_k N_k^g}{\sum_l R_l} \tag{9.7}$$

(see e.g. Kitaura et al. 2009, Jasche et al. 2009, for details).

120
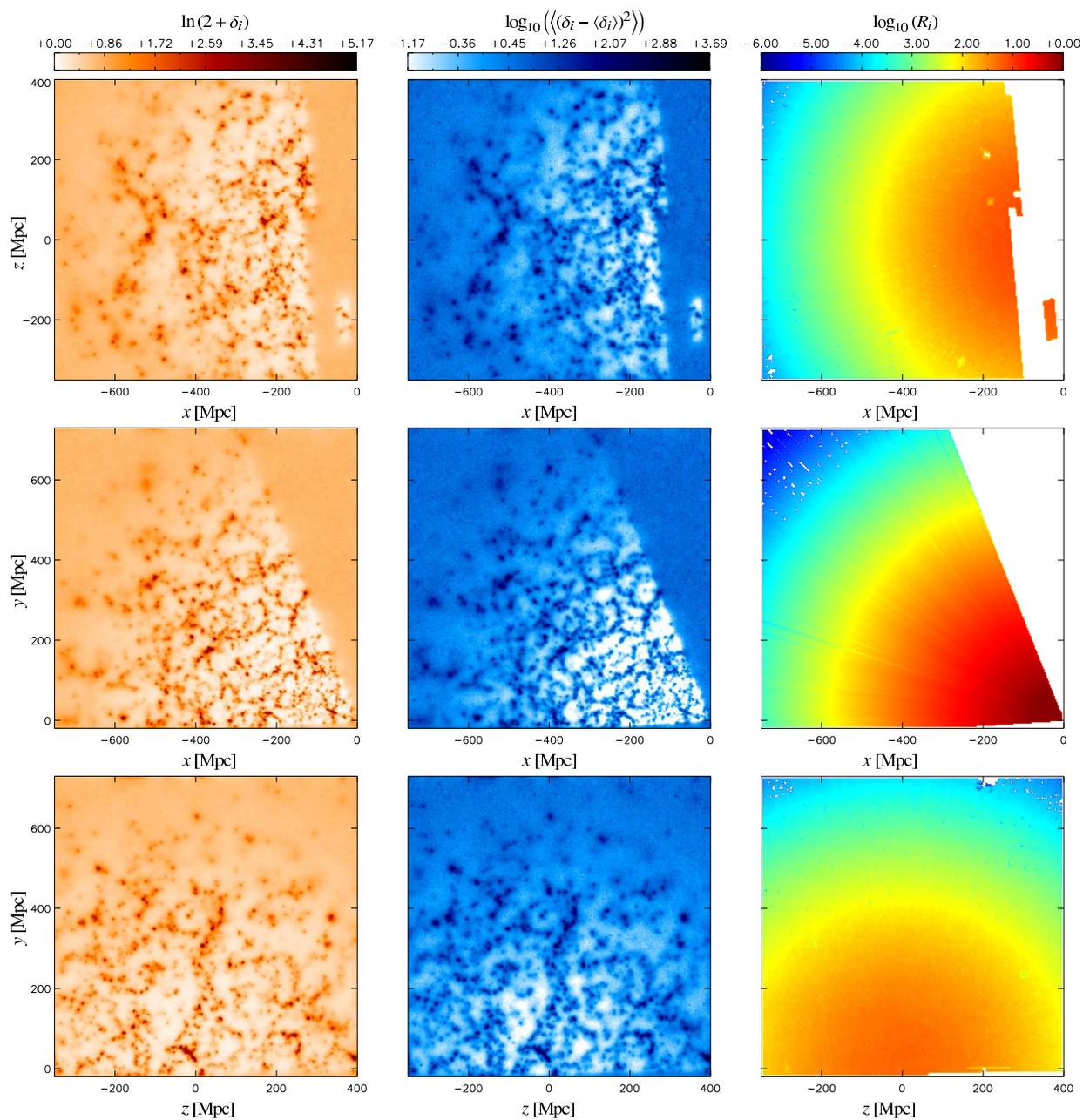
Figure 9.4.: Three different slices from different sides through ensemble mean density (left panels), ensemble variance (middle panels) and the three dimensional response operator R_i (right panels). Especially the variance plots demonstrate, that the method accounted for the full Poisonian noise structure introduced by the galaxy sample. One can also see the correlation between high density regions and high variance regions, as expected for Poissonian noise.

121

9.3.4. **Physical model**

Observations of galaxy redshifts do not permit direct inference of the underlying matter distribution. Various physical effects such as galaxy biasing and redshift space distortions must be taken into account for proper analyses. This is of particular relevance for the choice of power-spectrum required for the sampling procedure (see equation (8.1)). However, according to the discussion in Erdoğdu et al. (2004) and Kitaura et al. (2009) these effects can to a large extent be accounted for in a separate postprocessing step, once the continuous expected galaxy density field in redshift space has been obtained. For this reason, here we seek to recover the density field in redshift space permitting us to test various bias models and redshift space distortions correction methods in a subsequent step.

In particular, the relation between the true underlying dark matter density field and the expected continuous galaxy density contrast is generally very complicated and involves non-local and non-linear bias operators. Several non-local bias models have been presented, which mostly aim at correcting the large scale power in power-spectrum estimation procedures (see e.g. Tegmark et al. 2004, Seljak 2000, Peacock & Smith 2000, Hamann et al. 2008). As described in chapter 8 the Hamiltonian sampler is able to account for such bias models. Note however, that a specific bias model also fixes the model for the underlying dark matter distribution. Therefore, here, we prefer to follow the approach of previous work, setting the bias operator to a constant linear factor equal to unity (Erdoğdu et al. 2004, Kitaura et al. 2009). In this fashion, one obtains the expected continuous galaxy density contrast. As discussed in Kitaura et al. (2009), the according underlying dark matter distribution can then be simply obtained by deconvolving the results with a specific scale dependent bias model, permitting us to explore various different bias models.

In a similar manner, one can treat redshift-space distortion effects. These are mainly due to the peculiar velocities of galaxies, which introduce Doppler effects in the redshift measurement (see e.g. Kaiser 1987, Peacock & Dodds 1994, Hamilton 1998, Davis & Peebles 1983). This effect leads to a radial smearing of the observed density field in redshift-space and yields elongated structures along the line of sight, the so called *finger-of-God* effect.

Additionally, there exists a cosmological redshift-space effect which is sensitive to the global geometry of the Universe. In particular, the co-moving separation of a pair of galaxies at $z \gg 0.1$ is not determined only by their observable angular and redshift separations without specifying the geometry, or equivalently the matter content of the Universe (Magira et al. 2000). This effect yields anisotropies in the matter distribution especially at $z \geq 1$ (see e.g. Alcock & Paczyński 1979, Matsubara & Suto 1996, Ballinger et al. 1996, Popowski et al. 1998). However, for the volume considered in this thesis ($z \leq 0.27$), the dominant redshift-space distortions are due to non-linear peculiar motions of galaxies in large overdensities. This effect has pronounced consequences for the power-spectrum in redshift-space, since it suppresses power on small scales. As demonstrated in Erdoğdu et al. (2004), the redshift-space power-spectrum of a fully evolved non-linear matter distribution is very similar to a linear power-spectrum at the scales relevant for this work ($k \leq 2$ h/Mpc). Here, they used the non-linear power-spectrum fitting formula provided by Smith et al. (2003). However, the exact galaxy power-spectrum in redshift-space is not known. The work of Tegmark et al. (2006) indicates that the recovered power-spectrum of the SDSS main sample is close to a linear power-spectrum, which may be due to the fact that this galaxy sample is not strongly clustered. In this case, the redshift-space power-spectrum would be even closer to a linear power-spectrum. In any case assuming a linear power-spectrum will still permit physically accurate matter field inference in redshift-space (Erdoğdu et al. 2004). For this reason, in the absence of more precise information on the galaxy power-spectrum in redshift-space, here we will assume a linear power-spectrum, calculated according to the prescription provided by Eisenstein & Hu (1998) and Eisenstein & Hu (1999). One should also bear in mind that the data itself will govern the inference process. For this reason, power-spectra measured from the Hamiltonian samples will only be partially defined by the a priori power-spectrum guess but mostly by the data. However, we defer a more careful treatment of all physical effects including a joint inference of density field and power-spectrum to future work.

It is clear, that precise correction of these redshift-space effects requires knowledge about the peculiar velocities of all observed galaxies, which is usually not provided by galaxy redshift surveys. Therefore, precise correction of redshift-space distortions is very complicated and subject to ongoing research. In the linear regime, the theory behind the observed redshift-space distortions is well developed (Kaiser 1987, Hamilton 1998). However, in quasi-linear and non-linear regimes, we instead have to resort to making approximations or using fitting formulae based on numerical simulations (Percival & White 2009). The literature provides numerous approaches to alleviate these redshift-space distortions particularly in power-spectrum estimation. Most of these approaches aim at restoring the correct power by deconvolution with an redshift-space convolution kernel which takes into account the random pair velocities of galaxies in collapsed objects (see e.g Peacock & Dodds 1994, Ballinger et al. 1996, Jing et al. 1998,
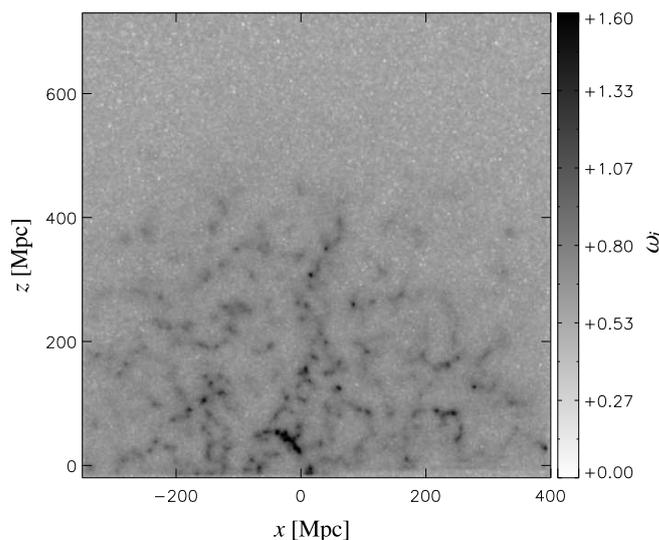
**Figure 9.5.:** The plot shows the relative density to standard deviation ratio ω_i. In comparison to the lower panels of figure 9.4 it indicates a high signal-to-noise ratio in regions of high density as expected for Poissonian noise.

Hamilton 1998, Kang et al. 2002, Jing & Börner 2004, Erdoğdu et al. 2004, Scoccimarro 2004, Cabré & Gaztañaga 2009, Percival & White 2009). Such techniques have been adopted to correct Wiener density reconstructions by applying a redshift distortion operator to the final result, in order to restore the correct power (Erdoğdu et al. 2004, Kitaura et al. 2009). However, it must be noted that this method does not account for the correction of phase information, and therefore only corrects the two-point statistics of the recovered density field.

Three dimensional density inference hence requires redshift-space distortions corrections which also account for phase information and would be dependent on the density or gravitational potential. In the linear regime it is possible to apply an inverse redshift-space operator which transforms the redshift-space density to its real-space counterpart (Taylor & Valentine 1999, D'Mellow & Taylor 2000). However, it does not account for the strongly non-linear regime which mostly generates the *finger-of-God* effect. For this reason Tegmark et al. (2004) proposed a *finger-of-God* compression method. Here they use a standard friends-of-friends algorithm to identify a cluster by taking into account different density thresholds, which set the linking length. They then measure the dispersion of galaxies about the cluster center along the line of sight and in transverse direction. If the radial dispersion exceeds the transverse dispersion, the cluster is compressed radially until the radial dispersion equals the transverse dispersion (Tegmark et al. 2004). However, it is not clear to what degree such a method would falsely isotropize filaments or under dense objects along the line of sight to spherical clusters. Such a method of isotropizing the density field, however, can also be applied in a post processing step, by noting that a density threshold refers to a linking length in the friends-of-friends algorithm.

Nevertheless, the above correction methods mask the fact that redshift-space distortions introduce statistical uncertainties. Thus unique recovery of the real-space density field is generally not possible. A full characterization of the joint uncertainties of the real-space density hence would require to carefully take into account the uncertainties introduced by redshift-space distortions or the lack of knowledge on peculiar velocities. This can be achieved by introducing a density dependent peculiar velocity sampling scheme to our method, as proposed by Kitaura & Enßlin (2008). However, we defer sampling of the peculiar velocities to future work.

9.4. Results

In this section we describe the results obtained from the large scale structure inference procedure.

9.4.1. Convergence test

HADES is a Markov Chain Monte Carlo sampler and hence we have to test, if the individual Hamiltonian samples really represent the lognormal Poissonian posterior. Convergence diagnostics for Markov chains are the subject of many discussions in the literature (see e.g. Heidelberger & Welch 1981, Gelman & Rubin 1992, Geweke 1992, Raftery & Lewis 1995, Cowles & Carlin 1996, Hanson 2001, Dunkley et al. 2005). However, here we apply the widely used Gelman & Rubin diagnostic, which is based on multiple simulated chains by comparing the variances within each chain and the variance between chains (Gelman & Rubin 1992). In particular, we calculate the potential scale reduction factor (PSRF) (see Jasche & Kitaura 2009). A large PSRF indicates that the inter chain variance is substantially greater than the intra chain variance and longer chains are required. Once the PSRF approaches unity, one can conclude that each chain has reached the target distribution. We calculated the PSRF for each voxel in our calculation domain. The result for this test is presented in Fig. 9.1. It indicates convergence of the Markov chain. However, it can be seen that some regions of the domain converge faster than others. This is due to the fact, that not all regions of the cubic volume have been observed equally. Regions which contain good observations converge faster, since there the probability distribution is narrower, while poorly or non observed regions converge slower, since the space of possible solutions is larger. Also note, that the Gelman & Rubin diagnostic is generally a conservative test, and other tests might indicate convergence much earlier. However, this test clearly demonstrates that the quality and amount of observational data can have a strong impact on the convergence behavior of the chain.

9.4.2. Hamiltonian samples

Since the Markov chain converges we can conclude, that the individual samples are really samples from the large scale structure posterior. At this point it is important to insist that the Hamiltonian samples are not the result of a filtering procedure. A filter generally suppresses the signal in low-signal to noise regions, and therefore produces biased estimates for the physical density field. This is not the case for the individual Hamiltonian samples. Since they are random realizations of the lognormal Poissonian posterior, they are unbiased density fields in the sense that they possess correct physical power throughout the entire cubic volume. As an example we present slices through an arbitrary density sample in Fig. 9.2. Already visually, one has the impression, that the density field has equal power throughout the entire domain, even in the unobserved regions. This is because the Hamiltonian sampler non-linearly augments the poorly or unobserved regions with statistically correct information. Each density sample is therefore a proper physical density field, from which physical quantities can be derived. To demonstrate this, we measure the power-spectra of some of these Hamiltonian samples. The result is presented in Fig. 9.3. As can be seen, the power-spectra of the individual samples, are very close to the assumed linear ΛCDM power-spectrum. The deviations at large scales and small scales are due to the impact of the data. At small scales the deviation can be explained by redshift space distortions, while at the largest scales cosmic variance is dominant. There is clearly no sign of artificial power loss due to the survey geometry. Since the individual samples are valid density field realizations, it is easy to derive meaningful physical quantities, such as the gravitational potential, cosmic flows or the tidal shear tensor as demonstrated in the remainder of this chapter.

