

**OPTIMISATION OF THE POPULATION  
MONTE CARLO ALGORITHM:  
APPLICATION TO COSMOLOGY**

Darell Moodley

University of KwaZulu-Natal

**OPTIMISATION OF THE POPULATION MONTE CARLO  
ALGORITHM: APPLICATION TO COSMOLOGY**

Darell Moodley

Submitted in fulfillment of the academic requirements for the degree of Doctor of Philosophy  
(Science) in the School of Mathematics, Statistics and Computer Science, University of  
KwaZulu-Natal.

As the candidate's supervisor I have approved this dissertation for submission.

Signed:

Name:

Date:

---

## ABSTRACT

---

In this thesis we study the Population Monte Carlo (PMC) algorithm and utilise simulations to improve the efficiency of the algorithm by optimising the algorithm parameters. We then apply these optimisation results to a cosmological parameter estimation problem, specifically that of determining the initial conditions for structure formation. We accomplish this by using cosmic microwave background (CMB) data to constrain models with an admixture of adiabatic and isocurvature modes.

We review the standard cosmological model and current cosmological probes used for cosmology and discuss the CMB anisotropy spectrum, which forms the basis for our cosmological parameter estimation study. We briefly outline linear perturbation theory and initial conditions that form the basis of the inflationary models considered in this thesis. We describe the adiabatic and isocurvature perturbations and investigate their effect on the CMB anisotropy spectrum.

We outline the Bayesian parameter estimation methodology adopted in our study and review Monte Carlo sampling, specifically the Markov Chain Monte Carlo (MCMC) and PMC algorithms explaining why these methods are used in Bayesian parameter estimation. We discuss recent developments to the PMC and MCMC algorithms and discuss various applications of these algorithms in cosmology.

We focus on optimising the performance of the PMC algorithm with respect to its algorithm parameters that are specified initially. However, we first define a measure of efficiency, related to the computational cost of the sampling algorithm and then use simulations to maximise this measure with respect to the algorithm parameters. These algorithm parameters include the sample size drawn at each iteration, the number of importance functions used, and the parameters that characterise the importance functions. Before this though, we will first investigate the optimisation of the PMC algorithm for a multivariate Gaussian target distribution, and present results for choosing the optimal algorithm parameters that maximise efficiency. We will also explore the performance of PMC on more complex distributions such as the banana shaped, bimodal and hypercube distribution, and discuss the advantages and shortfalls for these distributions.

We incorporate the results from the previous optimisation study by applying the PMC algorithm to a cosmological parameter estimation problem. We constrain models with an admixture of adiabatic and isocurvature perturbations using the nine-year data release from the Wilkinson Microwave Anisotropy Probe (WMAP) experiment. We discuss challenges faced in sampling such complex distributions, the modifications to the PMC sampler needed to achieve convergence, and the efficiencies achieved in sampling these distributions. We present results on the marginal and joint parameter distributions for all possible admixtures of adiabatic and isocurvature modes. We then perform a principal component analysis to determine the degeneracies that arise from the introduction of isocurvature modes. In comparison to similar studies undertaken with the WMAP one-year and three-year datasets, we find that the allowed isocurvature fraction is more tightly constrained than in previous studies.

---

## PREFACE

---

The study described in this thesis was carried out in the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal during the period June 2010 to July 2014. This dissertation was completed under the supervision of Prof. K. Moodley.

This study represents original work by the author and has not been submitted in any form for any degree or diploma to another tertiary institution. Where use was made of the work of others it has been duly acknowledged in the text.

Signed:

Name:

Date:

---

## DECLARATION 1 - PLAGIARISM

---

I, Darell Moodley, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
  - Their words have been re-written but the general information attributed to them has been referenced.
  - Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed:

Name:

Date:

---

## ACKNOWLEDGMENTS

---

I am extremely grateful to Prof. Kavilan Moodley for assisting me in completing this thesis. A value cannot be placed upon his guidance and expertise rendered to me. There were many difficult trials during the course of this research, and I thank him for keeping me focused and staying the course.

I also want to thank Dr. Simon Muya Kasanda and Dr. Moumita Aich for their ample support and insightful advice. Their constant willingness to assist me will not be forgotten.

To Kenda, Susan and Jethro, thank you for making me smile during difficult times. It was a great benefit sharing our pain experienced during our research, which reminded me I was not alone. I consider myself very fortunate to have friends like you.

Members of the Astrophysics and Cosmology Research Unit (ACRU) of the University of KwaZulu-Natal readily assisted with the smooth running of this project.

Many thanks must be extended to the Centre for High Performance Computing (CHPC), the South African Square Kilometer Array (SKA) project and the National Research Foundation (NRF) for their financial support.

---

# Contents

---

Abstract . . . . .	ii
Preface . . . . .	iv
Plagiarism Declaration . . . . .	v
Acknowledgments . . . . .	vi
List of Figures . . . . .	xiii
List of Tables . . . . .	xiv
<b>1 Introduction</b>	<b>1</b>
<b>2 The standard cosmological model</b>	<b>7</b>
2.1 The expansion of the universe . . . . .	8
2.2 The cosmic microwave background . . . . .	13
2.3 Large-scale structure . . . . .	14
2.4 Type Ia supernovae . . . . .	18
2.5 Baryon acoustic oscillations . . . . .	18
2.6 Background and perturbed Friedmann-Robertson-Walker models . . . . .	19

---

2.7	Einstein equations and energy momentum conservation . . . . .	22
2.8	Initial conditions . . . . .	26
2.9	CMB anisotropies . . . . .	28
2.9.1	Angular power spectrum $C_\ell$ . . . . .	29
2.10	Effect of cosmological parameters on the CMB temperature spectrum . . . . .	31
2.10.1	Baryon density $\Omega_b h^2$ . . . . .	31
2.10.2	Cold dark matter density $\Omega_c h^2$ . . . . .	31
2.10.3	Cosmological constant density $\Omega_\Lambda$ . . . . .	33
2.10.4	Optical depth $\tau$ . . . . .	34
2.10.5	The spectral index $n_s$ . . . . .	35
2.10.6	The scalar amplitude $A_s$ . . . . .	36
2.11	CMB dependence on initial conditions . . . . .	37
2.12	Isocurvature mode parametrisation . . . . .	40
<b>3</b>	<b>Monte Carlo sampling methods</b>	<b>44</b>
3.1	The general problem . . . . .	44
3.2	The Markov Chain Monte Carlo algorithm . . . . .	45
3.2.1	Markov Chains . . . . .	47
3.2.2	The Metropolis-Hastings Algorithm . . . . .	49
3.3	Population Monte Carlo algorithm . . . . .	52
<b>4</b>	<b>Optimisation of PMC algorithm parameters</b>	<b>58</b>
4.1	A sufficient convergence statistic . . . . .	58
4.2	Defining efficiency . . . . .	61
4.3	Simulations . . . . .	62
4.4	The Gaussian target distribution . . . . .	63

---

4.4.1	Sensitivity to mixture density parameters . . . . .	63
4.4.2	Optimal number of components . . . . .	65
4.4.3	Optimal sample size . . . . .	67
4.5	Complex target distributions . . . . .	72
4.5.1	Banana distribution . . . . .	72
4.5.2	Bimodal distribution . . . . .	75
4.5.3	Hypercube distribution . . . . .	78
<b>5</b>	<b>Application of PMC methods to cosmology</b>	<b>83</b>
5.1	The likelihood function for CMB data . . . . .	83
5.2	Parameter estimation: Adiabatic model . . . . .	86
5.2.1	Cosmological data . . . . .	87
5.2.2	Cosmological parameters . . . . .	87
5.2.3	Optimizing PMC for cosmological parameter estimation: Adiabatic model	88
5.3	Sampling models with the adiabatic mode correlated with isocurvature modes . .	92
5.3.1	Adiabatic mode correlated with a single isocurvature mode . . . . .	96
5.3.2	Adiabatic mode correlated with two isocurvature modes . . . . .	105
5.3.3	Adiabatic mode correlated with all three isocurvature modes . . . . .	113
<b>6</b>	<b>Conclusion</b>	<b>119</b>

---

## List of Figures

---

2.1	Blackbody spectrum of the CMB photon distribution. . . . .	15
2.2	The CMB temperature anisotropy spectrum from WMAP nine-year data. . . . .	16
2.3	Effect of baryon density on the CMB temperature anisotropy spectrum. . . . .	32
2.4	Effect of cold dark matter density on the CMB temperature anisotropy spectrum. . . . .	33
2.5	Effect of the cosmological constant density on the CMB temperature anisotropy spectrum. . . . .	34
2.6	Effect of the optical depth to the last scattering surface on the CMB temperature anisotropy spectrum. . . . .	35
2.7	Effect of the spectral index on the CMB temperature anisotropy spectrum. . . . .	36
2.8	Effect of the scalar amplitude on the CMB temperature anisotropy spectrum. . . . .	37
2.9	Comparison of CMB temperature power spectra produced by adiabatic and isocurvature modes . . . . .	38
4.1	Convergence statistic for PMC (ESS) versus convergence statistic for MCMC ( $r$ ) . . . . .	60

---

4.2	Efficiency of the PMC sampler against the algorithm parameters that initialise the mixture densities. . . . .	64
4.3	The efficiency against the minimum number of components needed for convergence. . . . .	66
4.4	The minimum number of iterations ( $T$ ) needed for convergence against sample size. The corresponding efficiency against sample size . . . . .	68
4.5	Fitting function for $T$ to the data for a three dimensional Gaussian target distribution against sample size. The corresponding fit to data of efficiency . . . . .	69
4.6	Three plots illustrating optimal sample size, computational cost and optimal iteration against dimension of the target distribution . . . . .	70
4.7	Banana distribution in two dimensions with two different complexities. Components of a converged sample are also illustrated. . . . .	73
4.8	Banana distribution results. The minimum number of components required to sample a banana distribution as a function of complexity. Efficiency as a function of complexity. . . . .	74
4.9	Contour plot of two dimensional symmetric bimodal distribution. Plot illustrates modes separated by different distances. . . . .	75
4.10	Bimodal study. Plot of efficiency against the distance between modes ( $a$ ) with corresponding cost of sampling the bimodal distribution . . . . .	76
4.11	Bimodal study. Distributions for various ranges of $a$ . . . . .	77
4.12	One dimensional hypercube study. Optimal positioning of components for hypercube target distribution and histogram of ESS from simulation. . . . .	79
4.13	Hypercube study. Distribution of a converged chain from a hypercube target distribution. . . . .	80

---

4.14	Two dimensional hypercube target distribution results. Converged joint distribution of a sample and histogram of ESS statistics obtained from the simulation . . .	81
4.15	Hypercube study. Progression of components with iteration when sampling a hypercube distribution. . . . .	82
5.1	Pure adiabatic model. Progression of ESS for runs that use different amount of components when sampling a pure adiabatic model. . . . .	89
5.2	Pure adiabatic model. Marginal distributions between selected pairs of cosmological parameters showing degeneracies. . . . .	91
5.3	Adiabatic plus CI. Marginal posterior of cosmological parameters for AD+CI mixed model. . . . .	98
5.4	Adiabatic plus NID. Marginal posterior of cosmological parameters for the AD+NID mixed model. . . . .	99
5.5	Adiabatic plus NIV. Marginal posterior of cosmological parameters for the AD+NIV mixed model. . . . .	100
5.6	Adiabatic plus CI. Joint distributions illustrating degeneracies between parameters.	102
5.7	Adiabatic plus NIV. Joint distributions illustrating degeneracies between parameters. . . . .	103
5.8	Adiabatic plus NID. Joint distributions illustrating degeneracies between parameters. . . . .	104
5.9	Marginal posterior of cosmological parameters for AD+CI+NID model. . . . .	107
5.10	Marginal posterior of cosmological parameters for AD+CI+NIV model. . . . .	108
5.11	Marginal posterior of cosmological parameters for AD+NID+NIV model. . . . .	109
5.12	Adiabatic plus two isocurvature modes. CMB temperature spectra for AD+CI+NID and AD+CI+NIV best-fit models against the spectrum of a pure adiabatic model.	110

---

5.13	Adiabatic plus CI+NID model. Marginal distributions between selected pairs of cosmological parameters showing degeneracies. . . . .	111
5.14	Adiabatic plus CI+NIV model. Marginal distributions between selected pairs of cosmological parameters showing degeneracies. . . . .	112
5.15	Adiabatic plus NID+NIV. Marginal distributions between selected pairs of cosmological parameters showing degeneracies. . . . .	113
5.16	Marginal posterior of cosmological parameters for AD+CI+NID+NIV model. . .	116
5.17	Adiabatic plus all three isocurvature modes. CMB anisotropy spectra for all isocurvature contributions for a model with large isocurvature fraction and likelihood. . . . .	117
5.18	Adiabatic plus CI+NID+NIV. Marginal distributions between selected pairs of cosmological parameters showing degeneracies. . . . .	118

---

## List of Tables

---

4.1	Gaussian target distribution. Optimal sample size for a given dimension of target distribution. . . . .	71
5.1	Prior constraints for cosmological parameters. . . . .	88
5.2	Pure adiabatic model. Summary of costs and efficiency when sampling a pure adiabatic model for different number of components. . . . .	90
5.3	Pure adiabatic model. Estimates from converged chain compared to previous analysis from WMAP nine-year data. . . . .	92
5.4	The cost and efficiency of sampling models with correlated adiabatic and isocurvature modes. . . . .	95
5.5	Adiabatic plus one isocurvature mode results. Statistics from converged sample. .	101
5.6	Adiabatic plus two isocurvature mode results. Statistics from converged sample. .	106
5.7	Adiabatic plus all three isocurvature mode results. Statistics from converged sample. . . . .	114

# CHAPTER 1

---

## Introduction

---

There has been a dramatic increase in both the quantity and quality of cosmological data from different observations. These observational data have improved our understanding of the universe by awarding merit to theoretical models that explain the data well and provide estimates for those quantities or parameters that these models depend on. Of the observed data, there are arguably three types that have revolutionised cosmology viz: the cosmic microwave background (CMB) radiation, specifically CMB anisotropies, observed by space-based and all-sky experiments such as the Wilkinson Microwave Anisotropy Probe (WMAP) [1, 2] and Planck [3, 4], standard candles such as Type Ia Supernovae [5, 6], and the detection of the imprint of baryon acoustic oscillations (BAO) in the early universe on galaxy clustering [7].

The CMB is the residual radiation from decoupling after the epoch of recombination when the universe was only a few hundred thousand years old. Having travelled virtually undisturbed ever since makes it a powerful probe of cosmology and early-universe physics. The tiny anisotropies reflect the small perturbations in the energy density of the universe at the initial time. Supernovae

data indicate that the expansion rate is increasing [6], thereby strongly motivating the case for dark energy.

All data points to the following: The strongly favoured standard cosmological model depicts a flat or nearly flat universe that is approximately 13.7 billion years old and made of just under 30% non-relativistic matter, of which only one-sixth is ordinary baryonic matter. Roughly 70% of the contents of the universe is dark energy, with the remaining contents being cold dark matter (CDM). Cold dark matter can only be observed by its gravitational interaction with baryonic matter [8]. It also explains the gravitational amplification of the small fluctuations, seen in the CMB, capable of forming the large scale structures that we see today. The dark energy component is believed to be in the form of vacuum energy density referred to as the cosmological constant,  $\Lambda$ , or a time varying component scalar field such as quintessence [9]. Dark energy is the mysterious component whose domination of the energy density is hypothesized to explain the accelerated expansion of the universe. There is much evidence in current observational data to support a model of dark energy with the equation of state (i.e., the ratio of pressure to energy density) exactly equal to -1 [10]. Relativistic components, such as photons and neutrinos make up a small fraction of the total density of the universe today.

With the large amount of cosmological data come new challenges requiring sophisticated data analysis methods. The scientific interpretation of such data requires a deep understanding of statistics. Over the last three centuries astronomy has aided the development of statistical methodology since many statistical theories were originally formulated as astronomical problems. Modern statistics in the field of astronomy, commonly known as astrostatistics, has been growing since the 1990's, stimulated by astronomical image processing [11, 12], galaxy clustering [13], Bayesian analysis [14], and astrostatistics in general [15]. Bayesian analysis plays a key role in cosmology and amongst its applications is the analysis of CMB data which has been providing insights into modern quantitative cosmology [16]. Some of the key statistical

applications in cosmology are model selection and parameter estimation.

Cosmologists often perform statistical inference by fitting theoretical models to cosmological data, establishing cosmological parameter estimates and confidence intervals to quantify the uncertainty in those estimates. In Bayesian statistics, estimates are obtained from evaluating the *posterior probability distribution* of the parameters [17], with recent cosmological data, enabling precision parameter estimation [18, 19, 20]. In cosmology, the posterior distribution is generally multidimensional and lacks an analytical expression. Therefore, evaluating the posterior involves integration over a large multidimensional parameter space, which is computationally expensive and time consuming. Evaluating these integrals can be done with the aid of efficient sampling algorithms.

Sampling algorithms provide a sequence of observations which are approximated from the posterior distribution when direct sampling is difficult. This chain sequence, or sample set, can then be used to approximate quantities with respect to the posterior distribution. Algorithms such as *nested sampling* [21] are widely used in the field of cosmology [22, 23, 24, 25] through publicly available software packages such as CosmoNest [22, 23] and MultiNest [26]. The most popular sampling algorithm is *Markov Chain Monte Carlo* (MCMC) [27] with its own publicly available package for cosmological applications called COSMOMC [28].

This thesis focuses on a hybrid *Adaptive Importance Sampling* (AIS) algorithm [29] called *Population Monte Carlo* (PMC) [30], which has already been applied to the field of cosmology [31, 32, 33] with its own publicly available code called CosmoPMC [34]. PMC is iterative in design, generating independent sample sets at each iteration, which can be used for parameter estimation. This enables easy parallelisation of a single run thereby distributing the time-consuming likelihood evaluation for these models and significantly reducing the wall-clock time. The algorithm has its own convergence criterion enabling the sampler to automatically stop when this criterion has been satisfied, thereby avoiding regular monitoring and analysis of the chains.

---

The PMC algorithm, although automated, must be initialised according to the desired target distribution. These inputs affect the performance of the sampler as measured by computational expense and wall-clock time used to attain the sample. It is therefore necessary to optimise PMC with regards to these inputs. We address this problem by deriving a statistic that will measure the effect that varying algorithm parameters will have on the performance of PMC. We undergo a systematic study to optimise the performance of PMC, thereby improving the efficiency, after which we apply our results to the parameter estimation problem in cosmology as in a similar study done for MCMC [35].

The MCMC algorithm has been extensively applied to cosmology including deriving constraints on cosmological parameters such as the baryon density ( $\Omega_b h^2$ ), cold dark matter density ( $\Omega_c h^2$ ), the energy density of the cosmological constant ( $\Omega_\Lambda$ ), the scalar spectral index ( $n_s$ ) and optical depth to reionisation ( $\tau$ ), using CMB data from WMAP-5 [36]. The power spectrum of galaxies, observed by the Sloan Digital Sky Survey (SDSS) [37] in conjunction with CMB data from WMAP have been used to provide extensive cosmological parameter estimations using MCMC [38]. MCMC has been used to provide constraints on the normalisation of the matter power spectrum,  $\sigma_8$ , and other applications of cosmological parameter estimation using baryon acoustic oscillations from SDSS [39, 40], and undertaken a comparison study between CMB data from WMAP nine-year [40, 41] and Planck [42, 43]. The population Monte Carlo algorithm, on the other hand, has recently been introduced to the cosmology community. It is not as widely used as MCMC. Nonetheless applications involving cosmological parameter estimation using CMB data from WMAP-5 for the standard  $\Lambda$ CDM model have already been done using PMC [31], and even provided a comparison to MCMC. PMC has also provided constraints on  $\sigma_8$  and the dark energy equation of state [44] and the publicly available CosmoPMC has been modified to incorporate various data such as CFHTLenS, BAO from SDSS [44, 45], with extended applications to model selection [33, 45]. The PMC algorithm, although relatively new in

the cosmology community, is increasing in popularity therefore making it a useful alternative to MCMC.

In addition to measuring the basic cosmological parameters, cosmologists are interested in the physical processes that played a role in the early universe, in particular inflation and the models that describe it. The simplest inflationary models predict purely adiabatic fluctuations. We know that a purely adiabatic model fits present data very well, but so does a mixture of adiabatic perturbations, correlated with isocurvature mode contributions [46]. A necessary study is therefore to determine how much isocurvature contribution is permitted by data. Early constraints [47, 48, 49, 50] on isocurvature perturbations were established shortly after the Cosmic Background Explorer (COBE) [51]. However, it was only after the precisely observed first acoustic peak in the CMB by Boomerang [52] and Maxima [53] that reasonable constraints on isocurvature perturbations [46] became possible. Some constraints before the WMAP first-year data release were carried out [54, 55, 56, 57] with some studies looking at the effect of isocurvature initial conditions on relaxing the constraints on  $\Omega_\Lambda$  [58, 59]. The WMAP first-year data [60] was used to constrain various mixtures of adiabatic and isocurvature perturbations [1, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75]. Results for similar studies were established using a combination of data from WMAP-5 [76], the Supernovae Legacy Survey (SNLS) [77] and SDSS in [78]. Constraints on adiabatic mode with an additional isocurvature mode using more recent data from the WMAP seven- and nine-year experiments as well as Planck have been carried out in [79, 80, 81]. Constraining models with mixtures of adiabatic and isocurvature perturbations becomes more difficult since the complexity of the distributions to be investigated increases as more isocurvature modes are added to the model. It is therefore necessary to establish a reliable and effective way of sampling these distributions if we are to determine which models are preferred by the data and how much isocurvature contribution is allowed. This thesis addresses this problem using PMC.

This thesis begins with Chapter two providing the foundation for the cosmology used in this study. We review the standard cosmological model and the relevant cosmological data used for cosmology. We outline the inflationary models explaining the initial conditions of the universe and how the CMB anisotropies is used to discriminate such models. Chapter 3 describes the role of parameter estimation and the MCMC and PMC algorithms in detail. Chapter 4 contains results of our study to optimise the PMC algorithm through the use of simulations. We describe the simulations and illustrate our findings for PMC against various target distributions. In Chapter 5, we apply our results from Chapter 4 to the cosmological parameter estimation problem of constraining mixed adiabatic and isocurvature perturbation models. In the last chapter, we present concluding remarks.

## CHAPTER 2

---

### The standard cosmological model

---

There are various cosmological models that aim to explain the universe we live in. Fortunately, observational data has improved our understanding of the universe by discriminating between these models. It is widely accepted that we live in a homogeneous and isotropic universe that has large scale structures such as galaxy clusters and CMB anisotropies originating from small initial fluctuations in the matter and radiation in the early universe, which grew due to gravitational instability. This is known as the Big Bang model. These effects can be explained mathematically and have well-established physics to describe them. The fact that all models are subject to mathematical scrutiny provides a concrete justification for models that can explain the data. In this Chapter, we outline the standard cosmological model along with underlying linear perturbation theory, which is the basis for understanding the formation of large scale structures and CMB anisotropies. We begin by introducing the Friedmann-Robertson-Walker model, which forms the basis of the Big Bang model, followed by the Einstein equations and the energy momentum conservation principle to derive the evolution equations for the different species of the universe.

Then follows a description of the initial conditions and the role of the CMB anisotropies as a probe into the early universe. We explain why the CMB is a useful tool in providing constraints on the cosmological parameters.

## 2.1 The expansion of the universe

The universe was first observed to be expanding by Edwin Hubble in 1929 [82], who related the recessional velocity  $v$  of a galaxy to its distance  $d$  away from us by

$$v = H(t)d \quad (2.1.1)$$

where  $H(t)$  is the Hubble parameter that represents the expansion rate as a function of time. The Hubble parameter is related to the scale factor  $a(t)$  describing the expansion of the universe according to

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \quad (2.1.2)$$

where  $\dot{a}$  refers to the derivative of  $a$  with respect to time  $t$ . At present time  $t_0$ , the Hubble parameter is measured to be  $H_0 = 67.4 \pm 1.4 \text{ km sec}^{-1} \text{ Mpc}^{-1}$  [3].

The evolution of the scale factor for a homogeneous and isotropic universe is described by

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\mathcal{K}c^2}{a^2} + \frac{\Lambda c^2}{3}, \quad (2.1.3)$$

which is known as the Friedmann equation, with  $\rho$  representing the mass-energy density. The symbol  $\Lambda$  represents the cosmological component and in the standard model, has been shown to produce the current accelerated expansion phase. The term  $\mathcal{K}$  is the curvature of the universe and

usually takes on three discrete values for the geometries, namely,

$$\mathcal{K} = \begin{cases} 1 & \text{Closed universe,} \\ 0 & \text{Flat universe,} \\ -1 & \text{Open universe.} \end{cases}$$

The symbol  $c$  represents the speed of light and can be scaled to 1 by using a geometric unit system. Therefore we re-write the Friedmann equation as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{\mathcal{K}}{a^2}, \quad (2.1.4)$$

where  $\Lambda = 0$ . Letting  $p$  represent the pressure resulting from mass-energy  $\rho$ , we have the local energy conservation equation

$$\dot{\rho} + 3H(\rho + p) = 0. \quad (2.1.5)$$

The second term in equation (2.1.5) corresponds to the dilution of  $\rho$  due to the Hubble expansion. The third term represents the work done by the pressure of the fluid. We also have mass-energy related to the pressure by

$$p = \omega\rho, \quad (2.1.6)$$

where  $\omega = \omega(\rho)$  depends only on the local energy density and is known as the linear equation of state. Equation (2.1.6) models the cosmological constant,  $\Lambda$ , with  $\omega = -1$  corresponding to  $\Lambda$  set to zero. For a flat universe ( $\mathcal{K} = 0$ ) dominated by one fluid with equation of state  $\omega = \text{constant}$ , we solve for  $\rho$  using equations (2.1.6) and (2.1.5) to obtain

$$\rho \propto a^{-3(1+\omega)}, \quad (2.1.7)$$

which relates the mass-energy to the scale factor. This relation is useful in determining the dominant component during any time period. These components are matter, radiation, and the

cosmological constant  $\Lambda$ .

$$\text{Matter: } \omega = 0, \quad \rho \propto a^{-3}, \quad (2.1.8)$$

$$\text{Radiation: } \omega = \frac{1}{3}, \quad \rho \propto a^{-4}, \quad (2.1.9)$$

$$\Lambda : \quad \omega = -1, \quad \rho \propto \text{constant}. \quad (2.1.10)$$

By differentiating equation (2.1.4) with respect to time we obtain

$$2\frac{\dot{a}}{a} \left( \frac{\ddot{a}a - \dot{a}^2}{a^2} \right) = \frac{8\pi G}{3} \dot{\rho} + 2\frac{\mathcal{K}\dot{a}}{a^3}. \quad (2.1.11)$$

Substituting for  $\dot{\rho}$  from (2.1.5) and using (2.1.4) again, equation (2.1.11) can be written as

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{3}(\rho + 3p), \quad (2.1.12)$$

which is known as the Raychaudhuri equation [83].