9.4.3. Ensemble mean and variance

Here we want to present the ensemble mean and variance for the set of 40000 Hamiltonian samples, each consisting of 256^3 voxels. For comparison with a single density sample the middle panels of Fig. 9.2 show the corresponding slices through the ensemble mean density field, which exhibits many interesting features. First of all, it renders remarkably well the filamentary structure of our cosmic neighborhood. Many clusters, filaments and voids can clearly be seen by visual inspection. In the unobserved regions the ensemble mean density amplitudes drop to the cosmic mean for the density contrast $\delta = 0$, just as required by construction. Structures close to the observer, at cartesian coordinates $(0, 0, 0)$, are more clearly visible than structures at larger distances. Especially, filaments and voids are less prominent at larger distances. This is due to the observational response operator R_i, which due to the radial selection function drops to very low values at large distances. Therefore, once a galaxy is detected far away from the observer, it must reside inside a large overdensity, and hence inside a cluster. This expectation is clearly represented by the ensemble mean density field. Another interesting point to remark is, that the borders to the unobserved regions are not very sharp. Some of the observed information is non-linearly propagated into the unobserved regions, since our method takes into account the correlation structure of the underlying signal. It can therefore be
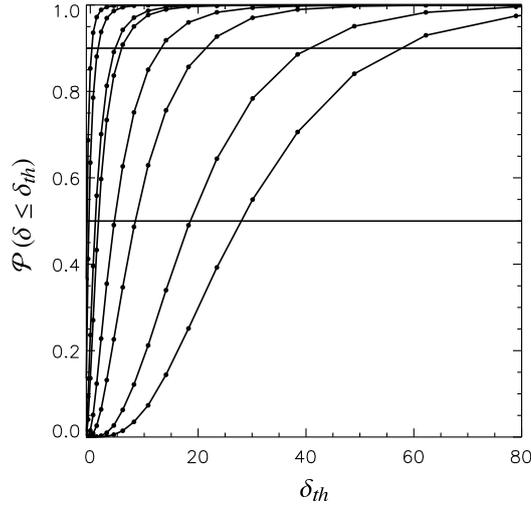
Figure 9.6.: Commulative probability distributions of the density at randomly chosen points in the volume. The cumulative probability distributions have been evaluated for 20 threshold values δ_{th}. The two horizontal lines indicate the $\mathcal{P}(\delta \leq \delta_{th}) = 0.5$ and 0.9 thresholds respectively.

seen, that some clusters and voids are interpolated further out into the unobserved regions. In comparison to the Wiener filter as previously applied to SDSS data by Kitaura et al. (2009), it seems that the Hamiltonian sampler is more conservative and less optimistic for the extrapolation of information into the unobserved region. This may be due to the fact that here we take into account the full Poissonian noise statistics rather than restricting the noise to a Gaussian approximation. Beside the ensemble mean, here we also calculate the ensemble variance per voxel, which is the diagonal of the full ensemble covariance matrix. Some slices through the ensemble mean, ensemble variance and the according slices through the observational response operator are presented in Fig. 9.4. Here the middle panels correspond to ensemble variance. At first glance, one can nicely see the Poissonian nature of the galaxy shot noise. High density peaks in the ensemble mean map correspond to high variance regions in the ensemble variance map, as expected for Poissonian noise. One can clearly see that the Hamiltonian sampler took into account the full three dimensional noise structure of the galaxy distribution. Additionally, with larger distance to the observer, the average variance increases, as is expected due to the radial selection function. It is also interesting to remark, that some voids have been detected with quite low variance, hence with high confidence. Note, however, although here we only plotted the diagonal of the density covariance matrix, the full non-diagonal covariance structure is completely encoded in the set of Hamiltonian samples, and can be taken into account for future analysis. Also note, that the variance slices show high variances in regions where many galaxies have been observed. This is a key feature of the Poisson statistics, because the standard deviation is equal to the square-root of the number of individual galaxies. That is, if there are N galaxies in each voxel, the mean is equal to N and the standard deviation is equal to $\sqrt{N}$. This makes the signal-to-noise ratio equal to $\sqrt{N}$ for such an homogeneous case. To emphasize the fact, that regions which show high variances have also high signal-to-noise ratios, we calculate the density to standard deviation ratio:

$$\omega_i = \frac{(1 + \langle \delta_i \rangle)}{\sqrt{\langle \delta_i^2 \rangle - \langle \delta_i \rangle^2}} \,. \tag{9.8}$$

The result of this calculation is presented in Fig. 9.5 for the case of the lower slices of Fig. 9.4. It clearly indicates high signal-to-noise ratios in high density regions. In addition, we also estimate the cumulative probabilities $\mathcal{P}(\delta_i \leq \delta_{th})$, at twenty different density threshold values δ_{th}, for the density found at each voxel. This cumulative probabilities are estimated from the Hamiltonian samples by:

$$\mathcal{P}(\delta_i \leq \delta_{th}) = \frac{\sum_{n=1}^{N_{\text{samp}}} \Theta(\delta_{th} - \delta_i)}{N_{\text{samp}}} \,, \tag{9.9}$$

125
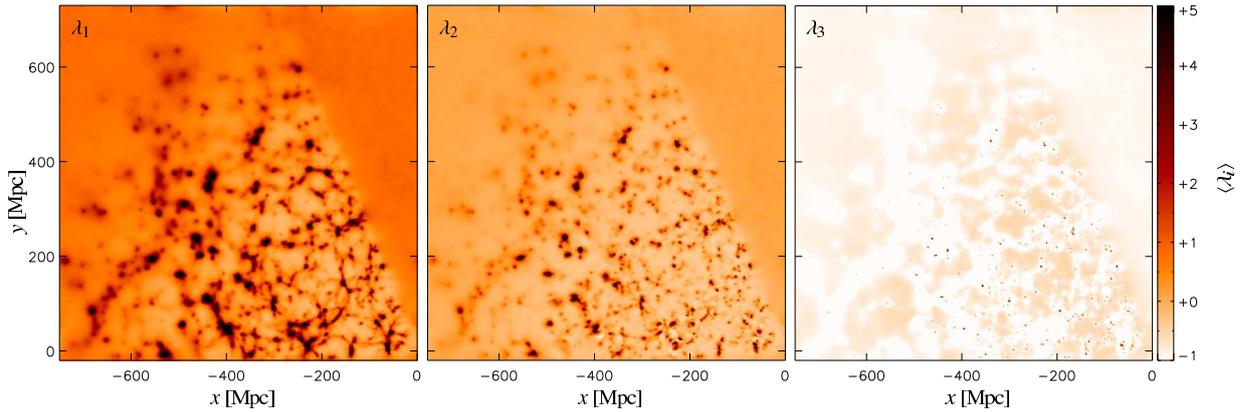
Figure 9.7.: Ensemble mean of the eigenvalues of the deformation tensor.

where n labels the individual Hamiltonian samples, N_{samp} is the total number of samples and $\Theta(x)$ is the Heaviside function. These cumulative probabilities allow for example to estimate the median density at each voxel, and can be useful, when analyzing galaxies in their cosmic environment as will be done in a following project. Some such cumulative probability distributions, chosen randomly, are shown in Fig. 9.6. As can be seen, the recovered density amplitudes extend over a large range, from small linear to very high non-linear values.

9.5. Web classification

Already in the introduction we mentioned that the results presented in section 9.4 are to be used for analyzing galaxy properties in the large-scale environment in a future work. Such analyses also require the classification of the large scale density field into different web type objects. Therefore, in order to characterize the environment of our SDSS galaxy population, here we apply the dynamic web classification procedure, as described in section 9.2, to the set of Hamiltonian samples. A similar analysis has been previously carried out by Lee & Erdogdu (2007) and Lee & Lee (2008) based on a Wiener mean density reconstructions of the 2MASS Redshift survey to study alignments of galaxy spins with the tidal field and the variation of galaxy morphological type with environmental shear.

Here we will follow a similar procedure to classify each individual voxel of a given Hamiltonian sample as one of the four web types T_i, with the different types being T_1 = void, T_2 = sheet, T_3 = filament, T_4 = halo. To do so, we perform the following three steps for an individual Hamiltonian sample:

1. Solve equation (9.1) for the deformation tensor T_{ij} by means of Fast Fourier Transform techniques

2. Solve the cubic characteristic equation for the three eigenvalues of the deformation tensor at each spatial position

3. Apply the rules given in table 9.1 to classify the web type at each spatial position for a given threshold value λ_{th}.

The result of this procedure for the nth sample is then a unit four vector $\vec{T}^n(\vec{x}_k)$ at each voxel position $\vec{x}_k$. All of the entries of this four vector are zero except for one, which indicates the web type. Applying the above method to all Hamiltonian samples will yield a set of classification four vectors, which encodes the information and uncertainty of the observations. Additionally, as an intermediate result, we obtain the set of the three eigenvalues for each individual Hamiltonian sample. Slices through their ensemble mean estimates are presented in Fig. 9.7.

However, rather than summarizing the results in terms of mean and variance here we want to estimate the full cosmic web posterior. This is achieved by counting the relative frequencies for each web type at each individual spatial coordinate within the set of Hamiltonian samples. With these definitions we yield the cosmic web posterior

for each web type as:

$$\mathcal{P}\left(\mathrm{T}_i(\vec{x}_k)|\{N_k^g\}, \lambda_{th}\right) = \frac{\sum_{n=1}^{N_{\mathrm{samp}}} \sum_{j=1}^{4} \delta_{\mathrm{T}_i(\vec{x}_k)\,\mathrm{T}_j^n(\vec{x}_k)}^{K}}{N_{\mathrm{samp}}} \, , \tag{9.10}$$

where n labels the individual Hamiltonian samples, N_{samp} is the total number of samples and δ_{ij}^{K} is the Kronecker delta. The cosmic web posterior incorporates all observational information and uncertainties, and enables us to determine how well different structures can be classified with respect to observational uncertainties.

We evaluate the cosmic web posterior for four different values of λ_{th}, with $\lambda_{th} = 0.0, 0.2, 0.4, 1.0$. Slices through the cosmic web posteriors for the four different cases are presented in Fig. 9.8. It can be clearly seen, that the properties of the survey geometry are represented by the four posterior distributions. While the web classification in the observed regions clearly follows the structure of the underlying density field, it obviously can not provide a clear classification of unobserved regions. Also with distance to the observer, the web classification becomes more and more uncertain. In this fashion, the cosmic web posterior renders the uncertainties introduced by the radial selection function and the resulting higher shot noise contribution at larger distances. The impact of the λ_{th} threshold can be observed when comparing the four cosmic web posteriors. In the case of $\lambda_{th} = 0.0$ the cosmic web consists of many small isolated voids, which occupy only a small fraction of the total area of the slice. With increasing threshold λ_{th}, voids become bigger and more connected until they completely dominate the cosmic web for the case $\lambda_{th} = 1.0$. The opposite behavior can be observed in case of the halo posteriors, as the number of clearly detected halos declines with increasing threshold λ_{th}. Following Forero-Romero et al. (2009), we also calculate the volume occupied by each web type (the volume filling fraction - VFF) and the fraction of mass contained in such a volume (mass filling fraction - MFF). The results are presented in Fig. 9.9, and show the same behavior as described in Forero-Romero et al. (2009). Figure 9.9 supports the visual impression, gained by inspection of Fig. 9.8, that especially the VFF and MFF for voids strongly depend on the threshold value λ_{th}. This shows that voids can serve as a sensitive monitor and indicator of the cosmic web (Forero-Romero et al. 2009). Unfortunately, Forero-Romero et al. (2009) do not provide an explicit gauging of the λ_{th} values from simulations. Such a gauging and hence a clear definition of the different cosmic web types would be very valuable for these types of analysis.

Having now a representation of the web type posterior we can for example calculate the odds $O_i(\vec{x}_k)$ ratio given as:

$$O_i(\vec{x}_k) = \frac{\mathcal{P}\left(\mathrm{T}_i(\vec{x}_k)|\{N_k^g\}, \lambda_{th}\right)}{1 - \mathcal{P}\left(\mathrm{T}_i(\vec{x}_k)|\{N_k^g\}, \lambda_{th}\right)} \frac{1 - \mathcal{P}\left(\mathrm{T}_i(\vec{x}_k)\right)}{\mathcal{P}\left(\mathrm{T}_i(\vec{x}_k)\right)} \, , \tag{9.11}$$

which tells us how much a specific web type is favored over all others. Here, the $\mathcal{P}\left(\mathrm{T}_i(\vec{x}_k)\right)$ can be obtained by averaging over all voxels in the volume. In example, this permits us to build a simple structure type map $m(\vec{x}_k)$ which can be used for visual analyses as presented in the next section. Such a map can be defined as:

$$m(\vec{x}_k) = \begin{cases} \mathrm{T}_i(\vec{x}_k) & \text{for } O_i(\vec{x}_k) \geq O_{th} \\ undecided & \text{else} \end{cases} \, , \tag{9.12}$$

where O_{th} is an odds threshold usually chosen larger than unity.

9.6. Galaxy properties versus LSS

In this section we present a preliminary, but intuitive examination of the correlations between the large-scale environment of galaxies and their physical properties. Here we consider two properties of galaxies: stellar mass $M_\star$ and $g - r$ color, and study how these are correlated with the overdensity δ of the large-scale environment and its type, which is one of the four web types as classified as halo, filament, sheet, and void. We will come back to this topic in a separate subsequent work by considering more physical properties of galaxies and performing more careful and quantitative analyses.

Our results are shown in figures 9.10 and 9.11 where we plot the galaxies in our sample with different stellar masses and $g - r$ colors, on top of a slice through the ensemble mean density field. In each figure the four panels correspond to four $M_\star$ intervals as indicated. The galaxies falling into a given $M_\star$ range are plotted in the corresponding panel, with red (blue) galaxies being shown as red (blue) dots. Here we classify each galaxy into red or blue population using its $g - r$ color and the luminosity-dependent divider as described in Li et al. (2006) (see their
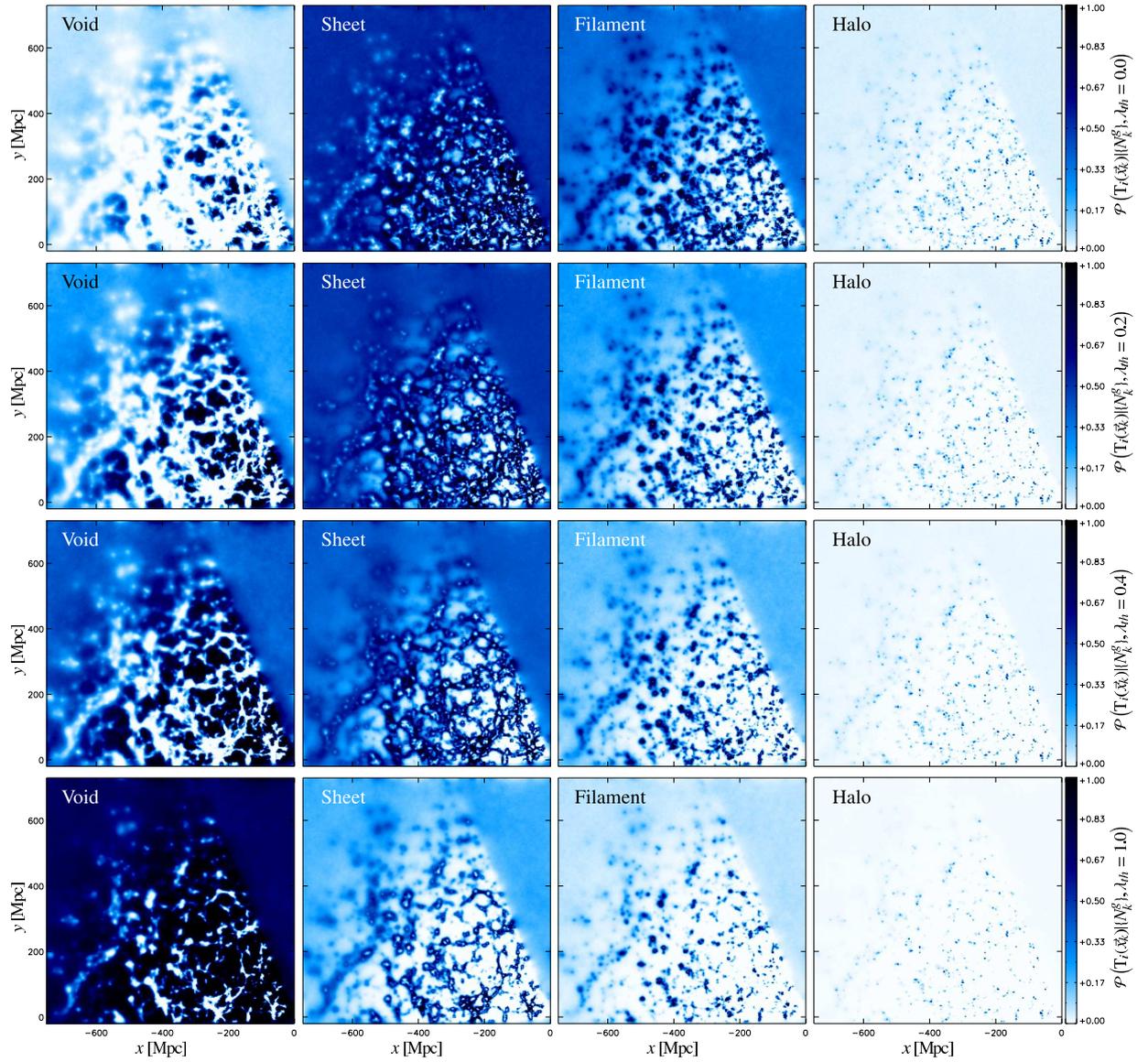
Figure 9.8.: Slices through the cosmic web posterior for the threshold values $\lambda_{th} = 0.0, 0.2, 0.4, 1.0$ (from top to bottom) for the four different web types. It is interesting to note, that sliced sheets look filamentary, while filaments piercing the slice appear as dots.
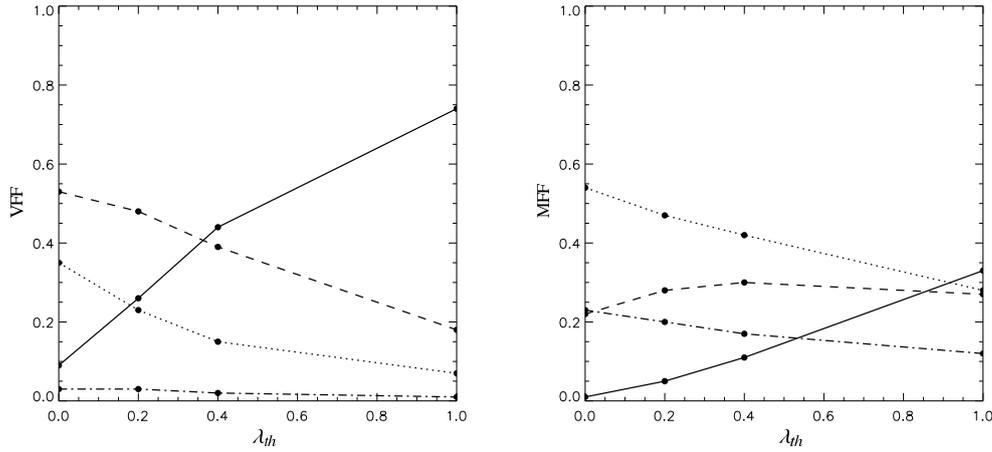
Figure 9.9.: Volume and Mass filling factors as a function of λ_{th}. Continuous lines denote voids, dashed lines sheets, dotted lines filaments and dot-dashed lines halos. Especially the void VFF and MFF respond strongly to a change in λ_{th} making them a sensitive measure of the cosmic web (Forero-Romero et al. 2009).

Eq. 7 and Table 4). The observer on Earth is at the bottom right-hand corner of the slice where $x = 0$ and $y = 0$ Mpc. The density field with $z = 302.16 \pm 4.5$ Mpc is projected onto the $x - y$ plane and is repeated in every panel. In Fig. 9.10 the background density field is coded by the mean overdensity, $\ln(2 + \langle \delta_i \rangle)$, averaged for each pixel over the z range probed and the 40,000 Hamiltonian samples. In Fig. 9.11 we present a structure type map as defined in equation (9.12) by choosing an odds threshold of $O_{th} = 1.55$ and $\lambda_{th} = 1.0$. Each pixel of this map is color-coded by the web type which is determined by our classification algorithm described above, with types of halo, filament, sheet and void being plotted in black, light grey, dark grey and white respectively.

Qualitatively the galaxies plotted in these figures appear to closely trace the underlying large-scale structure. This is not surprising because, by construction, the latter is reconstructed from the former. However, careful comparison of the different panels reveals a number of interesting trends. First, there exists a clear correlation between galaxy mass and the large-scale environment, regardless of how the environment is quantified. More massive galaxies tend to reside in regions with higher densities and more halo-like structures. At the highest masses, almost all galaxies are confined within regions of high densities, or those of halo and filament types. As $M_\star$ decreases, more and more galaxies are found in void-like regions. Second, at fixed stellar mass, galaxy color also appears to be correlated with large-scale environment. Red galaxies trace the density field more closely than blue galaxies. At all masses, the distribution of blue galaxies is more extended across the different types of structures. At low masses, the blue population dominates the galaxies in void-like environment.

These trends are consistent with recent similar studies by Lee & Lee (2008) and Lee & Li (2008), which were based on much shallower galaxy samples (thus smaller volume), and also with the clustering analyses of Li et al. (2006). More work is needed in order to have more quantitative characterization of the relationships between galaxy properties and the large-scale environment, and thus more powerful constrains on galaxy formation models. These results, in turn, can be fed back to the large scale structure inference and help to improve our cosmographical description of the Universe.

9.7. Summary and Conclusion

In this chapter the first application of the non-linear, non-Gaussian Bayesian large scale structure inference algorithm HADES to SDSS DR7 data is presented.

As already described in chapter 8, HADES is a numerically efficient implementation of a Hamiltonian Markov Chain sampler, which performs sampling in extremely high parameter spaces usually consisting of $\sim 10^7$ or more free parameters. In particular, HADES explores the lognormal Poissonian density posterior, which permits precision recovery of poorly sampled objects and density field inference deep into the non-linear regime (Jasche et al. 2009).

The large scale structure inference was conducted on a cubic equidistant grid with sidelength of 750 Mpc consist-
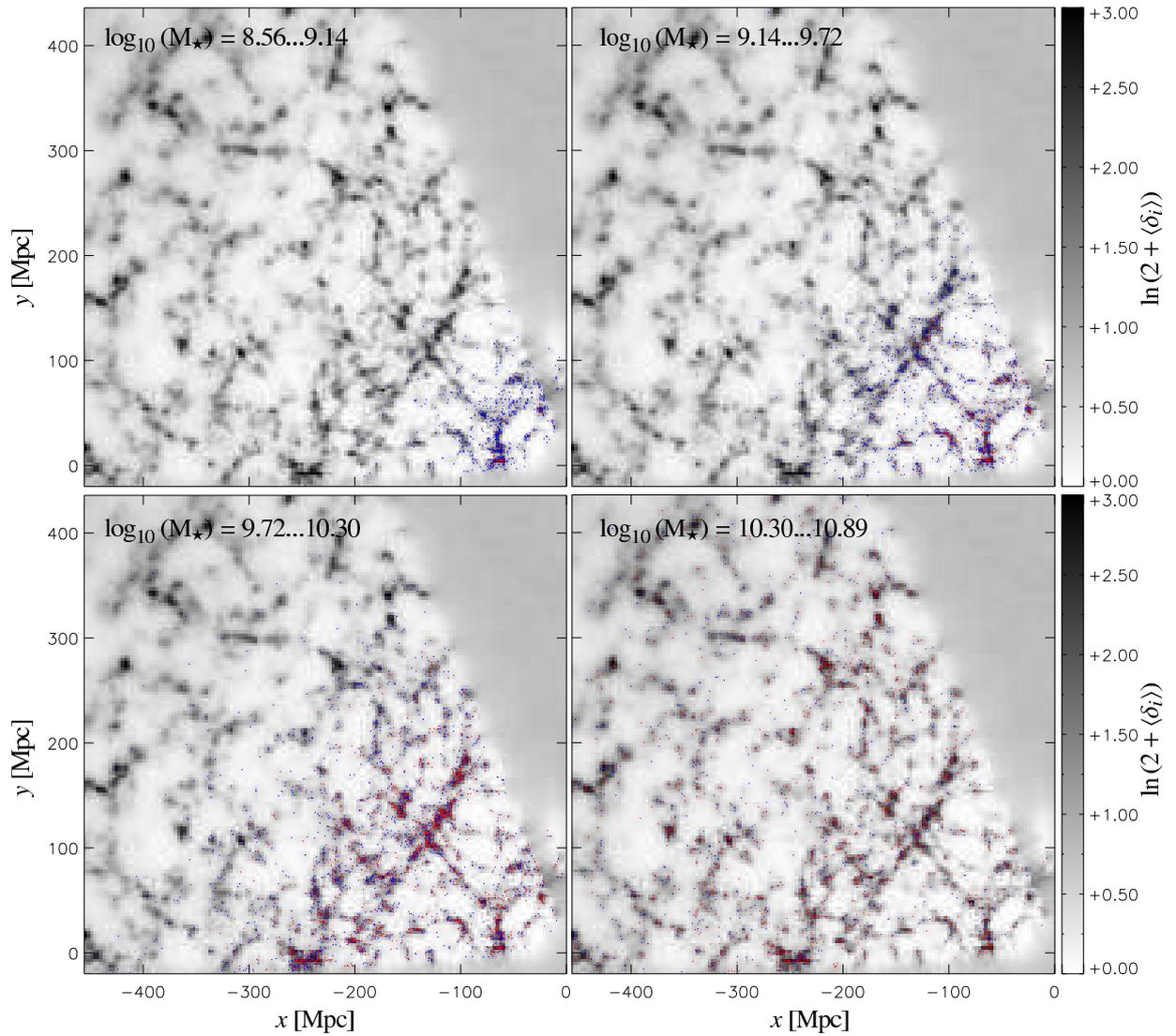
Figure 9.10.: SDSS galaxies overplotted on the ensemble mean density field. The blue and red dots denote blue and red galaxies respectively, and the different panels depict galaxies in different stellar mass $M_\star$ bins.
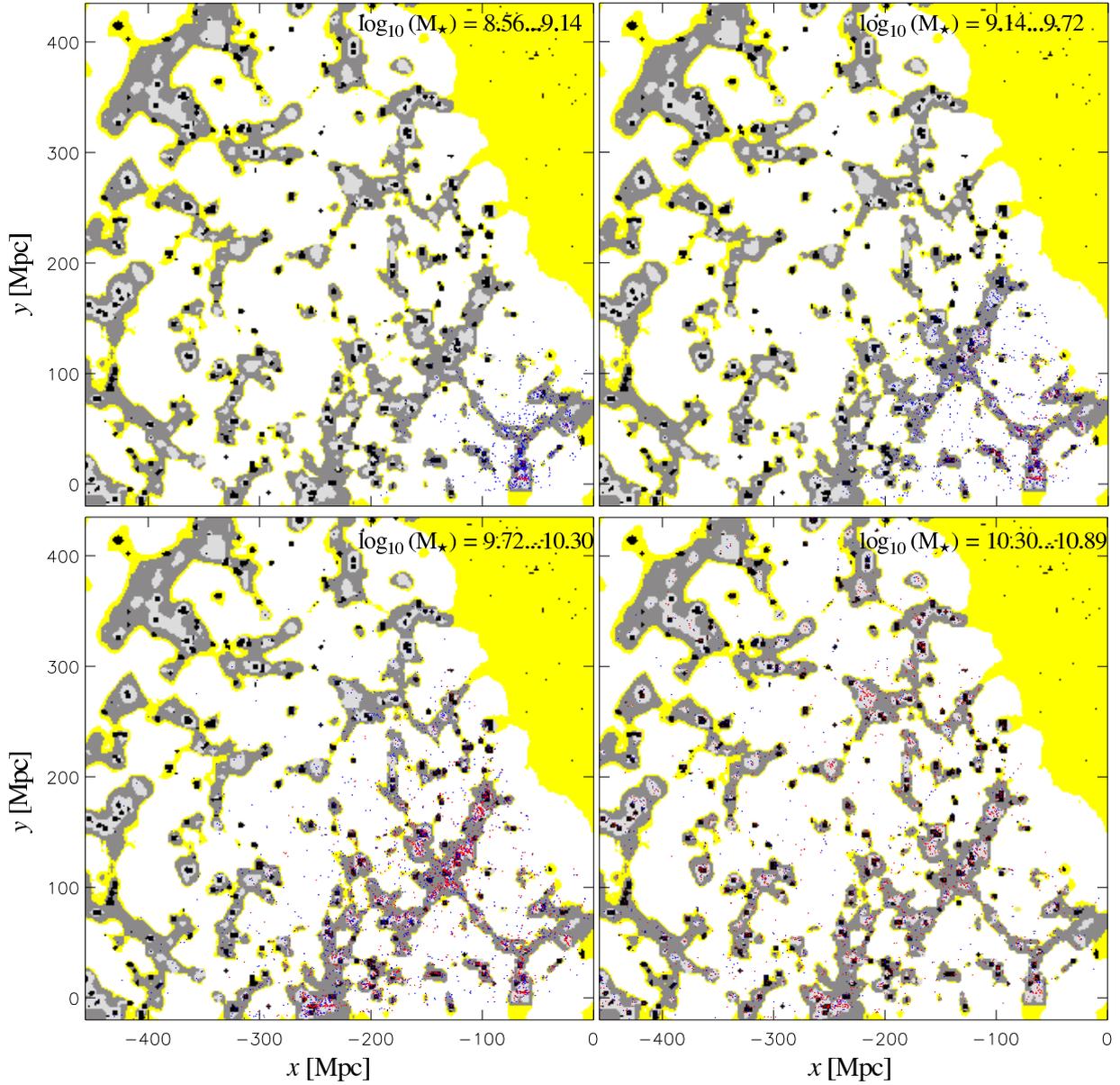
Figure 9.11.: Same as figure 9.10, but here the galaxies are overplotted on a structure type map as defined in section 9.5. The color coding denotes the web type: halo (black), filament (light grey), sheet (dark grey) and void (white). Regions, which are marked as undecided according to our criteria, equation (9.12) with $O_{th} = 1.55$, are colored yellow. Here we chose a threshold value $\lambda_{th} = 1.0$ for the web type classification.

ing of 256^3 voxels, yielding a grid resolution of about 3 Mpc. The large scale structure inference procedure correctly accounts for the survey geometry, completeness and radial selection effects as well as for the correct treatment of Poissonian noise. The analysis yielded about 3 TB of valuable scientific information in the form of full three dimensional density samples of the lognormal Poissonian density posterior. This set of density samples is thus a sampled representation of the full non-Gaussian density posterior distribution and therefore encodes all observational systematics and statistical uncertainties. Hence, all uncertainties and systematics can seamlessly be propagated to any finally inferred quantity, by simply applying the according inference procedure to the set of samples. In this fashion, the results permit us to make precise and quantitative statements about the large scale density field and any derived quantity.

We stress that our Hamiltonian samples are not the result of a filtering procedure. A filter generally suppresses the power of the signal in low signal-to-noise regions and therefore does not yield a physical meaningful density, since it lacks power in poorly or unobserved regions. However, each Hamiltonian density sample represents a complete physical matter field realization conditional on the observations, in the sense that it possesses correct physical power throughout the entire volume. Already visual inspection of these density samples shows a homogeneous distribution of power throughout the entire volume. This fact was emphasized by the demonstration of power-spectra measured from these density samples, which show no sign of being affected by lack of power or artificial mode coupling nor do they show any sign of being affected by an adaptive smoothing kernel as would be expected for filter applications. It should be noted that this fact marks the crucial difference of our method to previous filter based density estimation procedures.