We can relate the scale factor to time depending on the dominant component. It is convenient to normalise the scale factor so that it equals 1 at present time  $t_0$  i.e.  $a_0 = 1$ . We know from (2.1.9) that  $\rho \propto 1/a^3$  for matter domination. Fixing the proportionality constant by the present density  $\rho_0$  yields

$$\rho = \frac{\rho_0}{a^3}. \quad (2.1.13)$$

Substituting (2.1.13) into the Friedmann equation gives the following result

$$\dot{a}^2 = \frac{8\pi G}{3} \frac{\rho_0}{a}. \quad (2.1.14)$$

To solve this, we suppose  $a$  follows a power law,  $a \propto t^q$ ; therefore  $\dot{a} \propto t^{q-1}$ . In the right hand side we have  $a^{-1} \propto t^{-q}$ . These powers must be equal; therefore,  $q = 2/3$ , which means that

$$a(t) \propto \left( \frac{t}{t_0} \right)^{2/3}, \quad (2.1.15)$$

if matter is the dominant component. Similarly if radiation dominates,

$$a(t) \propto \left( \frac{t}{t_0} \right)^{1/2}, \quad (2.1.16)$$

so the universe expands faster if matter dominates instead of radiation. Similarly in a  $\Lambda$  dominated universe

$$a(t) \propto e^t \quad (2.1.17)$$

and we have exponential growth.

The density parameters are generally expressed as dimensionless quantities by comparing them to the critical density  $\rho_{crit} = \frac{3H^2}{8\pi a^2 G}$ . The critical density is the boundary value between a contracting universe and expanding universe. If the density of the universe is found to be greater than  $\rho_{crit}$ , then we have a closed model, whereas for a universe that has a density less than  $\rho_{crit}$  we have an open model. We use subscripts  $r$ ,  $m$ , and  $\Lambda$  to denote the radiation, matter, and dark energy component, respectively, to simplify the notation. The density parameters expressed as ratios of  $\rho_{crit}$ , depending on the dominant component are

$$\Omega_r = \frac{\rho_r}{\rho_{crit}}, \quad (2.1.18)$$

$$\Omega_m = \frac{\rho_m}{\rho_{crit}}, \quad (2.1.19)$$

and

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{crit}} = \frac{\Lambda}{8\pi G \rho_{crit}}, \quad (2.1.20)$$

with curvature given by

$$\Omega_{\mathcal{K}} = \frac{\rho_{\mathcal{K}}}{\rho_{crit}} = \frac{-3\mathcal{K}}{8\pi G a^2 \rho_{crit}}. \quad (2.1.21)$$

These quantities are constrained by

$$\Omega_r + \Omega_m + \Omega_\Lambda + \Omega_{\mathcal{K}} = 1, \quad (2.1.22)$$

which is a dimensionless form of the Friedmann equation. The current density  $\rho_0$ , excluding curvature, in terms of the current critical density is given by

$$\Omega_0 = \Omega_{r0} + \Omega_{m0} + \Omega_{\Lambda 0}, \quad (2.1.23)$$

where  $\Omega_{r0}$ ,  $\Omega_{m0}$ , and  $\Omega_{\Lambda 0}$  are the current density for radiation, matter and the dark energy component respectively. The Friedmann equation is an important equation in cosmology as it describes how the expansion of the universe evolves with time. The scale factor is related to the redshift of distant objects by

$$a = (1 + z)^{-1}, \quad (2.1.24)$$

where we have set  $a_0 = 1$  as mentioned previously. Using this relation, we can write the Friedmann equation in terms of redshift to yield

$$H^2(z) = H_0^2 [\Omega_m(1 + z)^3 + \Omega_r(1 + z)^4 + \Omega_\Lambda + \Omega_k(1 + z)^2]. \quad (2.1.25)$$

This equation explains how the expansion rates depends on the different components and the evolution of their energy densities with redshift. Since redshift is analogous with time, equation (2.1.25) shows that radiation was the dominant component during the early universe and that the cosmological constant will dominate eventually. Determining these densities is therefore, vital in our understanding of the evolution of our universe. Matter is made of two non-relativistic components, namely the visible baryonic and cold dark matter components such that  $\Omega_m = \Omega_b + \Omega_c$  where  $\Omega_b$  and  $\Omega_c$ , are the baryonic and cold dark matter densities relative to the critical density, respectively. Cold dark matter has many candidates, namely: WIMPs, Axions, WIMPzillas, and Primordial black holes [84, 85]. In a similar way, radiation is made up of relativistic photons and neutrinos, since we assume them to be massless as in the standard cosmological model. The CMB photons are the dominant source of photons in the universe. We therefore neglect those

generated from stars since their contribution to the density is very small. We denote the density of CMB photons and neutrinos as  $\Omega_{CMB}$  and  $\Omega_\nu$ , respectively with  $\Omega_r = \Omega_{CMB} + \Omega_\nu$ .

The standard cosmological model, including experiments depicts a flat universe with mean energy density of  $\bar{\rho}_{cr} = 9.9 \times 10^{-30} \text{ g cm}^{-3}$  at an age of  $13.75 \pm 0.11$  billion years old and made up of relativistic components, such as photons and neutrinos as well as non-relativistic ordinary matter consisting of baryons, dark matter, and dark energy. The matter that we see, i.e., baryonic matter, only constitutes 5% of the total content of the universe, whereas dark energy accounts for 68% [3].

The most compelling evidence for dark energy came from observations of Type Ia supernovae in 1998 by A. G. Riess *et al.* [5] and S. Perlmutter *et al.* [6] which indicated an increasing expansion rate. It is a hypothetical form of energy that tends to increase the expansion rate of the universe. Dark energy is supported by measurements of the CMB anisotropies [3]. There are two well known conjectures as to what dark energy is, namely the cosmological constant,  $\Omega_\Lambda$ , which is equivalent to the vacuum energy, filling space homogeneously with a constant energy density, and scalar fields with energy density changing in space and time.

The cold dark matter represents the remaining contents of the universe and was first introduced by Fritz Zwicky in 1934 to account for the "missing mass" in the dynamics of galaxy clusters [86, 87]. It is detectable through gravitational interaction with baryonic matter and has an impact on the CMB anisotropies [41].

## 2.2 The cosmic microwave background

The Big Bang model was strongly favoured since 1964 due to the discovery of the CMB radiation by Arno Penzias and Robert Wilson [88]. This radiation was predicted as a remnant of the Big Bang after 379,000 years. The CMB photons that free stream towards us and often regarded as an

image of the universe during its infant years. It is the moment of last scattering, when the photons scattered off the electrons and have been freely travelling through the universe ever since. Thus, we cannot see beyond the CMB since the photons were tightly coupled to the baryons.

Measurements by the Cosmic Background Explorer (COBE) satellite confirmed the perfect black body spectrum of the CMB with a temperature of 2.725 K and detected the fluctuations in the CMB at a level of  $10^{-5}$  K [89]. The CMB photons have a perfect blackbody spectrum, which is strong evidence to support the big bang theory over the Steady-State universe [90], because the photons had to be in equilibrium before they freestreamed towards us. This intensity as a function of frequency  $\nu$ , is given by

$$I_\nu = \frac{4\pi\hbar\nu^3/c^2}{\exp(2\pi\hbar\nu/k_B T) - 1}, \quad (2.2.26)$$

for  $\hbar$  the reduced Planck constant,  $k_B$  the Boltzmann constant and  $T$  its temperature. The spectrum is shown in figure 2.1. Following COBE, the Wilkinson Microwave Anisotropy Probe (WMAP) has produced more accurate measurements of the CMB anisotropy spectrum and better constraints on cosmological parameters that affect the CMB spectrum [41, 91].

The anisotropies of the CMB are a good way to compare theory with observations. An important CMB statistic is the two-point correlation function of the temperature distribution. The distribution is expanded in spherical harmonics since it is a two-dimensional field measured everywhere on the sky, therefore the two-point function of the CMB is a function of multipole moment  $l$ .

## 2.3 Large-scale structure

The universe on large scales is homogeneous and isotropic, however on small scales the universe is not as smooth, because there are observed structures such as planets, stars, and galaxies. The various forms of structure in the Universe are collectively referred to as large-scale structure

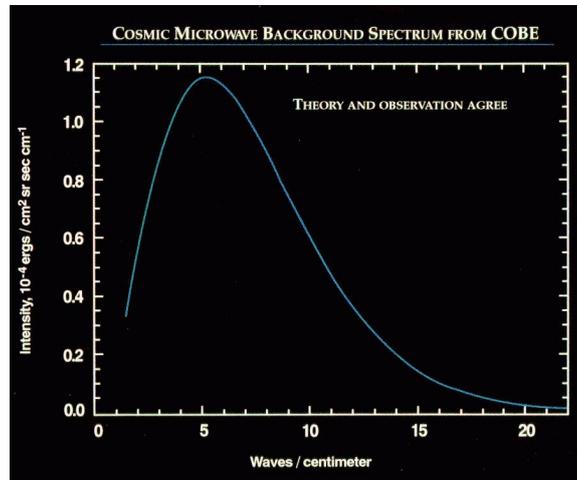


Figure 2.1: The intensity of the CMB radiation as a function of wavenumber measured from the Far InfraRed Spectrophotometer (FIRAS) [92] instrument on the COBE satellite.

(LSS). Galaxies are useful for understanding cosmology. Galaxies tend to cluster together by gravitational attraction in filaments with voids between them. Galaxy clusters can contain thousands of galaxies and also tend to cluster together with other galaxy clusters. The distribution of galaxies helps us to understand the universe, from the time they started forming, at  $z \approx 10$ , to the present universe, at  $z \approx 0$ . We now have a large enough sample of galaxies with corresponding redshifts to study galaxy clustering and the evolution of our universe.

The growth of structure under gravitational instability is key in explaining the evolution of structure. The overdensities in the matter distribution, which grows after matter-radiation equality, create regions with more mass than others. The perturbations grow in time under the influence of gravity to form the structures we see today. Gravitational instability is influenced by properties of the universe, such as the expansion rate and its material composition. Hence it can be used to estimate cosmological parameters.

Surveys of large-scale structure aim to provide a 3-dimensional image of the universe by digitizing the entire sky. A large enough sample of galaxies, quasars and stars with correspond-

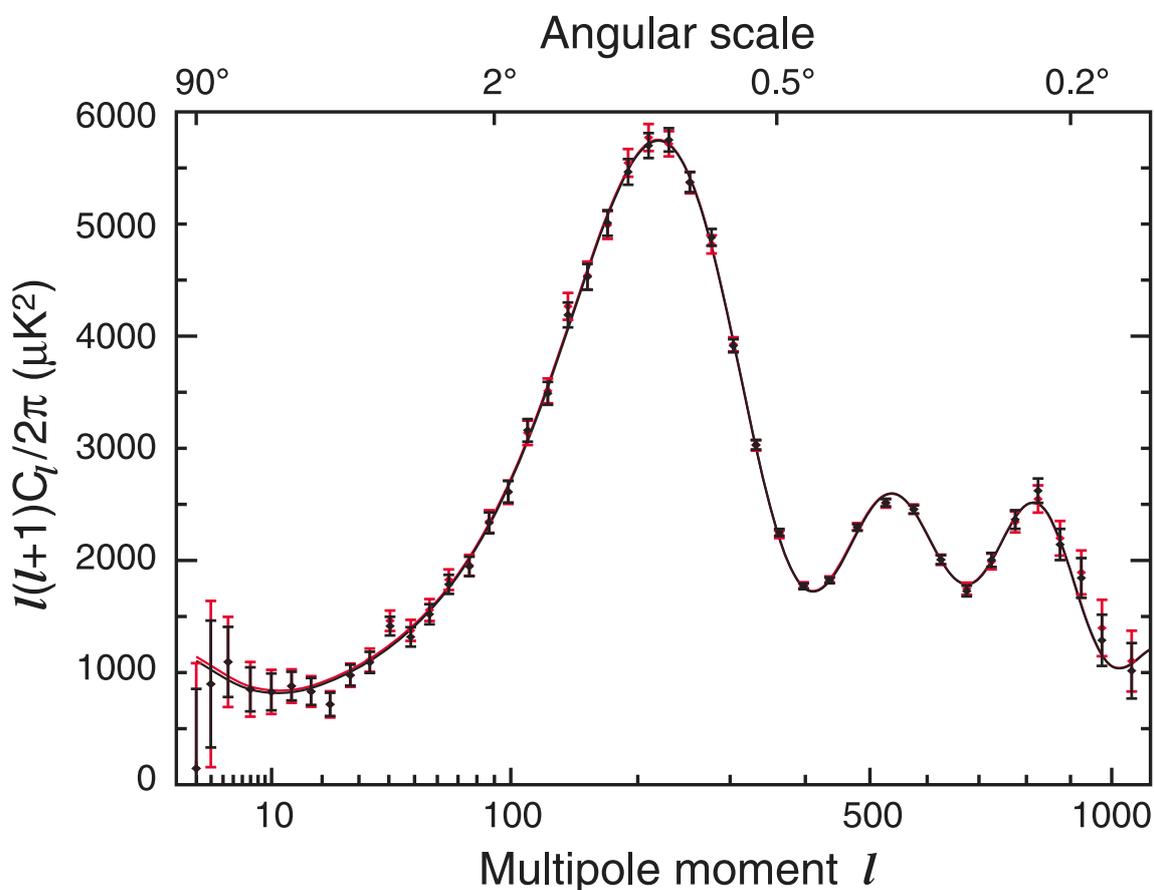


Figure 2.2: CMB temperature anisotropy spectrum. Taken from the 9-year WMAP data [41]. Two estimates of the WMAP nine-year power spectrum along with the best-fit model spectra obtained from each; *black*-a weighted spectrum and best-fit model; *red*-the same but for a Monte Carlo estimated spectrum and model.