In section 9.4.3, we estimated the ensemble mean and the corresponding variance from the 40000 density samples. The estimated ensemble mean nicely depicts the cosmic web consisting of filaments, voids and clusters extracted from the SDSS data. It is clear, that the ensemble mean represents the mean estimated from the lognormal Poissonian posterior conditional on the SDSS data. Therefore, it encodes the observational uncertainties and systematics. This can be seen by the fact, that the ensemble mean approaches cosmic mean density in poorly or not observed regions. Further, we plotted the according variance, which demonstrates that the non-Gaussian behavior and structure of the Poissonian shot noise was correctly taken into account in our analysis. Especially, the expected correlation between high mean density and high variance regions was clearly visible. We also estimated the cumulative probabilities for the density amplitude at each volume element, and demonstrated that the recovered density fields truly cover the broad range from linear to non-linear density amplitudes.

To characterize the environment of our galaxy sample, but also to demonstrate the advantages of the Hamiltonian samples, we performed an example cosmic web type classification in section 9.5. In particular, we followed the dynamical cosmic web classification approach of Hahn et al. (2007) with the extensions proposed by Forero-Romero et al. (2009). This procedure involves the calculation of the cosmic deformation tensor and its eigenvalues. We demonstrated that this procedure can easily be applied to the set of samples, since they represent full physical matter field realizations. As a byproduct of this procedure we estimated the ensemble mean for the three eigenvalues of the cosmic deformation tensor. Further, we classified the individual volume elements as one of the four different web types void, sheet, filament and halo. The classification into four discrete web types enabled us to explicitly estimate the cosmic web posterior, which provides the probability of finding a specific web type at a given point in the volume conditional on the SDSS data. This result is especially appealing from a Bayesian point of view, since it emphasizes the fact, that the result of a Bayesian method is a complete probability distribution rather than just a single estimate. Here we saw, that especially voids are a sensitive measure for the cosmic web. Of course, it is possible to repeat the cosmic web classification in a similar manner with any other classification procedure.

In the following section 9.6, we presented a preliminary examination of the correlation between the large-scale environment and physical properties of galaxies. In particular, we considered the stellar mass and $g - r$ color of galaxies in relation to the density contrast δ. A qualitative analysis revealed that there exist correlation between these galaxy properties and the large scale structure. In particular, massive galaxies are more likely to be found in massive structures, while low mass galaxies reside in void like regions. The plots also demonstrate the different clustering behavior of red and blue galaxies. Also note, that these observed trends are consistent with previous work (Lee & Lee 2008, Lee & Li 2008, Li et al. 2006). However, more work is required in order to provide quantitative statements. This will be done in a forthcoming publication.

The results presented in this chapter will be valuable for many subsequent scientific analyses of the dependence of galaxy properties on their cosmic environment. Herefore, particularly the Hamiltonian samples allow for a more intuitive handling of observational data, since they can be understood as full matter field realizations or different multiverses consistent with our data of the Universe we live in. Beside providing quantitative characterizations of the

large scale structure, the results also give us an intuitive understanding of the three dimensional matter distribution in our cosmic neighborhood. Future applications will also take into account non-linear bias models and peculiar velocity sampling procedures, to provide even more accurate density analyses.

In summary, this chapter demonstrates the potential of Bayesian large scale structure inference and its contribution to current and future precision analyses of our Universe.

10. Summary and outlook

Summary and conclusion

The main subject of this thesis is the development of two new Bayesian computer algorithms for the analysis of cosmological large scale structure data. Their scientific aim is to provide the joint posterior distribution of all signals, i.e. three dimensional density fields and their power-spectra, one seeks to recover from cosmological data sets. Since direct and unique inference of signals from data sets afflicted with noise and additional uncertainties is not possible, a Bayesian statistical approach is employed to fully account and correct for such sources of uncertainties and systematics, e.g. survey geometries, selection effects, biases, noise and cosmic variance. Both methods are based on very efficient implementations of Markov Monte Carlo Chains, which allow for exploring the extremely high dimensional parameter spaces of the posterior distributions (usually on the order of 10^6 or more free parameters) in computationally reasonable times. As a result, these methods provide a sampled representation of the according posterior distribution, which fully characterizes the statistical properties of the quantities inferred from the data along with their joint uncertainties. Any statistical summary, such as mean, mode or variance, can easily be reported and cosmological models can be tested via Bayesian model comparison. In this fashion, these methods allow for a statistical characterization of the observed large scale structure in the form of statistical summaries as well as for cosmographic descriptions in the form of three dimensional maps of the matter distribution in the Universe. The epistemological concepts, the mathematical framework of Bayesian statistics as well as numerical considerations are thoroughly discussed.

In Chapter 7, I present the Bayesian computer algorithm ARES (Algorithm for REconstruction and Sampling). ARES specifically aims at precision power-spectrum inference from galaxy redshift surveys. In particular, it was developed as a new approach to resolve artificial mode coupling introduced by survey geometry and selection effects. ARES jointly recovers the three dimensional density field and power-spectrum from observations by exploring their joint posterior via an iterative two step Gibbs sampling approach. Here, artificial mode coupling is resolved by solving the full three dimensional Wiener filtering equation. A numerically efficient operator based Krylov space method permits fast inversion of the Wiener filtering operator, which usually has on the order of $10^6 \times 10^6$ or more entries. Subsequent power-spectrum samples are efficiently generated from an inverse gamma distribution. Herewith it is possible to generate about 40000 full three dimensional density and power-spectra samples on a usual desktop machine within one day. The method has been thoroughly tested in a variety of mock cases. These mock observations have been designed to emulate some of the predominant features of the Sloan Digital Sky Survey, i.e. survey geometry. As a particular result it was possible to demonstrate that the maximal residual correlation due to artificial mode coupling can be reduced to less than 10%. Therefore, the method precisely corrects survey geometry and selection effects allowing for a high detectability of the BAO features in the power-spectrum. Further, the ensemble mean and the according confidence regions estimated from the set of Gibbs samples precisely depict the recovered baryon acoustic oscillations in the power-spectrum and its uncertainties. Based on the set of Gibbs samples, it is possible to provide an analytic representation of the power-spectrum posterior via the Blackwell-Rao estimator. This permits a variety of analytic analyses such as reporting higher order correlations, estimating cosmological parameters or Bayesian model comparison. Aside from power-spectra, ARES also provides three dimensional cosmographic descriptions of the three dimensional matter distribution and according uncertainties. This allows for studying the gravitational potential or the late integrated Sachs-Wolfe-effect as well as the properties of galaxies in their cosmic environment. The general framework of the Gibbs sampling approach offers the possibility to include additional parameters and uncertainties into the joint inference process. In this fashion, a set of parameters, such as density and peculiar velocity fields, power-spectra and biases, can be jointly infered from a cosmological data set.

Albeit the Gibbs sampling approach has proven to be extremely efficient and suited for three dimensional large scale structure inference, it still relies on a Gaussian approximation of the density posterior. In particular, it assumes normally distributed density fields and noise. While this is an adequate approximation for the largest scales, precision recovery of non-linear density structures requires the treatment of non-Gaussian effects, especially phase correlations. Precision recovery of poorly sampled objects, such as filaments or voids, further requires exact treat-

ment of the Poissonian shot noise generated by the discrete galaxy sample.

For these reasons, in Chapter 8, I develop a non-linear, non-Gaussian Monte Carlo sampler for the inference of the cosmic density field deep into non-linear regimes. The developed computer algorithm HADES (Hamiltonian Density Estimation and Sampling) explores the lognormal Poissonian posterior for the three dimensional density field conditional on a cosmological data set via an efficient Hamiltonian Monte Carlo sampler. Unlike conventional Metropolis Hastings algorithms, which move through the parameter space by random walk and therefore render computationally impractical in extremely high dimensions, the Hamiltonian Monte Carlo sampler suppresses random walk behavior by following a persistent motion through parameter space. As well as ARES, HADES also accounts for survey geometry, selection effects and cosmic variance. In particular, HADES can treat non-linear data models, e.g. non-local and non-linear bias models. HADES has been extensively tested in mock cases emulating the predominant features of the Sloan Digital Sky Survey. These tests clearly demonstrated the ability of HADES to recover the non-linear density field to great precision from observations. Especially tests with mock galaxy catalogs based on the Millenium run were able to recover the filamentary cosmic web with great accuracy. The method provides a sampled representation of the lognormal Poissonian distribution which characterizes the statistical properties of the inferred density field as well as its non-linear and non-Gaussian uncertainties. Aside from reporting statistical summaries and the possibility to conduct Bayesian model comparison, the availability of the sampled posterior distribution allows for straight forward non-Gaussian error propagation to any finally estimated quantity based on the inference results. Additionally, a very important result for cosmological data analysis is the demonstration that it is numerically feasible to sample in extreme high parameter spaces (e.g. 10^7 or more free parameters).

In Chapter 9 I present the application of HADES to the latest Sloan Digital Sky Survey data. This is the first non-linear, non-Gaussian, fully Bayesian large scale structure analysis of a cosmological data set conducted so far. The large scale structure inference has been performed on a cubic equidistant grid with a sidelength of 750 Mpc consisting of 256^3 voxels, yielding a grid resolution of 3 Mpc. It generated about 3 tera bytes of valuable scientific information in the form of $40,000$ Hamiltonian samples. The high resolution results show remarkable clear structures such as filaments, voids and clusters as predicted by theoretical simulations. Since individual samples represent full physical density field realizations, they can be treated as physical quantities allowing for simple straight forward estimation of physical quantities such as the gravitational potential, velocity fields and tidal shear tensors. The results of this procedure can be used in a variety of scientific applications, such as generation of integrated Sachs-Wolfe or weak lensing templates or estimation of the correlation between galaxy properties and their large scale environment. As a first step towards the latter application, I performed a dynamical cosmic web type classification based on the set of samples. This procedure involved estimation of the tidal shear tensor and its eigenvalues. In particular, it classified different voxels in the analyzed domain as either voids, sheets, filaments, or clusters. The application of this procedure yielded a probability distribution for finding a specific web type (void, sheet, filament, cluster) conditional on the SDSS data, the cosmic web type posterior. This procedure also serves as an example of simple non-linear and non-Gaussian error propagation to the finally inferred quantities, in this case the cosmic web types. A preliminary quantitative analysis revealed correlations between the properties of observed galaxies and the large scale structure. In particular, massive galaxies are more likely to be found in massive structures, while low mass galaxies reside in void like structures. In agreement with previous work, this analysis also demonstrates the different clustering behavior of red and blue galaxies.

In summary, this thesis describes the developement of two new Bayesian data analysis methods and the application to a real galaxy observation. These methods provide a full statistical characterization of the quantities to be inferred, i.e. full three dimensional density fields and power-spectra, along with their joint uncertainties. Both methods can be generalized to infer a larger set of parameters and their joint uncertainties from many different available cosmological data sets.

After having determined the parameters governing the homogeneous evolution of the Universe to the few percent level, precision cosmology needs to test the cosmological paradigm of inhomogeneous gravitational structure formation. New and more precise probes of the large scale structure together with detailed and precise data analysis methods will help to determine the physical mechanism governing the evolution of structure and might also reveal new physical phenomena. I am confident that the Bayesian methods presented in this thesis will contribute to this endeavor by opening up a new window to precision large scale structure analyses.

Outlook

The aim of the methods presented above is to provide a cosmographic as well as a statistical characterization of joint uncertainties. The resulting information can be used in a variety of cosmological analyses. Many of these analyses can be already conducted based on the results obtained during this PhD project.

Generation of templates

The three dimensional density samples allow for the calculation of the gravitational potential. On this basis templates for the integrated Sachs-Wolfe-effect conditional on the observed structures can be generated and uncertainty information can be propagated to the final estimates. These templates can be used to build matched filters and might thus enhance the detectability of the integrated Sachs-Wolfe-effect in cosmic microwave background data. Such templates are currently generated from the inferred SDSS density samples and will be used to study the integrated Sachs-Wolfe-effect in Planck data. In a similar fashion, weak lensing templates can be constructed, which in comparison to observed lensing maps might allow for testing different galaxy bias models. In addition, weak lensing signals in the sky can be predicted.

Testing correlations between galaxy properties and the large scale environment

It has been known for quite some time that physical properties of galaxies, such as morphological type, color, luminosity, spin parameter, star formation rate and concentration parameter are correlated with the large scale environment. In Chapter 9 a preliminary qualitative analysis was presented, which already indicated the existence of such correlations. A much more thorough and quantitative analysis will give information about the dependence of the parameters listed above on their cosmic environment and will reveal details about galaxy evolution. This information can then be used in future density analyses procedures to obtain more refined information on the cosmic large scale structure.

Constrained simulations

Various techniques to follow the density field back in time and to construct initial conditions for N-body simulations are described in literature. These techniques can be used to generate initial conditions from the recovered density fields. The full non-linear evolution of the density field within the simulation might then allow to follow in detail non-linear structure formation as it might have happened in our cosmic neighborhood. In addition, a set of such constrained N-body simulations, based on various density samples, would then encode the observational uncertainties, enabling us to determine the significance of the resimulated structures.

Peculiar velocity field sampling

Peculiar velocities of galaxies introduce redshift space distortions to the recovered density fields. In order to increase the precision of large scale structure inference, it is possible to extend the parameter space by the peculiar velocities of galaxies. In this fashion, peculiar velocities of galaxies would be sampled conditionally on the three dimensional density field allowing for a precise determination of the real space density field and peculiar velocities within the sampling framework. A similar technique can also be employed to process photometric redshift surveys.

Joining data sets

The Bayesian formalism provides a natural way to perform large scale structure inference from many different data sets, while at the same time taking into account the specific systematics and uncertainties of each individual data set. Joint inference from many data sets, especially those which complement the Sloan Digital Sky Survey, will greatly improve the accuracy of the inferred quantities.

Cosmological parameter sampling

An interesting approach is to eventually include a cosmological parameter sampling procedure in the Bayesian large scale structure inference framework. Such an inference machinery will then provide a full joint cosmological

analysis of cosmological datasets.

A. APPENDIX

A.1. Continuous Fourier transformation

We define the Fourier transformation as:

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(x)e^{-ikx}dx \tag{A.1}$$

and the according inverse Fourier transform is defined as:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(k)e^{ikx}dk \tag{A.2}$$

A.1.1. Convolution Theorem

The convolution theorem states that a convolution in real-space is a product in Fourier-space, and a convolution in Fourier-space is a product in real-space. This can be shown as follows.

Let $h(x)$ be the convolution of $f(x)$ with $g(x)$ defined as:

$$h(x) = \int_{-\infty}^{\infty} f(y)\,g(x-y)dy\,. \tag{A.3}$$

Applying the Fourier transform A.1 to $h(x)$ yields:

$$
\begin{aligned}
\hat{h}(k) &= \int_{-\infty}^{\infty} h(x)e^{-ikx}dx \\
&= \int_{-\infty}^{\infty} f(y) \int_{-\infty}^{\infty} e^{-ikx}\,g(x-y)dydx \\
&= \int_{-\infty}^{\infty} e^{-iky}f(y)dy \int_{-\infty}^{\infty} e^{-iky'}\,g(y')dy' \\
&= \hat{f}(k)\,\hat{g}(k)\,,
\end{aligned} \tag{A.4}
$$

where we substitute with $y' = x - y$ and used A.1 to write the last line. Hence, a convolution of the functions $f(x)$ and $g(x)$ in real-space is a product of the respective Fourier transforms $\hat{f}(k)$ and $\hat{g}(k)$ in Fourier-space.