ing redshifts collectively sample the universe as a whole. The Center for Astrophysics (CfA) Redshift Survey [93] was the first attempt to map the large-scale structure of the universe with the Sloan Digital Sky Survey [94] being the most recent, obtaining about 930 000 galaxies, 120 000 quasars, and 460 000 stars. Statistical measures are in place to extract information from the samples such as the two-point correlation function which measures the likelihood of two galax-

ies having a given separation,  $r$ , and its Fourier space analogue, the power spectrum,  $P(k)$ , that decomposes the pattern into waves with wavenumber  $k \equiv 2\pi/\lambda$  and specifies the amplitude of those waves. We derive the power spectrum by quantifying the perturbation as the difference in the density in a specific region compared to the average background density, given by

$$\delta = \frac{\rho(\mathbf{x}) - \bar{\rho}(\mathbf{x})}{\bar{\rho}(\mathbf{x})}, \quad (2.3.27)$$

where  $\rho(\mathbf{x})$  is the density of that region and  $\bar{\rho}(\mathbf{x})$  is the average density of the background. The density perturbation simplifies in Fourier space and we use the convention

$$A(\mathbf{x}) = \int \frac{1}{(2\pi)^3} A(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{x}) d^3\mathbf{k}, \quad (2.3.28)$$

to obtain

$$\delta(\mathbf{x}) = \int \frac{1}{(2\pi)^3} \delta(\mathbf{k}) \exp(i\mathbf{k}\cdot\mathbf{x}) d^3\mathbf{k}, \quad (2.3.29)$$

where  $\delta(\mathbf{k})$  is the Fourier component of  $\delta(\mathbf{x})$ . From isotropy, we have all the  $\mathbf{x}_i$  as being identical and the perturbation distribution can be characterised by the power spectrum,

$$P(k) = \langle |\delta(k)|^2 \rangle, \quad (2.3.30)$$

where  $\langle \cdot \rangle$  is the expectation over many realisations.

Cosmological models with different parameters produce predictions of these quantities which can be compared to the observed data. Hence by obtaining the best fit, we are able to constrain the cosmological parameters. Dark matter dominates the matter density and is required to produce the universe we see today and our theory of dark matter can be compared to observations to constrain it by producing compatible galaxy distributions. In addition to the matter density, other cosmological parameters such as the dark energy density, Hubble parameter, normalisation of density perturbations and even the dark energy equation of state can be constrained with this data [95, 96, 97].

## 2.4 Type Ia supernovae

The eventual fate of the universe can be determined by measuring the properties of distant type Ia supernovae since they provide a record of changes in the expansion over time. It is now possible to make precise measurements of distant type Ia supernovae and by extending our measurements of the expansion history to earlier times, we can indirectly observe the slowing down caused by the gravitational attraction due to the mass density. Mathematically we can illustrate this by re-writing the Friedmann equation in equation (2.1.25) as

$$H^2(z) = H_0^2 [\Omega_m(1+z)^3 + (1-\Omega_\Lambda)(1+z)^{3(1+\hat{w})}], \quad (2.4.31)$$

with

$$\hat{w}(z) \ln(1+z) = \int_0^z \frac{w(t)}{1+t} dt. \quad (2.4.32)$$

From equation (2.1.12), we have the redshift at the beginning of the acceleration epoch,

$$z_{\text{acc}} = \left[ (3w(z_{\text{acc}}) + 1) \frac{\Omega_m - 1}{\Omega_m} \right]^{-1/3\hat{w}(z_{\text{acc}})} - 1. \quad (2.4.33)$$

Observed supernovae of type Ia provide measurements for  $z_{\text{acc}}$  thus enabling us to constrain the matter density and dark energy [98, 99]. There are recent experiments such as the Supernova Legacy Survey (SNLS) [100], whose primary goal is to measure the equation of state of dark energy by measuring several hundred type Ia supernovae with remarkable precision at redshifts  $0.3 < z < 1$ . Another useful observation for cosmology is the baryon acoustic oscillations, which we discuss next.

## 2.5 Baryon acoustic oscillations

The baryon-photon fluid has regions of overdensity which promotes gravitational attraction; while the heat from the photon and matter interact to produce an outward pressure to counter

the force of gravity. These cancellation of forces produce oscillations similar to sound waves, that propagate in the photon-baryon fluid. These oscillations are referred to as baryon acoustic oscillations (BAO), which is a standard ruler because we can estimate its true size with high precision from measurements of the CMB acoustic peaks. The oscillations manifest themselves as regular periodic fluctuations in the density of the baryonic matter of the universe. The BAO matter clustering provides a standard ruler for a length scale in cosmology [101], that can be measured from the large-scale distribution of matter [7] and used to measure the expansion history of the universe, and hence the evolution of dark energy with cosmic time [102, 103, 104]. In addition to these constraints,  $\Omega_m, \Omega_\Lambda$ , and the geometry of the universe are constrained by the observed acoustic oscillations in the CMB and large-scale galaxy distributions [7, 105]. The Baryon Oscillation Spectroscopic Survey (BOSS) [106] is expected to map the baryon acoustic oscillation signature with unprecedented accuracy thereby improving the constraints on the Hubble expansion and dark energy.

## 2.6 Background and perturbed Friedmann-Robertson-Walker models

The FRW model describes a homogeneous and isotropic universe and is given by the line element,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = a^2(\tau) [-d\tau^2 + \gamma_{ij}(\vec{x}) dx^i dx^j], \quad (2.6.34)$$

where indices  $\mu$  and  $\nu$  range from 0 to 3,  $a$  being the scale factor accounting for the expansion rate of the universe and  $g_{\mu\nu}$  the unperturbed Robertson-Walker metric. The spatial coordinates are indicated by  $i$  and  $j$  and range from 1 to 3 with  $\gamma_{ij} = g_{ij}/a^2$  being the symmetric spatial part of the metric for a constant curvature space. The conformal time is represented by  $\tau$  and is related to proper time by  $dt = a(\tau)d\tau$ . The line element can also be written in spherical

coordinates with  $k$  representing the curvature of space in the following way:

$$ds^2 = a^2(\tau) \left( -d\tau^2 + \left[ \frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right] \right). \quad (2.6.35)$$

The constant  $k$  takes the values 1, 0 or -1, corresponding to a close (spherical), flat (euclidean), or open (hyperbolic) universe, respectively. Units are chosen such that the speed of light is unity.

In the presence of perturbations, there is no uniquely preferred coordinate system although the coordinates must reduce to equation 2.6.34 in the limit of zero perturbation. A coordinate system satisfying this criteria is called a gauge. The perturbed FRW metric accounts for the inhomogeneity of the universe and the most general first-order perturbation to the line element is [107],

$$ds^2 = a^2(\tau) \left[ -d\tau^2 + \gamma_{ij}(\vec{x}) dx^i dx^j + h_{\mu\nu}(\vec{x}, \tau) dx^\mu dx^\nu \right], \quad (2.6.36)$$

where  $h_{\mu\nu}$  is the metric perturbation and its components can be explicitly written as

$$h_{00} = -2A,$$

$$h_{0i} = -B_i,$$

$$h_{ij} = 2(D\gamma_{ij} + E_{ij}),$$

where  $E_{ij}$  is a traceless 3-metric with the trace of  $h_{ij}$  proportional to  $D$ . For a flat universe ( $k = 0$ ), a Cartesian coordinate system can be chosen such that  $\gamma_{ij} = \delta_{ij}$ . The term  $A$  is referred to as the lapse function and perturbs the conformal time component, which also specifies the relation between  $\tau$  and proper time along the threading. The term  $B_i$  is referred to as the shift function, and specifies the relative velocity between the threading and the world-lines orthogonal to the slicing and  $2(D\delta_{ij} + E_{ij})$  is the perturbation to the curvature of space.

The components of the metric perturbation are broken down further into scalar, vector and tensor perturbations by decomposing every symmetric tensor and every vector into longitudinal

and transverse parts. The tensor mode  $E_{ij}^T$  is gauge-invariant is the part of  $h_{ij}$  that is unattainable from the gradient of a scalar or vector. The vector mode are found in  $B_i^\perp$  and  $E_{ij}^\perp$  and corresponds to the transverse vector parts of the metric. There are two degrees of freedom to each part which can be eliminated by imposing a gauge condition. The scalar mode is spin-0 under spatial rotation and corresponds physically to Newtonian gravitation with relativistic modifications and is generally believed to be generated by the vacuum fluctuation of the inflation field. Any two of the scalar parts of the metric  $A$ ,  $D$ ,  $B_i^\parallel$  and  $E_{ij}^\parallel$  can be set to zero by a gauge transformation.

All three modes described above evolve independently in linear perturbation theory, therefore they can be examined individually. We consider only the scalar perturbations with a spatially flat background spacetime in this study. The line element is then

$$ds^2 = a^2(\tau)\{- (1 + 2A)d\tau^2 - B_i d\tau dx^i + [(1 + 2D)\delta_{ij} + 2E_{ij}]dx^i dx^j\}, \quad (2.6.37)$$

where  $A$ ,  $B_i$ ,  $D$  and  $E_{ij}$  are fixed according to the choice of gauge.

We first consider the conformal Newtonian gauge for scalar modes of the metric perturbations, which is characterised by scalar potentials  $\psi$  and  $\phi$ , the latter representing gravitational potential in the Newtonian limit. The four scalar perturbations to the metric are given by  $A = \psi$ ,  $D = \phi$  and  $B_i = E_{ij} = 0$ , with the line element

$$ds^2 = a^2(\tau)[-(1 + 2\psi)d\tau^2 + (1 + 2\phi)\delta_{ij}dx^i dx^j]. \quad (2.6.38)$$

The synchronous gauge includes scalar, vector and tensor modes with only the space-space component, ( $h_{00} = h_{0i} = 0$ ), of the metric tensor perturbed. The line element in this gauge is therefore

$$ds^2 = a^2(\tau)[-d\tau^2 + (\delta_{ij} + h_{ij})dx^i dx^j]. \quad (2.6.39)$$

## 2.7 Einstein equations and energy momentum conservation

The Einstein field equations relate the geometry of the universe to its matter contents by relating the Einstein tensor  $G_{\mu\nu}$  to the energy-momentum tensor  $T_{\mu\nu}$  according to

$$G_{\mu\nu} = R_{\mu\nu} - g_{\mu\nu} \frac{R}{2} = 8\pi G T_{\mu\nu}, \quad (2.7.40)$$

with  $R_{\mu\nu}$  and  $R \equiv g^{\mu\nu} R_{\mu\nu}$  representing the Ricci tensor and Ricci scalar, respectively. The Ricci tensor is given by

$$R_{\mu\nu} = \Gamma_{\mu\nu,\alpha}^{\alpha} - \Gamma_{\mu\alpha,\nu}^{\alpha} + \Gamma_{\beta\alpha}^{\alpha} \Gamma_{\mu\nu}^{\beta} - \Gamma_{\beta\nu}^{\alpha} \Gamma_{\mu\alpha}^{\beta}, \quad (2.7.41)$$

where  $\Gamma_{\alpha\beta}^{\nu}$  are the Christoffel symbols given by

$$\Gamma_{\alpha\beta}^{\mu} = \frac{g^{\mu\nu}}{2} \left[ \frac{\partial g_{\alpha\nu}}{\partial x^{\beta}} + \frac{\partial g_{\beta\nu}}{\partial x^{\alpha}} - \frac{\partial g_{\alpha\beta}}{\partial x^{\nu}} \right], \quad (2.7.42)$$

and commas denoting derivatives with respect to  $x$ .

The Einstein equations therefore relate the perturbations in the metric to the perturbations in matter and radiation [108]. From the Einstein equations, and considering a homogeneous universe with mean energy density  $\bar{\rho}(\tau)$  and pressure  $\bar{P}(\tau)$ , we obtain the evolution equations [109]:

$$H^2 = \left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi}{3} G a^2 \bar{\rho} - k, \quad (2.7.43)$$

$$\frac{d}{d\tau} H = \frac{d}{d\tau} \left( \frac{\dot{a}}{a} \right) = -\frac{4\pi}{3} G a^2 (\bar{\rho} + 3\bar{P}), \quad (2.7.44)$$

with the Hubble parameter shown as a function of conformal time. Equation 2.7.44 is called the acceleration equation, which expresses mathematically that the expansion of the universe is due to the density and pressure filling the universe with positive acceleration requiring a component with negative pressure  $\bar{P} < -\bar{\rho}/3$ .