The second part of the prove is analog to the first one presented above.

Let $\hat{h}(k)$ be the convolution of $\hat{f}(k)$ with $\hat{g}(k)$ defined as:

$$\hat{h}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(p)\,\hat{g}(k-p)dp\,. \tag{A.5}$$

Applying the inverse Fourier transform A.2 to $\hat{h}(k)$ yields:

$$
\begin{aligned}
h(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{h}(k)\,e^{ikx}dk \\
&= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \hat{f}(p) \int_{-\infty}^{\infty} e^{ikx}\,\hat{g}(k-p)\,dp\,dk \\
&= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} e^{ipx}\hat{f}(p)dp \int_{-\infty}^{\infty} e^{ik'x}\,g(k')dk' \\
&= f(x)\,g(x)\,,
\end{aligned} \tag{A.6}
$$

where we substituted with $k' = k - p$ and used A.2 to write the last line. Hence, a convolution of the functions $\hat{f}(k)$ and $\hat{g}(k)$ in Fourier-space is a product of the respective inverse Fourier transforms $f(x)$ and $g(x)$ in real-space.

A.1.2. Fourier transform of the sampling operator

The sampling function $\Pi(x)$ is given as an impulse train:

$$\Pi(x) = \sum_{m=-\infty}^{\infty} \delta^D(x - m\Delta x) \tag{A.7}$$

where Δx is the distance between two sampling positions, or the sampling interval. Applying the Fourier transform A.1 to the sampling function $\Pi(x)$ yields:

$$\hat{\Pi}(k) = \sum_{m=-\infty}^{\infty} e^{-ikm\Delta x}. \tag{A.8}$$

As can be easily seen, equation A.8 has the form of a Fourier series, in fact it is the Fourier series of a Dirac comb except for a constant factor, as we are going to show now. The Fourier series of a periodic function $f(z)$ with period Δz is defined as (Lang & Pucker 1998):

$$f(z) = \sum_{m=-\infty}^{\infty} C_m e^{-i\frac{2\pi}{\Delta z}mz} \tag{A.9}$$

with the series coefficients C_m being:

$$C_m = \frac{1}{\Delta z} \int_0^{\Delta z} f(z) e^{i\frac{2\pi}{\Delta z}mz} dz. \tag{A.10}$$

As a Dirac comb $\Pi(z)$ is periodic with period Δz, it can be represented by a Fourier series as:

$$\Pi(z) = \sum_{m=-\infty}^{\infty} \delta^D(z - m\Delta z) = \frac{1}{\Delta z} \sum_{m=-\infty}^{\infty} e^{-i\frac{2\pi}{\Delta z}mz}. \tag{A.11}$$

With this expression we can express $\hat{\Pi}(k)$ as:

$$\begin{aligned}\hat{\Pi}(k) &= p_s \sum_{m=-\infty}^{\infty} \frac{1}{p_s} e^{-ikm\Delta x} \\ &= p_s \sum_{m=-\infty}^{\infty} \delta^D(k - m\,p_s), \end{aligned} \tag{A.12}$$

where we introduced the periodicity length p_s in Fourier-space and the uncertainty principle $p_s\,\Delta x = 2\pi$. Equation A.12 shows that the Fourier transform of a Dirac comb is again a Dirac comb. But it is important to note that this proof is only valid under the assumption that the Dirac comb samples at a certain period. As soon as the sampling intervals of the Dirac comb become non-periodic this proof does not apply any longer.

A.1.3. Fourier transform pair of the ideal low pass filter

The ideal low pass filter is defined by:

$$\hat{W}(k) = \begin{cases} 1 & \text{for } |k| < p_{max} \\ 0 & \text{for } |k| \geq p_{max} \end{cases} \tag{A.13}$$

Applying the inverse Fourier transform A.2 we yield:

$$
\begin{aligned}
W(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{W}(k) e^{ikx} dk \\
&= \frac{1}{2\pi} \int_{-p_{max}}^{p_{max}} e^{ikx} dk \\
&= \frac{1}{i\,2\pi\,x} \left(e^{ip_{max}x} - e^{-ip_{max}x} \right) \\
&= \frac{sin(p_{max}x)}{\pi\,x} \\
&= \frac{p_{max}}{\pi} \, sinc(p_{max}x)
\end{aligned}
\tag{A.14}
$$

where we have introduced the sinus cardinal $sinc(x) = sin(x)/x$. Hence, the real-space representation of the ideal low pass filter is a sinc function.

A.1.4. Fourier transform of the finite sum sampling operator

The finite sum sampling operator is given as:

$$
\Pi_N(x) = \sum_{j=-N/2}^{N/2} \delta^D(x - j\Delta x).
\tag{A.15}
$$

Applying the Fourier transform A.1 then yields:

$$
\hat{\Pi}_N(p) = \sum_{j=-N/2}^{N/2} e^{-\sqrt{-1}\,p\,j\,\Delta x} = \sum_{j=-N/2}^{N/2} \left(e^{-\sqrt{-1}\,p\,\Delta x} \right)^j.
\tag{A.16}
$$

Using the closed form of the geometric series:

$$
\sum_{m=L}^{U} a^m = \frac{a^L - a^{U+1}}{1 - a}
\tag{A.17}
$$

with $a = e^{-\sqrt{-1}\,p\,\Delta x}$, $L = -N/2$ and $U = N/2$ yields:

$$
\begin{aligned}
\hat{\Pi}_N(p) &= \frac{e^{\sqrt{-1}\,p\,\Delta x \frac{N}{2}} - e^{-\sqrt{-1}\,p\,\Delta x (\frac{N}{2}+1)}}{1 - e^{-\sqrt{-1}\,p\,\Delta x}} \\
&= \frac{e^{\sqrt{-1}\,p\,\frac{\Delta x}{2}(N+1)} - e^{-\sqrt{-1}\,p\,\frac{\Delta x}{2}(N+1)}}{e^{\sqrt{-1}\,p\,\frac{\Delta x}{2}} - e^{-\sqrt{-1}\,p\,\frac{\Delta x}{2}}} \\
&= \frac{\sin\left(\frac{k\Delta x}{2}(N+1) \right)}{\sin\left(\frac{k\Delta x}{2} \right)} \\
&= \text{asinc}_N(p),
\end{aligned}
\tag{A.18}
$$

where we have defined the aliased sinc function $\text{asinc}_N(p) = \sin\left((k\Delta x)/2(N+1)\right) / \sin\left((k\Delta x)/2\right)$.

A.1.5. Ideal discretization kernel

To discretize the continuous and non-periodic function in real and Fourier space, one has to convolve the signal first with the replication kernel, and then apply a low-pass filter, or vice versa. Thus, the ideal discretization kernel will

be a convolution of the ideal low-pass filter, with the replication operator. This can be written as follows:

$$\Psi(x) = \int_{-\infty}^{\infty} dz W(x-z) \Pi_R(z). \tag{A.19}$$

According to the convolution theorem we can express the Fourier transform $\hat{\Psi}(p)$ as:

$$\begin{aligned}
\hat{\Psi}(p) &= \hat{W}(p)\hat{\Pi}_R(p) \\
&= \frac{2\pi}{L} \sum_{j=-\infty}^{\infty} \delta^D(p - j\Delta p)\, \hat{W}(p).
\end{aligned} \tag{A.20}$$

Applying an inverse Fourier transform then yields:

$$\begin{aligned}
\Psi(x) &= \frac{1}{L} \sum_{j=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ipx} \delta^D(p - j\Delta p)\, \hat{W}(p) dp \\
&= \frac{1}{L} \sum_{j=-\infty}^{\infty} e^{ij\Delta p x} \hat{W}(j\Delta p) \\
&= \frac{1}{L} \sum_{j=-M/2}^{M/2} \left(e^{i\Delta p x} \right)^j,
\end{aligned} \tag{A.21}$$

where we made use of the fact, that the ideal low-pass filter given by A.13 vanishes outside the base-band. One can now use the closed form of the geometric series given in equation (A.19) to yield:

$$\begin{aligned}
\Psi(x) &= \frac{1}{L} \frac{e^{-ix\Delta p M/2} - e^{ix\Delta p (M/2+1)}}{1 - e^{ix\Delta p}} \\
&= \frac{1}{L} \frac{e^{ix\frac{\Delta p}{2}}}{e^{ix\frac{\Delta p}{2}}} \frac{e^{-ix\Delta p(M/2+\frac{1}{2})} - e^{ix\Delta k(M/2+\frac{1}{2})}}{e^{-ix\frac{\Delta k}{2}} - e^{ix\frac{\Delta k}{2}}} \\
&= \frac{1}{L} \frac{sin\left(x\frac{\Delta p}{2}(M+1)\right)}{sin\left(x\frac{\Delta p}{2}\right)} \\
&= \frac{1}{L} asinc_M(x).
\end{aligned} \tag{A.22}$$

Thus the ideal discretization kernel is an aliased sinc function in real-space.

A.1.6. Discrete Fourier transformation

we define the numerical Fourier transformation as:

$$y_k = \hat{C} \sum_{j=0}^{N-1} x_j e^{-2\pi jk \frac{\sqrt{-1}}{N}} \tag{A.23}$$

and the inverse Fourier transform

$$x_j = C \sum_{k=0}^{N-1} y_k e^{2\pi jk \frac{\sqrt{-1}}{N}} = C \sum_{k=-(\frac{N}{2}-1)}^{\frac{N}{2}} y_k e^{2\pi jk \frac{\sqrt{-1}}{N}}. \tag{A.24}$$

The normalization coefficients C and $\hat{C}$ are chosen such that they fulfill the requirement:

$$C\hat{C} = \frac{1}{N}. \tag{A.25}$$

A.1.7. Discrete mode coupling function

Here we proof the equality

$$\hat{U}(\Delta p\,(i-j)) = \hat{C}\sum_{k=0}^{N-1} e^{2\pi\,\sqrt{-1}k\frac{(j-i)}{N}} = \hat{C}\,N\,\delta_{ij}^{K}, \tag{A.26}$$

by evaluating the sum. First we discuss the case $j \neq i$ where we will consider the sum

$$\begin{aligned}
\sum_{k=0}^{N-1} e^{2\pi k\frac{\sqrt{-1}}{N}(j-i)} &= \sum_{k=0}^{N-1}\cos\frac{2\pi}{N}k(j-i) + \sqrt{-1}\sin\frac{2\pi}{N}k(j-i) \\
&= 1 + \sum_{k=1}^{N-1}\cos\frac{2\pi}{N}k(j-i) + \sqrt{-1}\sum_{k=1}^{N-1}\sin\frac{2\pi}{N}k(j-i)
\end{aligned} \tag{A.27}$$

By making use of the trigonometric sums:

$$\sum_{k=1}^{N-1}\cos(kx) = \frac{\sin\left(\left(N-\frac{1}{2}\right)x\right)}{2\sin\left(\frac{x}{2}\right)} - \frac{1}{2} \tag{A.28}$$

and

$$\sum_{k=1}^{N-1}\sin(kx) = \frac{\sin(\frac{N-1}{2}x)\sin(\frac{N}{2}x)}{\sin(\frac{x}{2})} \tag{A.29}$$

we yield:

$$\begin{aligned}
&= 1 + \sum_{k=1}^{N-1}\cos\frac{2\pi}{N}k(j-i) + \sqrt{-1}\sum_{k=1}^{N-1}\sin\frac{2\pi}{N}k(j-i) \\
&= 1 + \frac{\sin\left(\left(N-\frac{1}{2}\right)\frac{2\pi}{N}(j-i)\right)}{2\sin(\frac{\pi}{N}(j-i))} - \frac{1}{2} + \sqrt{-1}\frac{\sin\left(\frac{N-1}{N}\pi(j-i)\right)\sin(\pi(j-i))}{sin(\frac{\pi}{N}(j-i))}
\end{aligned} \tag{A.30}$$

because of $\sin(\pi(j-i)) = 0$ for all j and i the last term vanishes, and we yield

$$= \frac{1}{2} + \frac{\sin\left(2\pi(j-i) - \frac{\pi}{N}(j-i)\right)}{2\sin(\frac{\pi}{N}(j-i))}. \tag{A.31}$$

Since the sinus is 2π-periodic we can write:

$$\sin\left(2\pi(j-i) - \frac{\pi}{N}(j-i)\right) = -\sin\left(\frac{\pi}{N}(j-i)\right) \tag{A.32}$$

Using equation A.32 in equation A.31 we yield:

$$= \frac{1}{2} - \frac{\sin\left(\frac{\pi}{N}(j-i)\right)}{2\sin(\frac{\pi}{N}(j-i))} = \frac{1}{2} - \frac{1}{2} = 0. \tag{A.33}$$

Hence, for the case $j \neq i$ the sum will always be zero. Now we discuss the case $j = i$, where the sum yields:

$$\sum_{k=0}^{N-1} e^{2\pi k\frac{\sqrt{-1}}{N}(j-i)} = \sum_{k=0}^{N-1} 1 = N. \tag{A.34}$$

Hence, we end up with the result:

$$\sum_{k=0}^{N-1} e^{2\pi k \frac{\sqrt{-1}}{N}(j-i)} = \begin{cases} N & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \tag{A.35}$$

Which yields:

$$\sum_{k=0}^{N-1} e^{2\pi k \frac{\sqrt{-1}}{N}(j-i)} = N\delta_{ij}^K. \tag{A.36}$$

Thus, equation (A.26) has been proven.

A.2. Change to FFT representation

For computational convenience, we would like to consider the set of coefficients $\hat{S}_{kl}$, which are the matrix coefficients of the Fourier transformed signal covariance matrix S. The signal covariance matrix coefficients S_{ij} are related to the Fourier matrix coefficients $\hat{S}_{kl}$ via the discrete Fourier transform defined in Appendix A.1.6:

$$\begin{aligned} S_{ij} &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi ik \frac{\sqrt{-1}}{N}} \hat{S}_{kl} \, e^{-2\pi jl \frac{\sqrt{-1}}{N}} \\ &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N}(ik-jl)} \hat{S}_{kl}, \end{aligned} \tag{A.37}$$

where N is the number of Volume cells which we consider in our calculation, and C is the normalization of the discrete Fourier Transform. The matrix elements of the Jacobi matrix for this coordinate transformation can then be written as:

$$\begin{aligned} \mathcal{J}_{(ij)(mn)} = \frac{\partial S_{ij}}{\partial \hat{S}_{mn}} &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N}(ik-jl)} \frac{\partial \hat{S}_{kl}}{\partial \hat{S}_{mn}} \\ &= C^2 \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N}(ik-jl)} \delta_{km}^K \delta_{ln}^K \\ &= C^2 e^{2\pi \frac{\sqrt{-1}}{N}(im-jn)} \\ &= C^2 A_{(ij)(mn)}, \end{aligned} \tag{A.38}$$

where we defined the transformation matrix $A_{(ij)(mn)} \equiv e^{2\pi \frac{\sqrt{-1}}{N}(im-jn)}$. The norm of the Jacobi determinant can then be calculated in matrix notation as:

$$\begin{aligned} |\det(\mathcal{J})| &= \left| \det\left(C^2 \, A\right) \right| \\ &= \left| \det\left(C^2 \, \mathbb{1}_{N^4} \, A\right) \right| \\ &= \left| \det\left(C^2 \, \mathbb{1}_{N^4}\right) \right| |\det(A)| \\ &= C^{2N^4} |\det(A)|, \end{aligned} \tag{A.39}$$

where we made use of the fact, that the Jacobi matrix is a quadratic $N^2 \times N^2$ matrix, which is true for usual Fast Fourier Transform implementations. Since A is a unitary matrix:

$$
\begin{aligned}
\left(A_{(ij)(mn)}\right)^\dagger &= A^*_{(mn)(ij)} \\
&= e^{-2\pi \frac{\sqrt{-1}}{N}(mi-nj)} \\
&= e^{-2\pi \frac{\sqrt{-1}}{N}(im-jn)} \\
&= A^{-1}_{(ij)(mn)},
\end{aligned}
\tag{A.40}
$$

the norm of the determinant $\det(A)$ is unity. Therefore, the norm of the Jacobi determinant is just a constant:

$$
|\det(\mathcal{J})| = C^{2N^4}.
\tag{A.41}
$$

With this definition we can perform the change of coordinates and express the probability distribution given in equation (7.44) in terms of the set of the matrix coefficients $\hat{\hat{S}}_{kl}$:

$$
\mathcal{P}(\{\hat{\hat{S}}_{kl}\}|\{s_i\}) = C^{2N^4}\,\mathcal{P}(\{S_{ij}\}|\{s_i\}) = C^{2N^4}\frac{\mathcal{P}(\{\hat{\hat{S}}_{kl}\})}{\mathcal{P}(\{s_i\})}\mathcal{P}(\{s_i\}|\{\hat{\hat{S}}_{kl}\}),
\tag{A.42}
$$

where $\mathcal{P}(\{\hat{\hat{S}}_{kl}\})$ is the signal covariance matrix prior in Fourier space, and $\mathcal{P}(\{s_i\}|\{\hat{\hat{S}}_{kl}\})$ is given as:

$$
\begin{aligned}
\mathcal{P}(\{s_i\}|\{\hat{\hat{S}}_{kl}\}) &= \mathcal{P}(\{s_i\}|S) \\
&= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2}\sum_i \sum_j s_i S_{ij}^{-1} s_j} \\
&= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2}\sum_i \sum_j s_i C^2 \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} e^{2\pi \frac{\sqrt{-1}}{N}(ik-jl)}\hat{\hat{S}}_{kl}^{-1} s_j} \\
&= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{C^2}{2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\sum_i s_i e^{2\pi \frac{\sqrt{-1}}{N}ik}\hat{\hat{S}}_{kl}^{-1}\sum_j s_j e^{-2\pi \frac{\sqrt{-1}}{N}jl}} \\
&= \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{C^2}{2\hat{C}^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\hat{s}_k^* \hat{\hat{S}}_{kl}^{-1}\hat{s}_l},
\end{aligned}
\tag{A.43}
$$

with $\hat{s}_l$ being the signal coefficients of the discrete Fourier transformed signal. To determine the determinant $\det(2\pi S)$ in Fourier space, we rewrite equation (A.37) as a matrix product:

$$
\begin{aligned}
S_{ij} &= \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} C\, e^{2\pi ik\frac{\sqrt{-1}}{N}}\,\hat{\hat{S}}_{kl}\, C\, e^{-2\pi jl\frac{\sqrt{-1}}{N}} \\
&= \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} CA_{ik}\,\hat{\hat{S}}_{kl}\, C\, A_{lj}^*
\end{aligned}
\tag{A.44}
$$

where we defined the unitary matrix $A_{ik} \equiv e^{2\pi ik\frac{\sqrt{-1}}{N}}$. The determinant can then be evaluated in matrix notation as:

$$
\begin{aligned}
\det(2\pi S) &= \det\left(2\pi C^2 A\hat{\hat{S}}A^{-1}\right) \\
&= \det\left(C^2\,\mathbb{1}_{N^2}A2\pi\hat{\hat{S}}A^{-1}\right) \\
&= \det\left(C^2\,\mathbb{1}_{N^2}\right)\det(A)\det\left(2\pi\hat{\hat{S}}\right)\det\left(A^{-1}\right) \\
&= C^{2N^2}\det\left(2\pi\hat{\hat{S}}\right).
\end{aligned}
\tag{A.45}
$$

A.3. Wiener Variance

In section 7.4.3 we described, that a signal sample can be obtained by adding a fluctuation term y_i to the Wiener mean reconstruction m_i. According to equations (7.38) and (7.42) we can rewrite the fluctuation term as:

$$
\begin{aligned}
y_i &= s_i^* - \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} d_l^* \\
&= s_i^* - \sum_j D_{ij}^{-1} \sum_m \sum_l K_{mj} N_{ml}^{-1} \left(\sum_k K_{lk}\, s_k^* + \epsilon_l^* \right) \\
&= \sum_j D_{ij}^{-1} \left[\sum_k D_{jk}\, s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \left(\sum_k K_{lk}\, s_k^* + \epsilon_l^* \right) \right] \\
&= \sum_j D_{ij}^{-1} \left[\sum_k \left(D_{jk} - \sum_m \sum_l K_{mj} N_{ml}^{-1} K_{lk} \right) s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \epsilon_l^* \right] \\
&= \sum_j D_{ij}^{-1} \left[\sum_k S_{jk}^{-1}\, s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \epsilon_l^* \right].
\end{aligned}
\tag{A.46}
$$

From this we immediately see that the mean of y_i:

$$
\langle y_i \rangle = \sum_j D_{ij}^{-1} \left[\sum_k S_{jk}^{-1} \langle s_k^* \rangle - \sum_m \sum_l K_{mj} N_{ml}^{-1} \langle \epsilon_l^* \rangle \right] = 0
\tag{A.47}
$$

vanishes by construction, as $\langle s_k^* \rangle = \langle \epsilon_l^* \rangle = 0$. Note, that the signal and noise covariance are given by:

$$
\langle s_i^* s_j^* \rangle = S_{ij}
\tag{A.48}
$$

and

$$
\langle \epsilon_i^* \epsilon_j^* \rangle = N_{ij}.
\tag{A.49}
$$

Also note that , by construction, there is no correlation between the two random variates s_i^* and ϵ_i^* as they have been created by two independent random processes. Therefore, we yield:

$$
\langle s_i^* \epsilon_j^* \rangle = 0.
\tag{A.50}
$$

Then we can calculate the variance as:

$$
\begin{aligned}
\langle y_i y_q \rangle &= \left\langle \sum_j D_{ij}^{-1} \left[\sum_k S_{jk}^{-1} s_k^* - \sum_m \sum_l K_{mj} N_{ml}^{-1} \epsilon_l^* \right] \sum_r D_{qr}^{-1} \left[\sum_s S_{rs}^{-1} s_s^* - \sum_t \sum_u K_{tr} N_{tu}^{-1} \epsilon_u^* \right] \right\rangle \\
&= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[\sum_{ks} S_{jk}^{-1} S_{rs}^{-1} \langle s_k^* s_s^* \rangle + \sum_{mltu} K_{mj} N_{ml}^{-1} K_{tr} N_{tu}^{-1} \langle \epsilon_l^* \epsilon_u^* \rangle \right] \\
&= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[\sum_{ks} S_{jk}^{-1} S_{rs}^{-1} S_{ks} + \sum_{mltu} K_{mj} N_{ml}^{-1} K_{tr} N_{tu}^{-1} N_{lu} \right] \\
&= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[\sum_k S_{jk}^{-1} \delta_{rk}^K + \sum_{mlt} K_{mj} N_{ml}^{-1} K_{tr} \delta_{tl}^K \right] \\
&= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} \left[S_{jr}^{-1} + \sum_{ml} K_{mj} N_{ml}^{-1} K_{lr} \right] \\
&= \sum_{jr} D_{ij}^{-1} D_{qr}^{-1} D_{jr}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_r D_{qr}^{-1} \delta_{ir}^K \\
&= D_{iq}^{-1}.
\end{aligned}
\tag{A.51}
$$

Therefore, the additive fluctuation term y_i provides the correct variance.

A.4. Hamiltonian Masses

The Hamiltonian sampler can be extremely sensitive to the choice of masses. To estimate a good guess of Hamiltonian masses we follow a similar approach as suggested in Taylor et al. (2008). According to the leapfrog scheme, given in equations (8.20), (8.21) and (8.22), a single application of the leapfrog method can be written in the form:

$$
p_i(t + \epsilon) = p_i(t) - \frac{\epsilon}{2} \left(\left. \frac{\partial \Psi(r)}{\partial r_i} \right|_{r(t)} + \left. \frac{\partial \Psi(r)}{\partial r_i} \right|_{r(t+\epsilon)} \right)
\tag{A.52}
$$

$$
r_i(t + \epsilon) = r_i(t) + \epsilon \sum_j M_{ij}^{-1} p_j(t) - \frac{\epsilon^2}{2} \sum_j M_{ij}^{-1} \left. \frac{\partial \Psi(r)}{\partial r_j} \right|_{r(t)}.
\tag{A.53}
$$

We will then approximate the forces given in equation (8.17) for $r_i << 1$:

$$
\begin{aligned}
\frac{\partial \psi(r)}{\partial r_l} &= \sum_j Q_{lj}^{-1} \left(r_j + \mu_j \right) - \left(\frac{N_l^g}{(1 + b (e^r - 1))} - R_l \bar{N} \right) b\, e^{r_l} \\
&\approx \sum_j Q_{lj}^{-1} \left(r_j + \mu_j \right) - \left[\left(N_l^g - R_l \bar{N} \right) b + \left(\left(N_l^g - R_l \bar{N} \right) b - N_l^g b^2 \right) r_l \right]. \\
&= \sum_j \left[Q_{lj}^{-1} - \left(\left(N_l^g - R_l \bar{N} \right) b - N_l^g b^2 \right) \delta_{lj}^K \right] r_j + \sum_j Q_{lj}^{-1} \mu_j - \left(N_l^g - R_l \bar{N} \right) b.
\end{aligned}
\tag{A.54}
$$

By introducing:

$$
A_{lj} = Q_{lj}^{-1} - \left(\left(N_l^g - R_l \bar{N} \right) b - N_l^g b^2 \right) \delta_{lj}^K
\tag{A.55}
$$

and

$$
D_l = \sum_j Q_{lj}^{-1} \mu_j - \left(N_l^g - R_l \bar{N} \right) b,
\tag{A.56}
$$

equation (A.54) simplifies to:

$$
\frac{\partial \psi(r)}{\partial r_l} = \sum_j A_{lj} r_j + D_l.
\tag{A.57}
$$

Introducing this approximation into equations (A.52) and (A.53) yields:

$$
\begin{aligned}
p_i(t + \epsilon) &= \sum_m \left[\delta_{im}^K - \frac{\epsilon^2}{2} \sum_j A_{ij} M_{jm}^{-1} \right] p_m(t) - \epsilon \sum_j A_{ij} \sum_p \left[\delta_{jp}^K - \frac{\epsilon^2}{4} \sum_m M_{jm}^{-1} A_{mp} \right] r_p(t) \\
&\quad - \frac{\epsilon}{2} \sum_m \left[\delta_{im}^K - \frac{\epsilon^2}{2} \sum_j A_{ij} M_{jm}^{-1} \right] D_m
\end{aligned}
\tag{A.58}
$$

and

$$r_i(t + \epsilon) = \epsilon \sum_j M_{ij}^{-1} p_j(t) + \sum_m \left(\delta_{im}^K - \frac{\epsilon^2}{2} \sum_j M_{ij}^{-1} A_{jm} \right) r_m(t) - \frac{\epsilon^2}{2} \sum_j M_{ij}^{-1} D_j \,. \tag{A.59}$$

This result can be rewritten in matrix notation as:

$$\begin{pmatrix} r(t + \epsilon) \\ p(t + \epsilon) \end{pmatrix} = T \begin{pmatrix} r(t) \\ p(t) \end{pmatrix} - \frac{\epsilon^2}{2} \begin{pmatrix} M^{-1} D \\ \epsilon \left[\mathbb{1} - \frac{\epsilon^2}{2} A M^{-1} \right] D \end{pmatrix}, \tag{A.60}$$

where the matrix T is given as:

$$T = \begin{pmatrix} \left[\mathbb{1} - \frac{\epsilon^2}{2} M^{-1} A \right] & \epsilon M^{-1} \\ -\epsilon A \left[\mathbb{1} - \frac{\epsilon^2}{4} M^{-1} A \right] & \left[\mathbb{1} - \frac{\epsilon^2}{2} A M^{-1} \right] \end{pmatrix}, \tag{A.61}$$

with $\mathbb{1}$ being the identity matrix. Successive applications of the leapfrog step yield the following propagation equation:

$$\begin{pmatrix} r^n \\ p^n \end{pmatrix} = T^n \begin{pmatrix} r^0 \\ p^0 \end{pmatrix} - \frac{\epsilon^2}{2} \left[\sum_{i=0}^{n-1} T^i \right] \begin{pmatrix} M^{-1} D \\ \epsilon \left[\mathbb{1} - \frac{\epsilon^2}{2} A M^{-1} \right] D \end{pmatrix}. \tag{A.62}$$

This equation demonstrates, that there are two criteria to be fulfilled if the method is to be stable under repeated application of the leapfrog step. First we have to ensure, that the first term of equation (A.62) does not diverge. This can be fulfilled if the eigenvalues of T have unit modulus. The eigenvalues λ are found by solving the characteristic equation:

$$det \left[\mathbb{1} \lambda^2 - 2\lambda \left(\mathbb{1} - \frac{\epsilon^2}{2} A M^{-1} \right) + \mathbb{1} \right] = 0 \,. \tag{A.63}$$

Note, that this is a similar result to what was found in Taylor et al. (2008). Our aim is to explore the parameter space rapidly, and therefore we wish to choose the largest ϵ still compatible with the stability criterion. However, any dependence of equation (A.63) also implies, that no single value of ϵ will meet the requirement for every eigenvalue to have unit modulus. For this reason we choose:

$$A = M \,. \tag{A.64}$$

We then yield the characteristic equation:

$$\left[\lambda^2 - 2\lambda \left(1 - \frac{\epsilon^2}{2} \right) + 1 \right]^N = 0 \,, \tag{A.65}$$

where N is the number of voxels. This yields the eigenvalues:

$$\lambda = \pm i \sqrt{1 - \left[1 - \frac{\epsilon^2}{2} \right]^2} + \left[1 - \frac{\epsilon^2}{2} \right], \tag{A.66}$$

which have unit modulus for $\epsilon \leq 2$. The second term in equation (A.62) involves evaluation of the geometric series $\sum_{i=0}^{n-1} T^i$. However, the geometric series for a matrix converges if and only if $|\lambda_i| < 1$ for each λ_i eigenvalue of T. This clarifies, that the nonlinearities in the Hamiltonian equations generally do not allow for arbitrary large pseudo time steps ϵ. In addition, for practical purposes we usually restrict the mass matrix to the diagonal of equation (A.55). For these two reasons, in practize, we choose the pseudo timestep ϵ as large as possible while still obtaining a reasonable rejection rate.

A.5. Gelman & Rubin diagnostic

The Gelman&Rubin diagnostic is a multichain convergence test (Gelman & Rubin 1992). It is based on analyzing multiple Markov chains by comparing intra chain variances, within each chain, and inter chain variances. A large deviation between these two variances indicates nonconvergence of the Markov chain. Let $\{\phi^k\}$, where $k = 1, ..., N$, be the collection of a single Markov chain output. The parameter ϕ^k is the kth sample of the Markov chain. Here, for notational simplicity, we will assume ϕ to be single dimensional. To test convergence with the Gelman&Rubin statistic, one has to calculate M parallel MCMC chains, which are initialized from different parts of the target distribution. After discarding the initial burn-in samples, each chain is of length n. We can then label the outputs of various chains as ϕ_m^k, with $k = 1, ..., n$ and $m = 1, ..., M$. The inter chain variance B can then be calculated as:

$$B = \frac{n}{M-1} \sum_{m=1}^{M} (\theta_m - \Omega)^2 \, , \tag{A.67}$$

where θ_m is given as:

$$\theta_m = \frac{1}{n} \sum_{k=1}^{n} \phi_m^k \, , \tag{A.68}$$

and Ω as:

$$\Omega = \frac{1}{M} \sum_{m=1}^{M} \theta_m \, . \tag{A.69}$$

Then the intra chain variance can be calculated as:

$$W = \frac{1}{M} \sum_{m=1}^{M} \Gamma_m^2 \, , \tag{A.70}$$

with :

$$\Gamma_m^2 = \frac{1}{n-1} \sum_{k=1}^{n} (\phi_m^k - \theta_m)^2 \, . \tag{A.71}$$

With the above definition the marginal posterior variance can be estimated via:

$$V = \frac{n-1}{n} W + \frac{M+1}{nM} B \, . \tag{A.72}$$

If all M chains have reached the target distribution, this posterior variance estimate should be very close to the intra chain variance W. For this reason, one expects the ratio V/W to be close to 1. The square root of this ratio is reffered to as the potential scale reduction factor (PSRF):

$$PSRF = \sqrt{\frac{V}{W}} \, . \tag{A.73}$$

If the PSRF is close to one, one can conclude that each chain has stabilized, and has reached the target distribution (Gelman & Rubin 1992).