If the matter components of the universe are treated as an ideal fluid at all times, then it can be described by the energy density contrast  $\delta$  and the velocity divergence  $\theta$ . However a full

treatment for the radiation component requires the use of the Boltzmann equation, however in the baryon-photon regime, the single fluid can be treated as a perfect fluid without loss of accuracy. For the case of non-massive neutrinos, the energy momentum tensor for such a fluid is given by [109]

$$T_\nu^\mu = \sum_n P_n g_\nu^\mu + (\rho_n + P_n) V^\mu V_\nu, \quad (2.7.45)$$

where  $V^\mu = \frac{dx^\mu}{\sqrt{-ds^2}}$  is the four-velocity of the fluid,  $\rho = \bar{\rho} + \delta\rho$  and  $P = \bar{P} + \delta P$  are the proper energy density and pressure in the fluid rest frame respectively, with the barred quantities referring to the background, and  $\delta\rho$  and  $\delta P$  are the density and pressure fluctuations, respectively. The components of the perturbed energy-momentum tensor are

$$T_0^0 = -(\bar{\rho} + \delta\rho) \quad (2.7.46)$$

$$T_i^0 = (\bar{\rho} + \bar{P})v_i = -T_0^i, \quad (2.7.47)$$

$$T_j^i = (\bar{P} + \delta P)\delta_j^i + \Sigma_j^i, \quad \Sigma_i^i = 0, \quad (2.7.48)$$

where  $v_i \equiv dx^i/d\tau$  is the coordinate velocity of the fluid (assumed non-relativistic), and  $\Sigma_j^i \equiv T_j^i - \delta_j^i T_k^k/3$  is the anisotropic shear perturbation which denotes the traceless component of  $T_j^i$ . The mass density and pressure of the FRW background universe are represented by  $\bar{\rho}(\tau)$  and  $\bar{P}(\tau)$  respectively.

Introducing variables  $\theta$  and  $\sigma$  such that,

$$(\bar{\rho} + \bar{P})\theta \equiv ik^j \delta T_j^0, \quad (2.7.49)$$

$$(\bar{\rho} + \bar{P})\sigma \equiv -(\hat{k}_i \hat{k}_j - \frac{1}{3}\delta_{ij})\Sigma_j^i, \quad (2.7.50)$$

where  $\theta = ik^j v_j$  is the divergence of the fluid velocity and  $\sigma$  is related to the shear stress. We also define the density contrast  $\delta \equiv \delta\rho/\bar{\rho} = -\delta T_0^0/\bar{\rho}$ .

Substituting the metric and stress-energy tensor in the Einstein equation, we obtain the field equations for the perturbations [109]:

$$k^2\eta - \frac{1}{2}\frac{\dot{a}}{a}\dot{h} = -4\pi Ga^2\delta\rho, \quad (2.7.51)$$

$$k^2\dot{\eta} = 4\pi Ga^2(\bar{\rho} + \bar{P})\theta, \quad (2.7.52)$$

$$\ddot{h} + 2\frac{\dot{a}}{a}\dot{h} - 2k^2\eta = -8\pi Ga^2\delta\rho, \quad (2.7.53)$$

$$\ddot{h} + 6\ddot{\eta} + 2\frac{\dot{a}}{a}(\dot{h} + 6\dot{\eta}) - 2k^2\eta = -24\pi Ga^2(\bar{\rho} + \bar{P})\sigma. \quad (2.7.54)$$

When  $\dot{a}$  approaches zero, i.e. the limit of no expansion, equation 2.7.52 reduces to the ordinary Poisson equation for gravity, viz  $\nabla^2\eta = -4\pi a^2 G\delta\rho$ . Terms that are proportional to  $\dot{a}$  account for the expansion and play an important role for modes with wavelengths ( $\lambda \sim 1/k$ ) comparable to or larger than the the Hubble radius,  $H^{-1}$ .

The subscript  $j = 1, 2, \dots, n$  is used to denote the different species. We define the critical density of the universe  $\bar{\rho}_{cr}$  as the total density required for a flat universe; therefore, the ratio of the density of each species to the critical density can be represented as  $\Omega_j \equiv \bar{\rho}_j/\bar{\rho}_{cr}$ . With this in mind, we can rewrite equations (2.7.52)-(2.7.54) as

$$k^2\eta - \frac{1}{2}\frac{\dot{a}}{a}\dot{h} = -\frac{3}{2}H^2\bar{\rho}_{cr}\sum_j\Omega_j\delta_j, \quad (2.7.55)$$

$$k^2\dot{\eta} = \frac{3}{2}H^2\bar{\rho}_{cr}\sum_j\Omega_j(1+w_j)\theta_j, \quad (2.7.56)$$

$$\ddot{h} + 2\frac{\dot{a}}{a}\dot{h} - 2k^2\eta = -9H^2\bar{\rho}_{cr}\sum_j\Omega_j c_{sj}^2\delta_j, \quad (2.7.57)$$

$$\ddot{h} + 6\ddot{\eta} + 2\frac{\dot{a}}{a}(\dot{h} + 6\dot{\eta}) - 2k^2\eta = -9H^2\bar{\rho}_{cr}\sum_j(1+w_j)\Omega_j\sigma_j, \quad (2.7.58)$$

where  $w_j = p_j/\rho_j$  is the equation of state for the  $j^{th}$  species and  $c_s^2 = \frac{\partial p_j}{\partial \rho_j}$  is the square of the sound speed. First order evolution equations for density and velocity perturbations can be

obtained from the conservation equations

$$T^{\mu\nu}_{;\mu} = \partial_\mu T^{\mu\nu} + \Gamma^\nu_{\alpha\beta} T^{\alpha\beta} + \Gamma^\alpha_{\alpha\beta} T^{\nu\beta} = 0, \quad (2.7.59)$$

which yield the time evolution equations for the density perturbations and the divergence of the fluid velocity

$$\dot{\delta} = -(1+w) \left( \theta + \frac{\dot{h}}{2} \right) - 3 \frac{\dot{a}}{a} \left( \frac{\delta P}{\delta \rho} - w \right) \delta, \quad (2.7.60)$$

$$\dot{\theta} = -\frac{\dot{a}}{a} (1-3w) \theta - \frac{\dot{w}}{1+w} \theta + \left( \frac{\delta P}{\delta \rho} \right) \frac{k^2 \delta}{1+w} - k^2 \sigma, \quad (2.7.61)$$

$$(2.7.62)$$

where  $w$  is given by the equation of state  $w \equiv P/\rho$ . For a constant equation of state, equations (2.7.61) and (2.7.62) simplify further since  $\delta P/\delta \rho \equiv c_s^2 = w$  and  $\dot{w} = 0$ , and are valid for the global fluid. The evolution equations for photons, baryons, cold dark matter and massless neutrinos are [109]

$$\dot{\delta}_\gamma = -\frac{4}{3} \theta_\gamma - \frac{2}{3} \dot{h}, \quad (2.7.63)$$

$$\dot{\delta}_b = -\theta_b - \frac{1}{2} \dot{h}, \quad (2.7.64)$$

$$\dot{\delta}_c = -\frac{1}{2} \dot{h}, \quad (2.7.65)$$

$$\dot{\delta}_\nu = -\frac{4}{3} \theta_\nu - \frac{2}{3} \dot{h}, \quad (2.7.66)$$

for the density contrasts  $\delta$ , and

$$\dot{\theta}_\gamma = k^2 \left( \frac{1}{4} \delta_\gamma - \sigma_\gamma \right) + a n_e \sigma_T (\theta_b - \theta_\gamma), \quad (2.7.67)$$

$$\dot{\theta}_b = -\frac{\dot{a}}{a} \theta_b + c_s^2 k^2 \delta_b + \frac{4 \bar{\rho}_\gamma}{3 \bar{\rho}_b} a n_e \sigma_T (\theta_\gamma - \theta_b), \quad (2.7.68)$$

$$\theta_c = 0, \quad (2.7.69)$$

$$\dot{\theta}_\nu = k^2 \left( \frac{1}{4} \delta_\nu - \sigma_\nu \right), \quad (2.7.70)$$

for the velocity divergences. The corrective terms in equations (2.7.63-2.7.66) are due to the baryons coupling to photons through Thomson scattering before recombination. The indices  $\gamma$ ,  $b$ ,  $c$  and  $\nu$  correspond to photons, baryons, cold dark matter, and relativistic neutrinos, respectively.  $\sigma_T$  is the Thomson cross section,  $n_e$  is the electron number density, and  $\bar{\rho}_\gamma$  and  $\bar{\rho}_b$  the density of photons and baryons, respectively. After the radiation component decouples from the matter component for a mode within the horizon, there is significant shear stress from the photons and neutrinos, which produces two additional equations for the quadrupole [109]

$$\dot{\sigma}_\gamma = \frac{4}{15}\theta_\gamma - \frac{3}{10}kF_{\gamma 3} + \frac{2}{15}\dot{h} + \frac{4}{5}\dot{\eta} - \frac{9}{10}an_e\sigma_T\sigma_\gamma, \quad (2.7.71)$$

$$\dot{\sigma}_\nu = \frac{4}{15}\theta_\nu - \frac{3}{10}kF_{\nu 3} + \frac{2}{15}\dot{h} + \frac{4}{5}\dot{\eta}, \quad (2.7.72)$$

where  $F_{\nu 3}$  and  $F_{\gamma 3}$  are the third moments of the momentum-averaged phase space densities for neutrinos and photons. The photon shear stress is particularly useful for studying the primary CMB polarisation.

## 2.8 Initial conditions

The standard cosmological model based on the smooth Friedmann-Robertson-Walker universe does not explain the existence of the large structures we see today such as galaxies and galaxy clusters. Gravitational instability can produce these structures but it requires some initial irregularities or primordial fluctuations to act upon, which cannot be explained with the standard model. The idea of inflation [110] was proposed, which provides the linear imprints in the matter and radiation content required for generating the irregularities needed to initiate structure formation during the early universe. Inflation posits that the universe underwent very rapid growth during the first fraction of a second before it settled down to its current rate of expansion. Quantum fluctuations in the energy density of the inflationary field were caught up in the expansion and

stretched to classical scales and these form the primordial fluctuations which seeded structure formation. In what follows, we discuss the types of fluctuations produced beginning with the adiabatic mode.

The adiabatic mode is generated in the simplest inflationary model [111] and produces a non-zero curvature perturbation. The perturbations can be classified in terms of an overall entropy perturbation,  $\phi$  and for the adiabatic mode  $\phi = 0$ . The densities of all species are perturbed in proportion at some initial time  $i$ , such that

$$\delta_{c,i} = \delta_{b,i} = \frac{3}{4}\delta_{\gamma,i} = \frac{3}{4}\delta_{\nu,i}, \quad (2.8.73)$$

or with relative entropy between two species  $x$  and  $y$

$$S_{xy} = \frac{\delta_x}{1+w_x} - \frac{\delta_y}{1+w_y} \quad (2.8.74)$$

for  $w$  being the equation of state for each of the respective species. In practice the entropy is conveniently defined relative to the photon density, i.e.  $y = \gamma$ . This is done because the time dependence of  $S_{x\gamma}$  is naturally captured.

Isocurvature modes on the other hand, correspond to perturbations in entropy with no perturbation in curvature. They are another possibility for initial conditions that are well defined. These isocurvature modes depict ratios of different species that are not spatially constant initially. There are four isocurvature modes: the CDM, baryon, neutrino density, and neutrino velocity isocurvature modes. For the CDM mode, which was first proposed by [112], the ratio of the CDM to photon varies spatially, and fluctuations are imprinted on the CDM component only, hence

$$\delta_{c,i} = 1, \quad \delta_{b,i} = \delta_{\gamma,i} = \delta_{\nu,i} = 0, \quad (2.8.75)$$

with corresponding relative entropies,

$$S_{c\gamma} = \delta_c - \frac{3}{4}\delta_\gamma \approx 1, \quad (2.8.76)$$

and all other entropies equal to zero. In a similar way we have for the baryon mode [113]

$$\delta_{c,i} = 0, \quad \delta_{b,i} = 1, \quad \delta_{\gamma,i} = \delta_{\nu,i} = 0, \quad (2.8.77)$$

with

$$S_{b\gamma} = \delta_b - \frac{3}{4}\delta_\gamma \approx 1. \quad (2.8.78)$$

The neutrino density and velocity isocurvature modes were introduced in [114] and are characterised to have no curvature and matter components but an initial perturbation in the neutrino density is balanced by the photons. The neutrino to photon ratio is initially spatially perturbed but the total density perturbation vanishes. The initial perturbation of each species is given by

$$\delta_{c,i} = \delta_{b,i} = 0, \quad \delta_{\gamma,i} = -\frac{R_\nu}{R_\gamma}\delta_{\nu,i}, \quad (2.8.79)$$

with  $R_\nu$  and  $R_\gamma$  representing the fractional contributions from the neutrino and photons respectively. Relative entropies are given by

$$\begin{aligned} S_{\nu\gamma} &= \frac{3}{4}\delta_\nu - \frac{3}{4}\delta_\gamma \\ &= -\frac{3}{4}\delta_\gamma \left(1 + \frac{R_\gamma}{R_\nu}\right) \end{aligned} \quad (2.8.80)$$

with all other entropies equal to zero. For the neutrino velocity isocurvature mode, the neutrino velocity divergence is initially perturbed but balanced by the baryon-photon velocity. The initial perturbation for the neutrino velocity mode is given by

$$\theta_{c,i} = 0, \quad \theta_{b,i} = \theta_{\gamma,i} = -\frac{R_\nu}{R_\gamma}\theta_{\nu,i}. \quad (2.8.81)$$

## 2.9 CMB anisotropies

The CMB power spectrum measures the temperature anisotropies over a range of angular scales. The temperature variations are small enough that the early universe appears homogeneous during

the formation of the CMB. One of the useful properties of the CMB is its ability to constrain models because of its sensitivity to the cosmological parameters. In practice these dependencies can be studied with the aid of publicly available software such as CMBFAST [115] or CAMB [28]. Instead of angular scale  $\theta$ , the anisotropies are measured as a function of multipole which relates to angular scales as follows  $\ell \approx \frac{180^\circ}{\theta}$ , therefore large angular scales correspond to small  $\ell$  and vice versa. There are statistics in place that enable us to extract the relevant information from the CMB, which we discuss briefly in what follows.

### 2.9.1 Angular power spectrum $C_\ell$

The temperature distribution across the sky is the fundamental quantity that we measure for the CMB. If we denote the temperature over the sky as  $T(\theta, \phi)$ , with spherical angular coordinates on the celestial sphere  $\theta, \phi$ , then the dimensionless temperature anisotropy is given by

$$\frac{\Delta T}{T}(\theta, \phi) = \frac{T(\theta, \phi) - \bar{T}}{\bar{T}}, \quad (2.9.82)$$

with mean temperature of the CMB photons,  $\bar{T}$ . In order to simplify the mathematics, spherical harmonics  $Y_m^\ell$ , which is the analogue of a Fourier series for the surface of a sphere is used instead such that

$$\frac{\Delta T}{T}(\theta, \phi) = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_m^\ell(\theta, \phi). \quad (2.9.83)$$

The coefficient  $a_{\ell m}$  is in Fourier space and is the amplitude of the irregularities on different scales, and hence quantifies the anisotropies. The  $a_{\ell m}$  generally follows a Gaussian probability distribution with mean zero and variance dependent on  $\ell$ . This brings us to an important statistic which is the radiation angular spectrum  $C_\ell$ , given by

$$C_\ell = \langle |a_{\ell m}|^2 \rangle, \quad (2.9.84)$$

averaged over all possible universes or equivalently all observers in our universe. The  $C_\ell$  depends only on  $\ell$ . The difference between our region of the universe compared to the average region of the universe is known as *cosmic variance* and is quite significant for large angular scales. The most important regions for cosmological parameter estimation is  $\ell \geq 2$  because the dipole moment ( $\ell = 1$ ) is believed to be caused mainly by the relative motion of the earth with respect to the CMB.

Before last-scattering the photons are tightly coupled to the baryons, but there is still some small degree of movement of the photons through the baryon-photon fluid. The fluid is oscillating with constant compression and rarefaction under the combined influence of gravitational collapse and pressure, and calculating the anisotropies involves these physical processes. When the photons decouple from the fluid, the baryons are able to gravitationally collapse leading to the subsequent formation of galaxies and galaxy clusters whereas the photons begin to travel freely towards us. The distribution of photons is retained as they free stream and this reflects the fluctuations in the baryon-photon fluid at the time. This manifests as acoustic characteristics in the power spectrum of the CMB.

There are regions of the CMB spectrum that behave differently and can be explained by different physical processes. Large scales for  $\ell \leq 15$  were first probed by the COBE satellite and is known as the Sachs-Wolfe plateau, which is caused by the gravitational potential between regions. The first peak at  $\ell \approx 200$  was confirmed by BOOMERANG [116], an earth based balloon experiment designed to detect the CMB photons from the upper stratosphere. Experiments on earth have limited sky coverage and also limited by the atmosphere, thereby motivating the Wilkinson Microwave Anisotropy Probe (WMAP) satellite [91] and Planck Surveyor [3, 4] projects.

## 2.10 Effect of cosmological parameters on the CMB temperature spectrum

The CMB anisotropy is sensitive to changes to the cosmological parameters, such as the matter density  $\Omega_m h^2$ , which consists of baryon and cold dark matter density contributions. Therefore changes to  $\Omega_b h^2$  and  $\Omega_c h^2$  result in changes to the temperature spectrum. Apart from these parameters, the cosmological constant density  $\Omega_\Lambda$ , the optical depth to reionisation  $\tau$ , the scalar spectral index  $n_s$ , and scalar amplitude  $A_s$  also affect the CMB anisotropy spectrum. In this section we briefly explain the physical significance of these parameters and their effect on the CMB anisotropy spectrum. The significance of these effects is that it can be used to compare theory with observation enabling us to constrain cosmological models and obtain estimates for the cosmological parameters, by finding best-fit likelihood estimates with their corresponding uncertainty.

### 2.10.1 Baryon density $\Omega_b h^2$

The baryon density influences the propagation of the acoustic waves in the baryon-photon fluid before last scattering. Therefore, if there are more baryons in the fluid, the sound speed of the waves decreases, resulting in changes to the location of peaks and troughs in the CMB spectrum. Odd numbered peaks are enhanced while even peaks are lowered if  $\Omega_b h^2$  increases. Figure 2.3 illustrates the baryon density effect on the  $C_\ell$ s.

### 2.10.2 Cold dark matter density $\Omega_c h^2$

The total matter density is made up of both the baryon and cold dark matter density contributions, i.e.  $\Omega_m h^2 = \Omega_b h^2 + \Omega_c h^2$ , and since the CMB spectrum is sensitive to the matter density, we can keep the baryon density constant and vary  $\Omega_c h^2$  to trace its effect on the spectrum. Increasing the

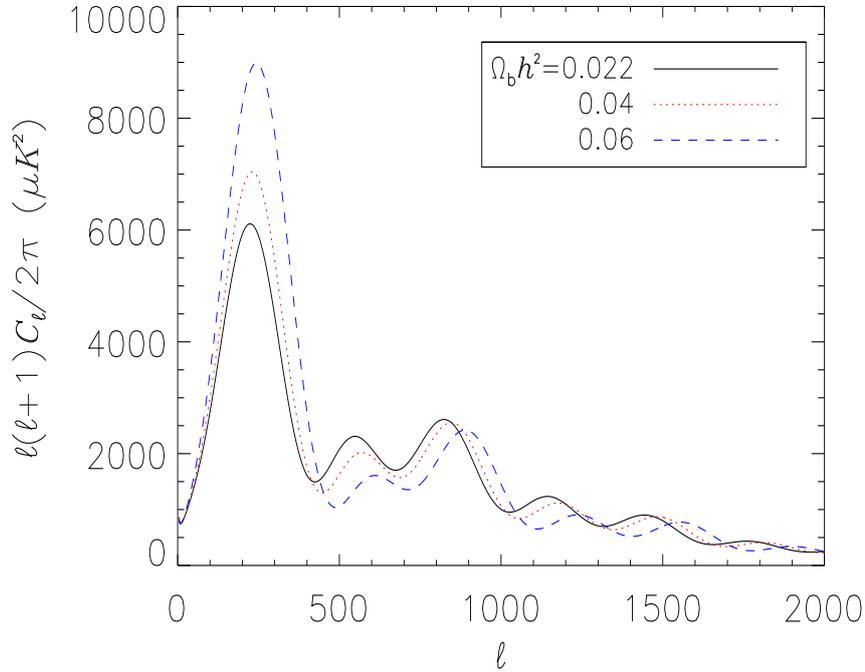


Figure 2.3: The effects on the temperature angular power spectrum for various baryon densities  $\Omega_b h^2$ .

CDM density slightly suppresses the entire angular spectrum, with the first two peaks decreasing by a large factor. The third and following peaks however are enhanced with respect to the first two peaks. From equations (2.1.9) and (2.1.16), we know that increasing the matter density will reduce the age of the universe bringing the epoch of matter-radiation equality  $\tau_{eq}$  closer to recombination. For a lower matter density, there is less occurrence of gravitational instability at recombination and the gravitational potential is reduced. This implies that radiative pressure is higher therefore increasing oscillations, inducing photon-density perturbations that are larger than in a universe with higher matter density.

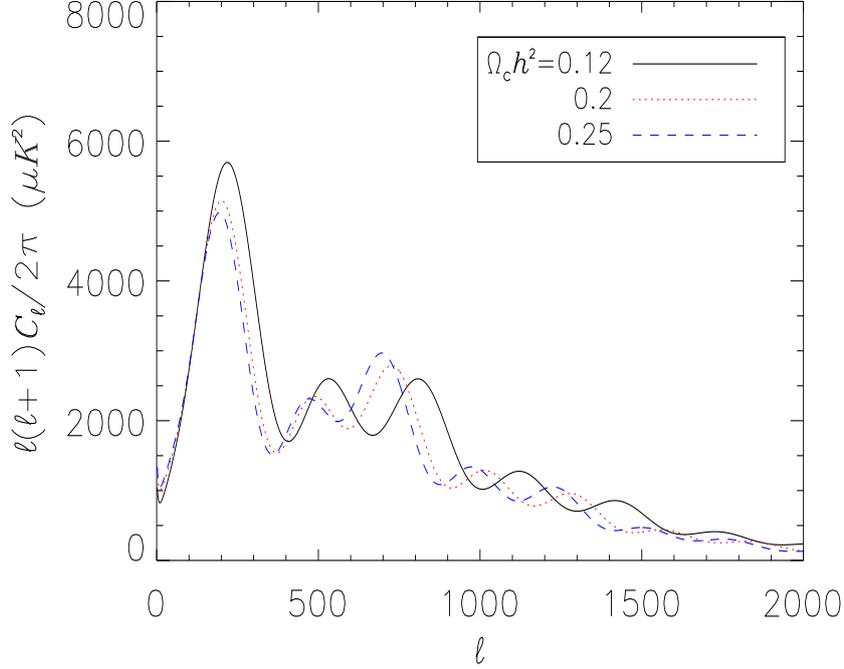


Figure 2.4: Sensitivity of the angular power spectrum to the CDM density  $\Omega_c h^2$ .

### 2.10.3 Cosmological constant density $\Omega_\Lambda$

The angular diameter distance,  $d_A$ , is a useful way of measuring distances based on the angle subtended by an object,  $\varphi$ , of known size,  $d$ , such that  $d_A = d/\varphi$ . The cosmological constant and curvature only affect the CMB spectrum through the angular diameter distance and late-time integrated Sachs-Wolfe (ISW) effect [117]. In our study the dark energy is restricted to a cosmological constant. Therefore, dark energy contributed negligibly during last scattering, and has affected only the photons free-streaming towards us since then. For a flat universe  $\Omega_K \equiv 1 - \Omega_\Lambda - \Omega_m = 0$  and  $\varphi = \frac{d}{a}/\Delta\tau$  for comoving distance  $\Delta\tau$ . We therefore have [108]

$$d_A^{flat} = a\Delta\tau = \frac{\Delta\tau}{1+z}, \quad a_0 = 1. \quad (2.10.85)$$

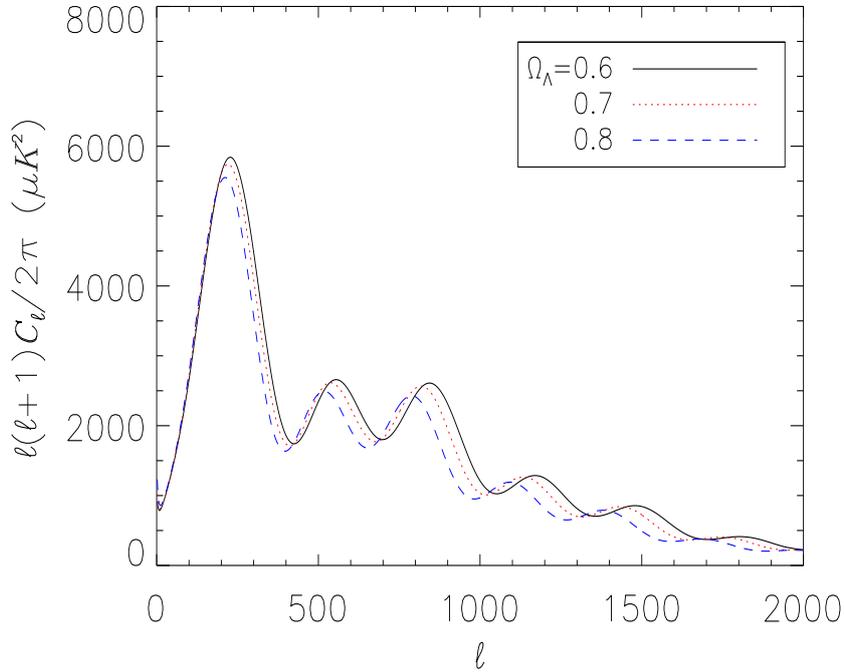


Figure 2.5: The angular power spectrum for various dark matter densities  $\Omega_\Lambda$ .

The CMB spectrum is shifted to lower  $\ell$  for higher  $\Omega_\Lambda$ , since at low  $z$  the angular diameter distance reduces to the comoving distance.