Bibliography

Abazajian K. N., et al., 2009, ApJS, 182, 543

Albrecht A., Steinhardt P. J., 1982, Physical Review Letters, 48, 1220

Alcock C., Paczyński B., 1979, Nature, 281, 358

Andrieu C., de Freitas N., Doucet A., Jordan M. I., 2003, Machine Learning, 50, 5

Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, A&A, 474, 315

Atwood W. B., et al., 2009, ApJ, 697, 1071

Ballinger W. E., Heavens A. F., Taylor A. N., 1995, MNRAS, 276, L59

Ballinger W. E., Peacock J. A., Heavens A. F., 1996, MNRAS, 282, 877

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

Bardeen J. M., Steinhardt P. J., Turner M. S., 1983, Phys. Rev. D, 28, 679

Bartelmann M., Schneider P., 2001, Physics Reports, 340, 291

Beichl I., Sullivan F., 2000, Computing in Science and Engg., 2, 65

Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Physics Reports, 367, 1

Bernardi M., Nichol R. C., Sheth R. K., Miller C. J., Brinkmann J., 2006, AJ, 131, 1288

Bertschinger E., 1987, ApJL, 323, L103

Bertschinger E., Dekel A., 1989, ApJL, 336, L5

Bertschinger E., Dekel A., 1991, in Latham D. W., da Costa L. A. N., eds, ASP Conf. Ser. 15: Large-Scale Structures and Peculiar Motions in the Universe Mapping Large-Scale Flows in Three Dimensions: Method

Bistolas V., Hoffman Y., 1998, ApJ, 492, 439

Blake C., Glazebrook K., 2003, ApJ, 594, 665

Blanton M. R., Brinkmann J., Csabai I., Doi M., Eisenstein D., Fukugita M., Gunn J. E., Hogg D. W., Schlegel D. J., 2003, AJ, 125, 2348

Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, ApJ, 629, 143

Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, AJ, 125, 2276

Blanton M. R., Roweis S., 2007, AJ, 133, 734

Boerner G., Deng Z., Xia X., Zhou Y., 1991, A&A Suppl., 180, 47

Brandenberger R., Kahn R., 1984, Phys. Rev. D, 29, 2172

Brandenberger R., Kahn R., Press W. H., 1983, Phys. Rev. D, 28, 1809

Buchert T., 1996, in S. Bonometto, J. R. Primack, & A. Provenzale ed., Dark Matter in the Universe Lagrangian Perturbation Approach to the Formation of Large-scale Structure. pp 543–564

Buchert T., 2000, Shanghai Observatory Annals, 21, 85

Buchert T., Domínguez A., 2005, A&A, 438, 443

Cabré A., Gaztañaga E., 2009, MNRAS, 396, 1119

Carroll S. M., 2004, Spacetime and geometry. An introduction to general relativity. San Francisco, CA, USA: Addison Wesley

Carroll S. M., Press W. H., Turner E. L., 1992, ARA&A, 30, 499

Chib S., Greenberg E., 1995, The American Statistician, 49, 327

Choi Y., Park C., Vogeley M. S., 2007, ApJ, 658, 884

Colberg J. M., et al., 2008, MNRAS, 387, 933

Colberg J. M., Sheth R. K., Diaferio A., Gao L., Yoshida N., 2005, MNRAS, 360, 216

Cole S., et al., 2005, MNRAS, 362, 505

Coles P., Jones B., 1991, MNRAS, 248, 1

Colless M., et al., 2001, MNRAS, 328, 1039

Colombi S., 1994, ApJ, 435, 536

Cowles M. K., Carlin B. P., 1996, Journal of the American Statistical Association, 91, 883

Crocce M., Scoccimarro R., 2006, Phys. Rev. D, 73, 063519

Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., Navarro J. F., Yoshida N., 2006, MNRAS, 365, 11

Cui W., Liu L., Yang X., Wang Y., Feng L., Springel V., 2008, ApJ, 687, 738

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

De Lucia G., Blaizot J., 2007, MNRAS, 375, 2

Dekel A., Lahav O., 1999, ApJ, 520, 24

Del Popolo A., 2007, Astronomy Reports, 51, 169

d'Inverno R., 1992, Introducing Einstein's Relativity. Oxford University Press, Oxford 1992

D'Mellow K. J., Taylor A. N., 2000, in S. Courteau & J. Willick ed., Cosmic Flows Workshop Vol. 201 of Astronomical Society of the Pacific Conference Series, Generalising the Inverse Redshift-Space Operator: Vorticity in Redshift-Space. p. 298

Doran M., Wetterich C., 2003, Nuclear Physics B Proceedings Supplements, 124, 57

Dressler A., 1980, ApJ, 236, 351

Duan Q., Angelini E., Song T., Laine A., 17-21 Sept. 2003, Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE, 2, 1192

Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Physics Letters B, 195, 216

152

Dunkley J., Bucher M., Ferreira P. G., Moodley K., Skordis C., 2005, MNRAS, 356, 925

Ebeling H., Wiedenmann G., 1993, PRE, 47, 704

Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1P

Efstathiou G., et al., 2002, MNRAS, 330, L29

Einasto J., 2009, arXiv:0901.0632v1 [astro-ph.CO]

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 778-786., pp 778–786

Eisenstein D. J., Annis J., et al., 2001, AJ, 122, 2267

Eisenstein D. J., Hu W., 1998, ApJ, 496, 605

Eisenstein D. J., Hu W., 1999, ApJ, 511, 5

Enßlin T. A., Frommert M., Kitaura F. S., 2009, Phys. Rev. D, 80

Erdoğdu P., et al., 2004, MNRAS, 352, 939

Erdoğdu P., Lahav O., Huchra J., et al. 2006, MNRAS, 373, 45

Erdoğdu P., Lahav O., Zaroubi S., et al. 2004, MNRAS, 352, 939

Eriksen H. K., Huey G., Banday A. J., Górski K. M., Jewell J. B., O'Dwyer I. J., Wandelt B. D., 2007, ApJL, 665, L1

Eriksen H. K., O'Dwyer I. J., Jewell J. B., Wandelt B. D., Larson D. L., Górski K. M., Levin S., Banday A. J., Lilje P. B., 2004, ApJS, 155, 227

Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23

Ferguson R. S., 2001, Practical algorithms for 3D computer graphics. A. K. Peters, Ltd., Natick, MA, USA

Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995, MNRAS, 272, 885

Fisher K. B., Scharf C. A., Lahav O., 1994, MNRAS, 266, 219

Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, MNRAS, 396, 1815

Freedman W. L., Madore B. F., Gibson B. K., Ferrarese L., Kelson D. D., Sakai S., Mould J. R., Kennicutt R. C., Ford H. C., Graham J. A., Huchra J. P., Hughes S. M. G., Illingworth G. D., Macri L. M., Stetson P. B., 2001, ApJ, 553, 47

Friedmann A., 1922, Z. Phys., 10, 377

Friedmann A., 1924, Z. Phys., 32, 326

Frieman J. A., et al., 2008, AJ, 135, 338

Frigo M., Johnson S. G., 2005, Proceedings of the IEEE, 93, 216

Frommert M., Enßlin T. A., Kitaura F. S., 2008, MNRAS, 391, 1315

Galassi M., Davies J., Theiler J., Gough B., Jungman G., Booth M., Rossi F., 2003, Gnu Scientific Library: Reference Manual. Network Theory Ltd.

Gamerman D., Lopes H. F., 2006, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall/CRC

Ganon G., Hoffman Y., 1993, ApJL, 415, L5

Gaztanaga E., Yokoyama J., 1993, ApJ, 403, 450

Gelfand A. E., Smith A. F. M., 1990, 85, 398

Gelman A., 2008, Bayesian Analysis, 3, 445

Gelman A., Rubin D., 1992, Statistical Science, 7, 457

Geman S., Geman D., 1984, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721

Geweke J., 1992, in IN BAYESIAN STATISTICS Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. University Press, pp 169–193

Gibbons G. W., Hawking S. W., 1977, Phys. Rev. D, 15, 2752

Gómez P. L., et al., 2003, ApJ, 584, 210

Goss M., Wu K., 2000, Tech.rep., HP Labs, 12

Goto T., Yamauchi C., Fujita Y., Okamura S., Sekiguchi M., Smail I., Bernardi M., Gomez P. L., 2003, MNRAS, 346, 601

Gunn J. E., et al., 1998, AJ, 116, 3040

Gunn J. E., et al., 2006, AJ, 131, 2332

Guth A. H., 1981, Phys. Rev. D, 23, 347

Guth A. H., Pi S.-Y., 1982, Physical Review Letters, 49, 1110

Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489

Hajian A., 2007, Phys. Rev. D, 75, 083525

Hamann J., Hannestad S., Melchiorri A., Wong Y. Y. Y., 2008, Journal of Cosmology and Astro-Particle Physics, 7, 17

Hamilton A. J. S., 1997a, MNRAS, 289, 285

Hamilton A. J. S., 1997b, MNRAS, 289, 295

Hamilton A. J. S., 1998, in Hamilton D., ed., The Evolving Universe Vol. 231 of Astrophysics and Space Science Library, Linear Redshift Distortions: a Review

Hannestad S., 2003, Journal of Cosmology and Astro-Particle Physics, 5, 4

Hanson K. M., 2001, in Sonka M., Hanson K. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4322 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Markov chain Monte Carlo posterior sampling with the Hamiltonian method. pp 456–467

Harrison E. R., 1970, Phys. Rev. D, 1, 2726

Hastings W. K., 1970, Biometrika, 57, 97

Hawking S. W., 1982, Physics Letters B, 115, 295

Heavens A. F., Taylor A. N., 1995, MNRAS, 275, 483

Heidelberger P., Welch P. D., 1981, Commun. ACM, 24, 233

Ho Y.-C., Pepyne D. L., 2002, Cybernetics and Sys. Anal., 38, 292

Hockney R. W., Eastwood J. W., 1988, Computer simulation using particles. Taylor & Francis, Inc., Bristol, PA, USA

Hoffman Y., 1994, in Balkowski C., Kraan-Korteweg R. C., eds, Unveiling Large-Scale Structures Behind the Milky Way Vol. 67 of Astronomical Society of the Pacific Conference Series, Wiener Reconstruction of the Large-Scale Structure in the Zone of Avoidance

Hoffman Y., Ribak E., 1991, ApJL, 380, L5

Hogg D. W., 1999, arXiv:astro-ph/9905116v4

Hu W., Eisenstein D., Tegmark M., 1998, Phys. Rev. Lett., 80, 5255

Hubble E., 1934, ApJ, 79, 8

Jasche J., Kitaura F. S., 2009, arXiv:0911.2496v1 [astro-ph.CO]

Jasche J., Kitaura F. S., Wandelt B. D., Enßlin T. A., 2009, arXiv:0911.2493v1 [astro-ph.CO]

Jewell J., Levin S., Anderson C. H., 2004, ApJ, 609, 1

Jewell J. B., Eriksen H. K., Wandelt B. D., O'Dwyer I. J., Huey G., Górski K. M., 2009, ApJ, 697, 258

Jing Y. P., 2005, ApJ, 620, 559

Jing Y. P., Börner G., 2004, ApJ, 617, 782

Jing Y. P., Mo H. J., Boerner G., 1998, ApJ, 494, 1

Kaiser N., 1987, MNRAS, 227, 1

Kang X., Jing Y. P., Mo H. J., Börner G., 2002, MNRAS, 336, 892

Kayo I., Taruya A., Suto Y., 2001, Astrophys. J., 561, 22

Kemeny J., Snell J., 1960, Finite Markov Chains. Van Nostrand, Princeton, N.J., U.S.A.

Kendall M. G., Stuart A., 1969, Advanced Theory of Statistics. Griffin, London, UK

Kibble T. W. B., 1976, Journal of Physics A Mathematical General, 9, 1387

Kitaura F. S., Enßlin T. A., 2008, MNRAS, 389, 497

Kitaura F. S., Jasche J., Li C., Enßlin T. A., Metcalf R. B., Wandelt B. D., Lemson G., White S. D. M., 2009, MNRAS, 400, 183

Kitaura F. S., Jasche J., Metcalf R. B., 2009, arXiv:0911.1407v1 [astro-ph.CO]

Klypin A., Hoffman Y., Kravtsov A. V., Gottlöber S., 2003, ApJ, 596, 19

Kolb E. W., Turner M. S., 1990, Frontiers in Physics, 69

Komatsu E., et al., 2009, ApJS, 180, 330

Kuehn F., Ryden B. S., 2005, ApJ, 634, 1032

Lahav O., 1994, in Balkowski C., Kraan-Korteweg R. C., eds, ASP Conf. Ser. 67: Unveiling Large-Scale Structures Behind the Milky Way Wiener Reconstruction of All-Sky Spherical Harmonic Maps of the Large-Scale Structure

Lahav O., Fisher K. B., Hoffman Y., Scharf C. A., Zaroubi S., 1994, ApJL, 423

Lahav O., Rees M. J., Lilje P. B., Primack J. R., 1991, MNRAS, 251, 128

Lahav O., Suto Y., 2004, Living Reviews in Relativity, 7, 8

Landau L. D., Lifshitz E. M., 1975, The classical theory of fields. Course of theoretical physics - Pergamon International Library of Science, Technology, Engineering and Social Studies, Oxford: Pergamon Press, 1975, 4th rev.engl.ed.

Lang C. B., Pucker N., 1998, Mathematische Methoden in der Physik. Spektrum, Akad. Verl., Berlin

Larson D. L., Eriksen H. K., Wandelt B. D., Górski K. M., Huey G., Jewell J. B., O'Dwyer I. J., 2007, ApJ, 656, 653

Layzer D., 1956, AJ, 61, 383

Lee J., Erdogdu P., 2007, ApJ, 671, 1248

Lee J., Lee B., 2008, ApJ, 688, 78

Lee J., Li C., 2008, arXiv:0803.1759v1 [astro-ph]

Lemson G., Kauffmann G., 1999, MNRAS, 302, 111

Lewis I., et al., 2002, MNRAS, 334, 673

Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, MNRAS, 368, 21

Libeskind N. I., Yepes G., Knebe A., Gottlöber S., Hoffman Y., Knollmann S. R., 2010, MNRAS, 401, 1889

Linde A. D., 1982, Physics Letters B, 108, 389

Magira H., Jing Y. P., Suto Y., 2000, ApJ, 528, 30

Marschner S. R., Lobb R. J., 1994, in VIS '94: Proceedings of the conference on Visualization '94 An evaluation of reconstruction filters for volume rendering. IEEE Computer Society Press, Los Alamitos, CA, USA, pp 100–107

Martel H., 2005, arXiv:astro-ph/0506540v4

Martínez V. J., Saar E., 2002, Statistics of the Galaxy Distribution. Chapman & Hall/CRC

Martinez-Vaquero L. A., Yepes G., Hoffman Y., Gottlöber S., Sivan M., 2009, MNRAS, 397, 2070

Matsubara T., Suto Y., 1996, ApJL, 470, L1

Matsumoto M., Nishimura T., 1998, ACM Trans. Model. Comput. Simul., 8, 3

Meneux B., et al., 2008, A&A, 478, 299

Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E., 1953, J. Chem. Phys., 21, 1087

Michel R., Kirchhoff K., 1999, Nachweis-, Erkennungs- und Vertrauensgrenzen bei Kernstrahlungsmessungen. TÜV Media GmbH TÜV Rheinland Group

Misner C. W., Thorne K. S., Wheeler J. A., 1973, Gravitation. W.H. Freeman and Co., San Francisco, 1973

Mitchell D. P., Netravali A. N., 1988, SIGGRAPH Comput. Graph., 22, 221

Mukhanov V., 2005, Physical Foundations of Cosmology. Cambridge University Press, Cambridge, U.K.