#### 2.10.4 Optical depth $\tau$

After last scattering the photons from the CMB have been free-streaming towards us, but at some point in time later the photons once again interacted with the free electrons through Thomson-scattering. This second interaction erases information carried by the photons about the anisotropies and causes a dampening in the CMB temperature by a factor of  $e^\tau$ , where  $\tau$  is the optical depth to reionisation. As a result of this temperature reduction, the  $C_\ell$  is also reduced by  $e^{-2\tau}$ . Therefore, all modes within the horizon at reionisation are affected by the second photon-electron interaction, while modes outside are not. Figure 2.6 shows the sensitivity of the temperature angular

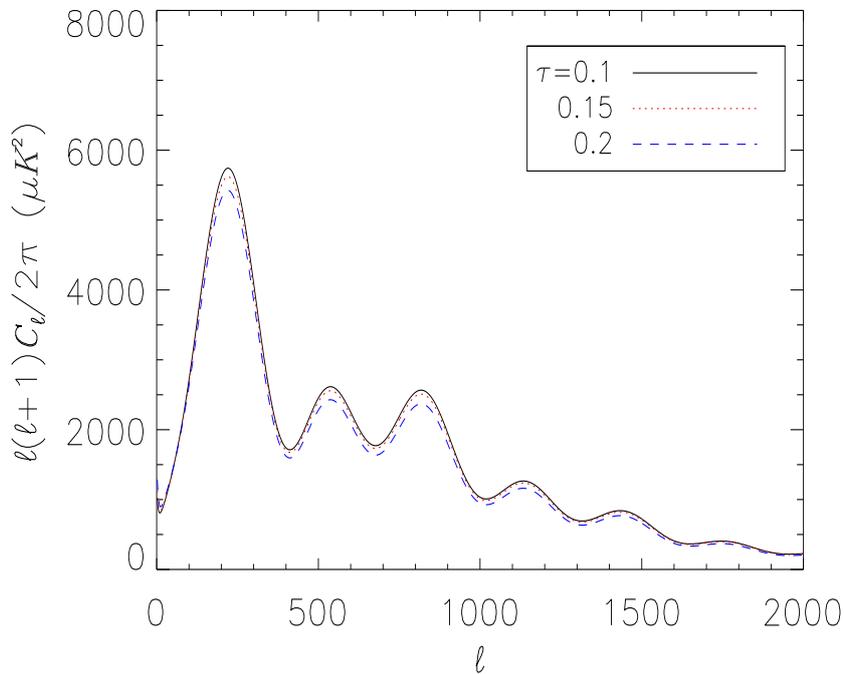


Figure 2.6: Small shifts of the angular power spectrum for various  $\tau$ .

spectrum to  $\tau$ .

### 2.10.5 The spectral index $n_s$

The primordial fluctuations are typically quantified by a power spectrum, which gives the power of the variations as a function of spatial scale. If we denote the temperature perturbation at reionization by  $\Theta$ , then we can write the CMB angular power spectrum as

$$C_\ell = \frac{2}{\pi} \int_0^\infty k^2 P(k) |\Theta_\ell(k)|^2 dk, \quad (2.10.86)$$

where  $P(k)$  is the primordial power spectrum for wavenumber  $k$ . Many inflationary models predict that the scalar component of the fluctuations obeys a power-law

$$P(k) \propto k^{n_s-1} \quad (2.10.87)$$

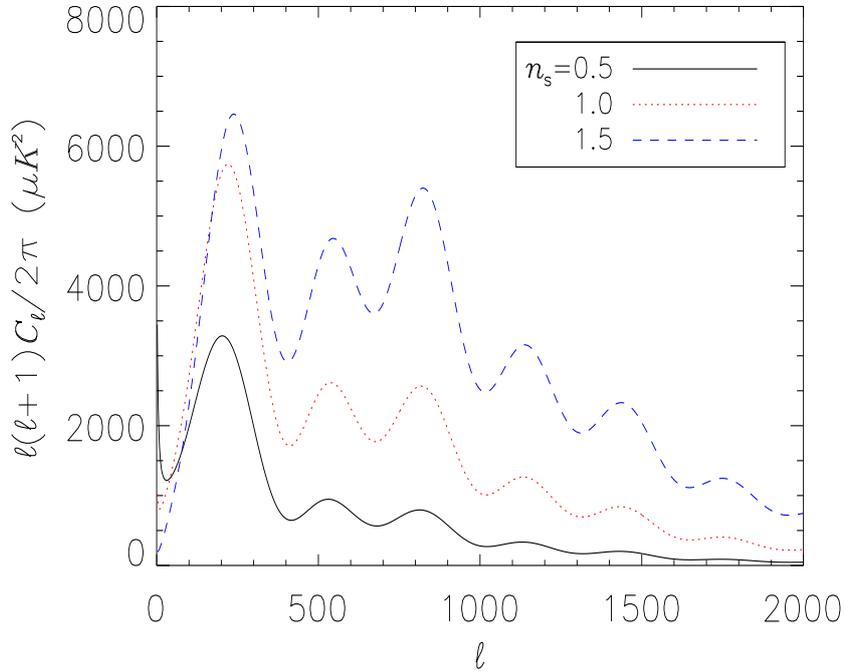


Figure 2.7: Large shifts of the angular power spectrum for various  $n_s$ .

where  $n_s$  is known as the scalar spectral index. From equation (2.10.87) we can see that the slope of the  $C_\ell$  is controlled by  $n_s$ , with  $n_s < 1$  lowering small-scale anisotropy and  $n_s > 1$  producing the opposite effect. The special case that  $n_s = 1$  corresponds to scale-invariant fluctuations. Figure 2.7 illustrates large variations for  $n_s$ .

### 2.10.6 The scalar amplitude $A_s$

The comoving curvature perturbations, denoted by  $\mathcal{R}$  vary with scale, and for an inflationary model we can expand  $\ln \mathcal{R}(\ln k)$  in a Taylor series whose terms decrease in size. For the simplest inflationary model, the initial condition for density perturbations can be described by the amplitude and slope of the power spectrum  $\langle |\mathcal{R}|^2 \rangle \propto k^n$  defined,

$$\Delta_{\mathcal{R}}^2 \equiv (k^3/2\pi^2) \langle |\mathcal{R}|^2 \rangle. \quad (2.10.88)$$

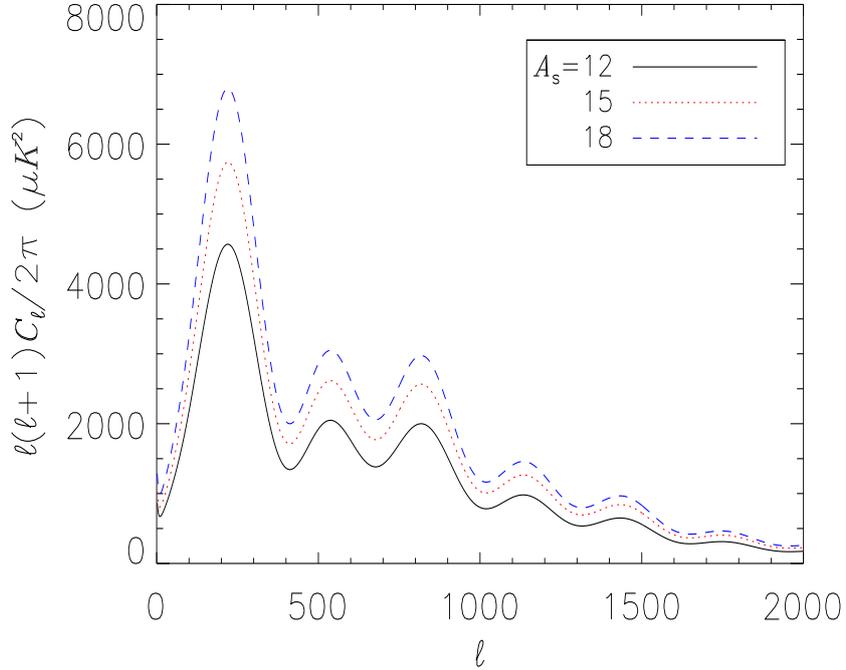


Figure 2.8: Shifts of the angular power spectrum for various  $A_s$ .

The scalar amplitude is given by

$$A_s^2 \equiv \Delta_{\mathcal{R}}^2(k_0), \quad (2.10.89)$$

with  $k_0 = 0.05 \text{Mpc}^{-1}$ . There are many definitions of  $A_s$  which can be cosmology dependent [118, 119]. For inflationary potential  $V(\phi)$ , the slope  $n$  involves  $V^n$  therefore  $A_s$  and  $n$  can in principle constrain potentials. The effect of the scalar amplitude on the angular power spectrum is shown in 2.8.

## 2.11 CMB dependence on initial conditions

In Section 2.8 we highlighted and briefly discussed the density contrasts for the baryons and photons as well as the baryon-photon common velocity divergence and gravitational potential

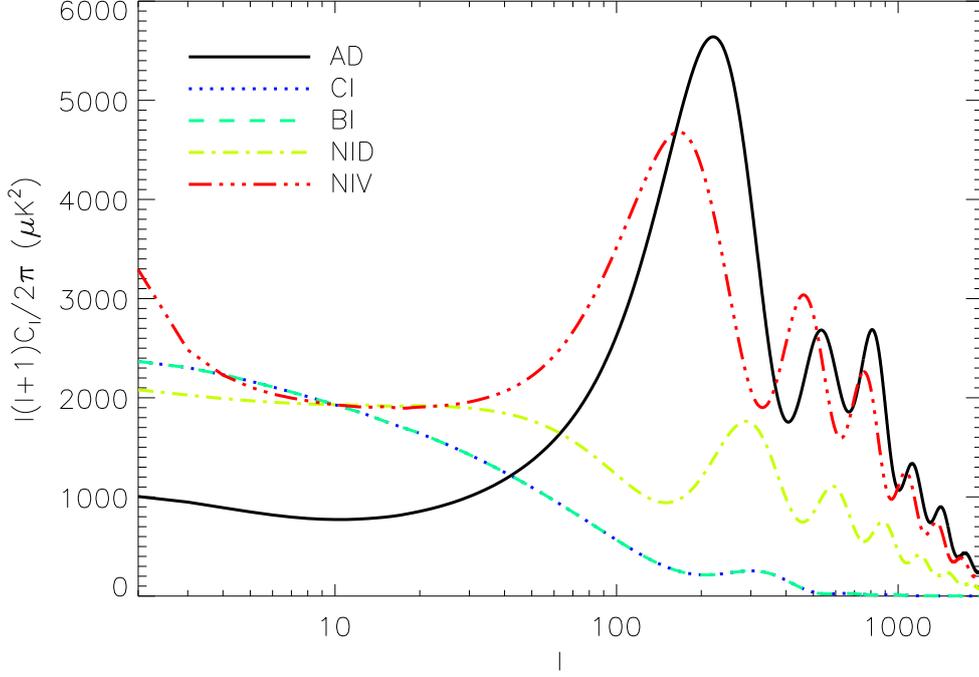


Figure 2.9: Comparison of CMB temperature power spectra for AD, CI, BI, NID and NIV modes for the same cosmological model. Spectra for all modes have a flat  $\Lambda$ CDM universe with  $\Omega_b h^2 = 0.022$ ,  $\Omega_c h^2 = 0.125$  and  $h = 0.7$ . The isocurvature spectra are normalized to have the same power as the adiabatic spectrum at  $\ell = 10$ . These spectra are obtained using CAMB. Figure courtesy of [120].

evolution during the tight-coupling regime. We know that the  $C_\ell$ s were dependent on the multipole moments  $\Theta_\ell(\tau, k)$ , which by the line of sight integral approach are given by

$$\Theta_\ell(k, \tau_0) = \int_0^{\tau_0} S(k, \tau) j_\ell[k(\tau - \tau_0)] d\tau, \quad (2.11.90)$$

where  $j_\ell$  is a spherical Bessel function [121]. Assuming the contribution to the source function  $S(k, \tau)$  due to the polarization is small, we conclude that

$$S = e^{-\tau_e} (\dot{\eta} + \ddot{\alpha}) + g(\tau) \left[ \frac{\delta_\gamma}{4} + \frac{\dot{\theta}_b}{k^2} + 2\dot{\alpha} \right] + \dot{g}(\tau) \left[ \frac{\theta_b}{k^2} + \alpha \right], \quad (2.11.91)$$

where  $g(\tau)$  is the visibility function, and

$$\alpha = -\frac{R_\nu}{15 + 4R_\nu}\tau. \quad (2.11.92)$$

Each mode has a different source function due to the differing behaviour of photon perturbations resulting from the distinct initial conditions. Each mode, therefore, has a characteristic contribution to the CMB angular spectrum, enabling us to fit the best-spectrum for a given set of data, thus discriminating between adiabatic, isocurvature, and mixed adiabatic isocurvature models.

Figure 2.9 illustrates the CMB temperature power spectra for the cold dark matter (CI), neutrino velocity (NIV), neutrino density (NID), and baryonic (BI) mode for the same cosmological model. We include the adiabatic (AD) mode spectrum for comparison against the isocurvature mode spectra, which are normalised to have the same power at  $\ell = 10$ , such that  $C_{10}^{\text{ISO}} = 2.5C_{10}^{\text{AD}}$ . Each mode has clearly distinctive features regarding positions and amplitudes of the acoustic peaks, the height of the Sachs-Wolfe plateau, and the steepness of the spectrum on small scales.

The spectrum of the NIV mode is the most similar, regarding shape, to the AD spectrum as compared to the other isocurvature modes. Both spectra illustrate virtually the same Sachs-Wolfe plateau and ISW effect. The NIV spectrum is dampened at small scales (high  $\ell$ ) with the acoustic peaks shifted to larger scales (lower  $\ell$ ). This is due to the NIV mode exciting a pure  $\sin kr_s$  harmonic, unlike the driven  $\cos kr_s$  harmonic for the AD mode, where  $r_s$  denotes the sound horizon.

The peaks of the NID spectrum are positioned at the troughs of the NIV mode, since both modes are free oscillations with cosine (NID) and sine (NIV) like phases. The NID mode however has more power at lower  $\ell$  with a trough before the first peak. The first peak and Sachs-Wolfe plateau have almost the same height.

The CI and BI mode have the same CMB spectrum which is indistinguishable since their photon evolution equations are similar, therefore we choose to discuss only the CI mode since

this mode is considered in our mixed adiabatic with isocurvature mode models in Chapter 5. The CI mode excites a  $\sin kr_s$  like harmonic therefore the peak positions are close to the troughs of the AD spectrum. The CI spectrum decreases at low  $\ell$  and does not have a Sachs-Wolfe plateau.