Neal R. M., 1993, Technical Report CRG-TR-93-1, Probabilistic Inference using Markov Chain Monte Carlo Methods. Dept. of Computer Science, University of Toronto

Neal R. M., 1996, Bayesian Learning for Neural Networks (Lecture Notes in Statistics), 1 edn. Springer

Norberg P., et al., 2002, MNRAS, 332, 827

Novikov D., Colombi S., Doré O., 2006, MNRAS, 366, 1201

O'Dwyer I. J., Eriksen H. K., Wandelt B. D., Jewell J. B., Larson D. L., Górski K. M., Banday A. J., Levin S., Lilje P. B., 2004, ApJL, 617, L99

O'Hagan A., 2000, Kendall's advanced theory of statistics. Vol. 2B, Arnold, London, U.K.

Padmanabhan T., 1993, Structure Formation in the Universe. Cambridge University Press, Cambridge, U.K.; New York, U.S.A.

Park C., Choi Y., Vogeley M. S., Gott J. R. I., Blanton M. R., 2007, ApJ, 658, 898

Peacock J. A., 1999, Cosmological Physics. Cambridge University Press, Cambridge, U.K.

Peacock J. A., Dodds S. J., 1994, MNRAS, 267, 1020

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Peebles P., 1980, Large Scale Structure of the Universe. Princeton University Press, Princeton, U.S.A.

Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815

Pen U.-L., 1998, ApJ, 504, 601

Percival W. J., 2005, MNRAS, 356, 1168

Percival W. J., et al., 2001, MNRAS, 327, 1297

Percival W. J., et al., 2002, MNRAS, 337, 1068

Percival W. J., Verde L., Peacock J. A., 2004, MNRAS, 347, 645

Percival W. J., White M., 2009, MNRAS, 393, 297

Pope A. C., et al., 2004, ApJ, 607, 655

Popowski P. A., Weinberg D. H., Ryden B. S., Osmer P. S., 1998, ApJ, 498, 11

Postman M., Geller M. J., 1984, ApJ, 281, 95

Pueblas S., Scoccimarro R., 2009, Phys. Rev. D, 80, 043504

Raftery A. E., Lewis S. M., 1995, in In Practical Markov Chain Monte Carlo (W.R. Gilks, D.J. Spiegelhalter and The number of iterations, convergence diagnostics and generic metropolis algorithms. Chapman and Hall, pp 115–130

Robert C. P., Casella G., 2005, Monte Carlo Statistical Methods (Springer Texts in Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA

Rojas R. R., Vogeley M. S., Hoyle F., Brinkmann J., 2005, ApJ, 624, 571

Rubin V. C., Ford Jr. W. K., 1970, ApJ, 159, 379

Sako M., et al., 2008, AJ, 135, 348

Sánchez A. G., Cole S., 2008, MNRAS, 385, 830

Saslaw W. C., 1985, Gravitational physics of stellar and galactic systems. Cambridge University Press, Cambridge, U.K.

Saunders W., Ballinger W. E., 2000, in Kraan-Korteweg R. C., Henning P. A., Andernach H., eds, Mapping the Hidden Universe: The Universe behind the Mily Way - The Universe in HI Vol. 218 of Astronomical Society of the Pacific Conference Series, Interpolation of Discretely-Sampled Density Fields

Saunders W., et al., 2000, in R. C. Kraan-Korteweg, P. A. Henning, & H. Andernach ed., Mapping the Hidden Universe: The Universe behind the Mily Way - The Universe in HI Vol. 218 of Astronomical Society of the Pacific Conference Series, The IRAS View of the Local Universe. p. 141

Schäfer B. M., 2009, International Journal of Modern Physics D, 18, 173

Schmoldt I. M., Saar V., Saha P., Branchini E., Efstathiou G. P., Frenk C. S., Keeble O., Maddox S., McMahon R., Oliver S., Rowan-Robinson M., Saunders W., Sutherland W. J., Tadros H., White S. D. M., 1999, ApJ, 118, 1146

Scoccimarro R., 2004, Phys. Rev. D, 70, 083007

Seljak U., 1998, ApJ, 503, 492

Seljak U., 2000, MNRAS, 318, 203

Seo H.-J., Eisenstein D. J., 2003, ApJ, 598, 720

Shannon C. E., 1948, Bell System Technical Journal, 27, 379

Shannon C. E., 1949, Proc,Inst.Radio.Eng, 37, 10

Sheth R. K., 1995, MNRAS, 277, 933

Short C. J., Coles P., 2006, Journal of Cosmology and Astro-Particle Physics, 12, 12

Silk J., 1968, ApJ, 151, 459

Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A., Efstathiou G., Couchman H. M. P., 2003, MNRAS, 341, 1311

Smith S. W., 2002, Digital Signal Processing: A Practical Guide for Engineers and Scientists. Newnes, Burlington, MA, USA

Spergel D., Verde L., Peiris H., Komatsu E., Nolta M., Bennett C., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S., Page L., Tucker G., Weiland J., Wollack E., Wright E., 2003, Astrophys. J. Suppl. Ser., 148, 175

Spergel D. N., et al., 2007, ApJS, 170, 377

Spergel D. N., Verde L., Peiris H. V., Komatsu E., Nolta M. R., Bennett C. L., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S. S., Page L., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2003, ApJS, 148, 175

Springel V., Hernquist L., 2002, MNRAS, 333, 649

Springel V., Yoshida N., White S. D. M., 2001, New Astronomy, 6, 79

Starobinsky A. A., 1982, Physics Letters B, 117, 175

Stuart A., Ord J. K., 1994, Kendall's advanced theory of statistics. Vol.1: Distribution theory. Hodder Arnold, London, UK

Sugiyama N., 1995, ApJS, 100, 281

Sunyaev R. A., Zeldovich Y. B., 1970, A&SS, 7, 3

Tadros H., et al., 1999, MNRAS, 305, 527

Tanner M. A., 1996, Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, 3rd edn. Springer Series in Statistics, Springer Verlag

Taylor A., Valentine H., 1999, MNRAS, 306, 491

Taylor J. F., Ashdown M. A. J., Hobson M. P., 2008, MNRAS, 389, 1284

Tegmark M., 1995, ApJ, 455, 429

Tegmark M., Bromley B. C., 1999, ApJL, 518, L69

Tegmark M., et al. 2004, ApJ, 606, 702

Tegmark M., et al., 2004, Phys. Rev. D, 69

Tegmark M., et al., 2006, Phys. Rev. D, 74, 123507

Tegmark M., Taylor A. N., Heavens A. F., 1997, ApJ, 480, 22

Theußl T., Hauser H., Gröller E., 2000, in VVS '00: Proceedings of the 2000 IEEE symposium on Volume visualization Mastering windows: improving reconstruction. ACM, New York, NY, USA, pp 101–108

Trotta R., 2008, Contemporary Physics, 49, 71

van de Weygaert R., Schaap W., 2001, in Banday A. J., Zaroubi S., Bartelmann M., eds, Mining the Sky Tessellation Reconstruction Techniques

Verde L., et al., 2003, ApJS, 148, 195

Wandelt B. D., 2004, arXiv:astro-ph/0401623v1

Wandelt B. D., Larson D. L., Lakshminarayanan A., 2004, Phys. Rev. D, 70, 083511

Webster M., Lahav O., Fisher K., 1997, MNRAS, 287, 425

Wetterich C., 1995, A&A, 301, 321

Whitmore B. C., Gilmore D. M., Jones C., 1993, ApJ, 407, 489

Wolberg G., 1997, in , The Computer Science and Engineering Handbook. pp 1270–1299

Wolpert D. H., Macready W. G., 1997, IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, 1, 67

Wolpert R. L., 2004, Statistical Science, 19, 205

Xia X. Y., Deng Z. G., Zhou Y. Y., 1987, in A. Hewitt, G. Burbidge, & L. Z. Fang ed., Observational Cosmology Vol. 124 of IAU Symposium, Large-Scale Distribution of Galaxies with Different Luminosities /. pp 363–+

Yamamoto K., 2003, ApJ, 595, 577

Yanny B., et al., 2009, AJ, 137, 4377

York D. G., et al., 2000, AJ, 120, 1579

Yoshida N., Springel V., White S. D. M., Tormen G., 2000, ApJL, 544, L87

Zaninetti L., 1995, A&A Suppl. Ser., 109, 71

Zaroubi S., 2002, MNRAS, 331, 901

Zaroubi S., Hoffman Y., Dekel A., 1999, ApJ, 520, 413

Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, ApJ, 449, 446

Zel'Dovich Y. B., 1970, A&A, 5, 84

Zeldovich Y. B., 1972, MNRAS, 160, 1P

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

Acknowledgements

Something that has always puzzled me all my life is why, when I am in special need of help, the good deed is usually done by somebody on whom I have no claim.

William Feather

It is a pleasure to thank the many people who have made this thesis possible. Foremost, I wish to thank my advisor Prof. Dr. Simon D. M. White for giving me the opportunity to study cosmology at the Max Planck Institute for Astrophysics. Simon is one of the scientists I admire most. His comments, scientific criticism and judgment have always been driving impulses for my work and eventually lead to the final results of this thesis. I further thank Torsten A. Enßlin, my PhD supervisor, for giving me the opportunity to work on this project, although the initial task of my PhD thesis focussed on a completely different topic. Torsten showed me how to attack difficult mathematical problems by breaking them down and understanding their basic behaviors. This approach helped me many times in the course of my PhD, when I got lost in neverending analytic calculations. I also enjoyed the many scientific discussions with him, often leading to new interesting approaches and projects. Torsten and I share the same interest in cosmic ray physics, and I am very thankful for his strong support and encouragement to persue this research interest aside from my main PhD task. His encouragement and trust in my abilities enabled me to follow my own independent research route and to develop my individual scientific style. I am looking forward to reap the fruit of our labour. Also, I am very thankful to Benedetta Ciardi, my former diploma thesis supervisor, who taught me the art of good scientific working. I am grateful to Benjamin D. Wandelt for many illuminating remarks and commentaries about the general and specific behavior of Gibbs sampling algorithms. His advises guided me to the successful completion of this project. I would also like to thank Benton R. Metcalf and Andreas Faltenbacher for many useful discussions and suggestions. It is a special honor for me to thank Francisco-Shu Kitaura Joyanes. Francisco introduced me to the world of Bayesian statistics and its possibilities. His enormous enthusiasm and creativity flashed over to me very quickly and effectively influenced my scientific approaches. He has been a major guide throughout my PhD thesis and I am happy to say that we have become good friends. Further, I want to thank Björn Malte Schäfer. Björn left MPA when I just arrived and he passed his Mensa-card on to me, which actually gave me one free lunch before I had to recharge. Throughout my entire thesis time, Björn supported me with encouraging conversations, interesting project proposals and many numerical tools. Particular thanks go to Cheng Li. He has made available his support in a number of ways, by providing me with the required data products and by being available with his enormous expertise. I would also like to thank Rainer Moll, my roommate, for interesting scientific discussions and exhilarant conversation. He has constantly supported me with many valuable numerical gadgets as well as useful advises. It is a pleasure to thank Immanuel Maurer for many funny lunch conversations and the playful suggestion to misuse Bayesian statistics in order to take online gambling casinos to the cleaner's. I am still thinking about this... seriously! Further, I would like to thank Klaus Dolag, who was always ready to listen to me. Now I am looking forward to starting our common projects. I also thank the many students at MPA, in particular, André Waelkens, Mona Frommert, Cornelius Weig, Henrik Junklewitz, Petr Kuchar and Herbert Kaiser for many interesting and funny conversations.

Special thanks go to my fiancée María Ángeles for her enduring encouragement, support and love. During all this time you have been very patient with me although you had to forgo my company much too often. I thank you for persistently reminding me that there is also a life beyond work. Without you the last years would have been much less joyful. Thanks also to my sisters Janina and Laura for their encouraging support and for being a constant source of joy in my life. Finally, I am very grateful to my parents for their unconditional support and trust throughout my entire life. I thank my father, who aroused my technical and scientific interests by continuously showing me that every problem has a solution as soon as the underlying principle is understood. Nevertheless, I owe my deepest gratitude to my mother, for her patience and strength throughout all these years. With her love and trust she cleared many obstacles and opened up new possibilities for me. Without my parents this thesis would not have been possible.

Publications

published in refereed journals

1. *Cosmic rays and the primordial gas*:
   J. Jasche, B. Ciardi, T. A. Enßlin, 2007
   MNRAS 380,417 (arXiv:0705.4541v1 [astro-ph])

submitted

1. *Digital Signal Processing in Cosmology*:
   J. Jasche, F. S. Kitaura, T. A. Enßlin:
   arXiv:0901.3043v1 [astro-ph.IM], submitted to MNRAS, 20.January.2009

2. *Bayesian power-spectrum inference for Large Scale Structure data*:
   J. Jasche, F. S. Kitaura, B. D. Wandelt, T. A. Enßlin:
   arXiv:0911.2493v1 [astro-ph.CO], submitted to MNRAS, 22.May.2009

3. *Fast Hamiltonian sampling for large scale structure inference*:
   J.Jasche, F. S. Kitaura:
   arXiv:0911.2496v1 [astro-ph.CO], submitted to MNRAS, 21.September.2009

4. *Bayesian non-linear large scale structure inference of the Sloan Digital Sky Survey data release 7*:
   J.Jasche, F. S. Kitaura, C. Li , T. A. Enßlin:
   arXiv:0911.2498v1 [astro-ph.CO], submitted to MNRAS, 12.November.2009

Curriculum Vitae

Personal data

Name: Jens Jasche
Address: Max-Planck-Institut für Astrophysik,
Karl-Schwarzschild-Straße 1,
85741 Garching, Germany
E-mail: jjasche@mpa-garching.mpg.de
Birth: 27.Mar.1981 in Essen, Germany
Nationality: German citizenship

Education

1987 - 1988 Katholische Grundschule Anholt, Germany
1988 - 1989 Grundschule Langenberg, Germany
1989 - 1991 Grundschule Herrenhauser Straße Mettmann, Germany
1991 - 2000 Konrad-Heresbach-Gymnasium Mettmann, Germany
June, 2000 Abitur at Konrad-Heresbach-Gymnasium Mettmann, Germany
2000 - 2001 Military service at the German army headquarters in Koblenz, Germany

Academic record

2001 - 2006 Study of physics at Leibniz-Universität Hannover, Germany
September, 2003 Vordiplom in physics awarded by Leibniz-Universität Hannover, Germany
2005 - 2006 Diploma student at Max-Planck-Institut für Astrophysik, Garching, Germany
October, 2006 Diplom in physics awarded by Leibniz-Universität Hannover, Germany
thesis: "On the coupling between cosmic rays and primordial gas"
supervisors: Prof. Dr. S.D.M. White and Prof. Dr. O. Lechtenfeld
2006 - 2010 PhD student at Max-Planck-Institut für Astrophysik, Garching, Germany