In the next Section we parametrise the isocurvature and adiabatic contributions so that we can constrain these contributions using the effects of the isocurvature mode features in the CMB spectrum. In our approach we determine the isocurvature fraction because the sub-dominant contribution impacts studies of other parameters, such as dark energy [122].

## 2.12 Isocurvature mode parametrisation

The standard cosmological model depicts a spatially flat universe containing baryons, CDM, photons, neutrinos, and a cosmological constant  $\Lambda$ ; with a primordial spectrum of Gaussian adiabatic density perturbations, described by a monomial power law [41, 91, 123, 124]. This is consistent with current CMB anisotropy and LSS data. However, models with an appreciable isocurvature component may also explain the data sufficiently.

Predictions of the CMB anisotropies and LSS data may be derived entirely from the perturbations of the above constituents. Therefore, models with a significant isocurvature component may explain the data just as well as a model with a purely adiabatic mode. If the ratios of density contrasts are not constant, the fluctuations are said to be generated from isocurvature initial conditions [114]; therefore, it is worthwhile to determine whether models having an isocurvature component can also account for the current data. Doing this entails providing constraints on models which in addition to the adiabatic mode, include either one, two, or three isocurvature modes as well as cross-correlations between the allowed modes.

There are four types of isocurvature perturbations as mentioned in Section 2.8, namely the baryon, CDM (CI), neutrino-density (NID), and neutrino-velocity (NIV). The baryon and CDM

isocurvature modes imprint almost the same perturbations on the CMB, therefore we do not consider them separately. Instead we consider only CI, NID, and NIV modes. Considering only quadratic observables, perturbations in all modes, including the adiabatic (AD) mode, can be completely characterised by a matrix valued power spectrum [114, 125]

$$\langle X_I(\mathbf{k})X_J^*(\mathbf{k}') \rangle = A_{IJ}(|\mathbf{k}|)\delta^2(\mathbf{k} - \mathbf{k}'), \quad (2.12.93)$$

where the indices  $I, J = 1, 2, 3, 4$  label the modes AD, CI, NID, NIV respectively, and  $X_I(\mathbf{k})$  represents the amplitude of the  $I$ th mode with wavenumber  $\mathbf{k}$ . The overall  $C_\ell$  is a sum over normalised mode spectra with amplitudes  $A_{IJ}$  such that

$$C_\ell = \sum_{IJ} A_{IJ} C_\ell^{IJ}. \quad (2.12.94)$$

Auto-correlations are represented by the diagonals, such that  $A_{IJ}$  is always positive definite, with cross-correlations represented by off-diagonal elements.

Previous work in providing constraints for adiabatic and mixed adiabatic with isocurvature modes has been done in [71, 80], using the popular MCMC sampling algorithm. We closely follow the parametrization and methodology of their work and implement the optimised PMC algorithm. A full analysis of all four modes is studied in this thesis using the WMAP nine-year released data.

Various models have been proposed that produce a mixture of adiabatic and isocurvature perturbations. One possibility is that during inflation, there exists a calculable field of mass that is relatively small to the Hubble scale, giving significant power on large scales. Hence this field may be used to establish various initial conditions [126, 127]. Another model explains that inflation is driven by two scalar fields [125, 128, 129]. Perturbations in the direction of motion of the dominant field correspond to the adiabatic perturbations. While the orthogonal perturbations correspond to the isocurvature perturbations [130]. All other trajectories generate

correlated adiabatic isocurvature perturbations. The amplitudes for auto- and cross-correlations are specified by the respective model. These are two such possibilities for models that generate primordial perturbations. Other possibilities are discussed in [114].

We use a covariance matrix in equation (2.12.93) which provides a complete description of the primordial perturbations for Gaussian initial conditions, with the amplitudes represented by  $A_{IJ}$ . To ensure the matrix is nonnegative definite, the cross correlation spectral indices are set to the mean of the corresponding auto-correlation spectral indices. We define a random variable  $z$ , to represent the contribution from each mode to the power spectrum such that

$$A_{IJ} \propto z_{IJ} \quad (2.12.95)$$

where

$$\text{tr}(zz^T) = \sum_{I,J=1}^N z_{IJ}^2 = 1. \quad (2.12.96)$$

The symmetric matrix  $z$  contains all information regarding fractional contributions by the auto- and cross-correlations, therefore, there are only  $D = N(N + 1)/2$  independent correlations and the coefficients  $z_{IJ}$  cover the unit sphere of dimension  $d = D - 1$ , on which we use the measure corresponding to the usual volume element. The transformation from  $w_i$  to the  $z_{IJ}$  parameters is accomplished by mapping the surface of the  $d$ -dimensional sphere,  $S^d$  into a  $d$ -dimensional ball  $B^d$ , using the volume preserving mapping

$$r(\theta) = \left[ d \int_0^\theta d\bar{\theta} \sin^{(d-1)} \bar{\theta} \right]^{1/d}, \quad (2.12.97)$$

where  $\theta$  is the angle with respect to the north pole of  $S^d$ , the point corresponding to a purely adiabatic model.  $r$  is the radial coordinate of  $B^d$ . In order to transform the exploration of  $B^d$  into the sphere  $S^d$  we have

$$\begin{aligned}
z_{11} = \cos \theta, \quad z_{12} &= \frac{\sin \theta}{\sqrt{2}} \frac{w_1}{r}, \dots, \quad z_{1N} = \frac{\sin \theta}{\sqrt{2}} \frac{w_{N-1}}{r}, \\
z_{22} &= \sin \theta \frac{w_N}{r}, \dots, \quad z_{2N} = \frac{\sin \theta}{\sqrt{2}} \frac{w_{2N-2}}{r}, \\
&\vdots \\
z_{NN} &= \sin \theta \frac{w_d}{r},
\end{aligned}$$

for symmetric  $z_{IJ}$ , where  $w_i$  are the Cartesian coordinates of the Euclidean space into which  $B^d$  has been embedded with  $r = |\mathbf{w}|$ , and  $\theta = \theta(w)$ . Following the framework of Bucher *et al.* [114] we define the non-adiabatic fraction as

$$f_{\text{ISO}} = \frac{z_{\text{ISO}}}{z_{\text{ISO}} + z_{\langle \text{AD}, \text{AD} \rangle}} \quad (2.12.98)$$

with  $z_{\text{ISO}}$  representing the fraction of isocurvature contribution to the data given by

$$z_{\text{ISO}} = \sqrt{1 - z_{\langle \text{AD}, \text{AD} \rangle}^2} \quad (2.12.99)$$

for both auto- and cross-correlations. This is the parametrisation that we will use when constraining adiabatic and isocurvature modes in Chapter 5. In Chapter 5 we briefly discuss the cosmological parameters that we wish to constrain using PMC and their effect on the CMB spectrum. There are two cases to which we apply the PMC algorithm, namely, the pure adiabatic model and mixed models with adiabatic and isocurvature modes. In both these cases we apply the PMC algorithm, discussing the results and possible challenges in sampling the target parameter space. However, we must first examine the Monte Carlo algorithms used for sampling, before proceeding to the cosmological parameter estimation study.

## CHAPTER 3

---

### Monte Carlo sampling methods

---

#### 3.1 The general problem

In order to quantify our beliefs in certain propositions and events, we assign a real number to that degree of belief, which is proportional to how strong that belief is. This number is referred to as a probability [131]. For a particular event,  $A$ , we denote the probability of that event given some background information  $I$  by  $P(A|I)$ . To ensure consistency, rules of probabilities are in place, such that for events  $A$  and  $B$

$$P(A|I) + P(\bar{A}|I) = 1 \quad (3.1.1)$$

and

$$P(A, B|I) = P(A|B, I) \times P(B|I) \quad (3.1.2)$$

where  $\bar{A}$  is the event that  $A$  does not occur, and  $P(A, B|I)$  is the joint probability of events  $A$  and  $B$ . Using this foundation we have *Bayes' theorem* [17],

$$P(A|B, I) = \frac{P(B|A, I)P(A|I)}{P(B|I)}. \quad (3.1.3)$$

By replacing  $A$  with a set of parameters  $x$ , and  $B$  with data set  $D$ , we have:

$$\pi(x|D) = \frac{P(D|x)P(x)}{P(D)}. \quad (3.1.4)$$

The term  $\pi(x|D)$ , is known as the *posterior* distribution of the model parameters and  $P(D|x)$  and  $P(x)$  are known as the *likelihood* and *prior* probability respectively. The denominator in equation (3.1.4) is called the *evidence* and is relevant to model selection but not to parameter estimation, therefore we do not consider it in this study. We omit the notation for dependence on data for convenience and write  $\pi(x)$  instead throughout this thesis. The prior distribution encodes all knowledge of  $x$  from previous experiments, whereas the likelihood is the probability that we obtained the data had the set of parameters,  $x$ , been chosen. We can thereafter provide inference on  $\pi(x)$  using the likelihood. It is generally difficult to evaluate the posterior directly since we most often lack an analytical expression for it, therefore we indirectly use the likelihood to evaluate the posterior since it only depends on the given data set. We rely on using a sample from this distribution to evaluate any integrals related with  $\pi(x)$ . That is,

$$\tau = \int f(x)\pi(x)dx \quad (3.1.5)$$

can be approximated empirically by,

$$\tau \approx \frac{1}{N_s} \sum_{i=1}^{N_s} f(x_i) \quad (3.1.6)$$

where  $(x_1, x_2, \dots, x_{N_s})$  is a sample of size  $N_s$  drawn from  $\pi(x)$ . A sampling algorithm is thus required for this process.

## 3.2 The Markov Chain Monte Carlo algorithm

It was not until the introduction of computers that simulations of random processes such as Monte Carlo were widely used to solve mathematical problems involving integration, optimisation, stochastic exploration, to name a few. Very often in statistics, the Monte Carlo process was

used to solve expectations of certain quantities with respect to a probability density function, specifically for a probability distribution function  $p$  of observation  $x$ , we have the expectation of some function  $f$  given by the integral

$$I = \int_{\Omega} p(x) f(x) dx, \quad (3.2.7)$$

where  $\Omega$  is the real space of  $p$ . The Monte Carlo algorithm works by drawing random numbers  $X_1, X_2, \dots, X_n$  where  $X_j$  are random variables drawn from density  $p$ , i.e.,  $X_j \sim p$  for  $j = 1, 2, \dots, n$ . A Monte Carlo estimator of equation (3.2.7) is then given by

$$\hat{I}_n = \frac{1}{n} \sum_{j=1}^n f(X_j). \quad (3.2.8)$$

Thus we have  $\hat{I}_n \rightarrow I = E_p[f(X)] = \int_{\Omega} p(X) f(X) dX$  for  $n \rightarrow \infty$ .

To quantify the error in our estimation, we use the central limit theorem (CLT). Let  $X_1, X_2, \dots, X_n$  for independent identically distributed (i.i.d.) random variables with  $E(X_i) < \infty$  and variance  $\text{Var}(X_i) = \sigma^2 < \infty$  for all  $i = 1, 2, \dots, n$ . The central limit theorem states that if for a population mean  $\mu$  we have

$$Z_n = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sigma/\sqrt{n}}, \quad (3.2.9)$$

then  $Z_n$  converges to a standard normal variable, i.e.  $Z_{\infty} \sim N(0, 1)$  which is equivalent to

$$p(Z_n) = \frac{1}{\sqrt{2\pi}} e^{-Z_n^2/2}, \quad (3.2.10)$$

for  $n \rightarrow \infty$ . Since we do not know the population mean  $\mu$ , we can estimate it from a sample

$$\hat{\mu} \approx \frac{1}{n} \sum_{i=1}^n X_n + \xi \sigma / \sqrt{n}, \quad (3.2.11)$$

where  $\xi \sim N(0, 1)$ . The second term in equation (3.2.11) represents the error in the approximation. We are therefore able to determine the error if we can estimate the standard deviation

$\sigma$  of the sample, enabling us to calculate confidence intervals for  $\mu$ . In practice the maximum likelihood estimate is used, i.e.,

$$\sigma_{\text{ML}}^2 = \frac{1}{n-1} \sum_j^n (X_j - \bar{X}_n)^2. \quad (3.2.12)$$

This illustrates that the error is inversely proportional to the square root of the sample size; hence, to decrease the error by a factor of ten say, we need to increase the sample size by a factor of a hundred.

### 3.2.1 Markov Chains

The ‘‘classical Monte Carlo’’ algorithm generates an independent series of random numbers, which are specified by a target distribution  $\pi$ . This can however be difficult for high dimensional problems or complicated target distributions. Markov chains are designed to overcome this problem, specifically the Markov Chain Monte Carlo (MCMC) algorithm, which was introduced in 1953 [27] to simulate a liquid in equilibrium with its gas phase. It was only after 1990 that MCMC became popular amongst statisticians.

Markov chains are a series of random variables  $X_1, X_2, \dots$  that are assumed to have a finite number of states  $1, 2, \dots, T$ , i.e.,  $X_i$  can only assume values  $S_1^1, S_2^2, \dots, S_T^T$ . In a Markov chain given a set of states  $\{i\}$ , the probability of the next state depends only on the current state which is equivalent to

$$P(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_0 = i_0) = P(X_n = i_n | X_{n-1} = i_{n-1}), \quad (3.2.13)$$

which is known as a transition probability.

We assume homogeneity with regards to steps, which implies  $P(X_n = i_n | X_{n-1} = i_{n-1})$  does not depend on  $n$ . We assume there are  $d$  states such that for each  $j = 1, 2, \dots, d$  state,

$$\pi_j^{(k)} = P(X_k = j) \quad (3.2.14)$$

is the probability to find the chain in state  $j$  at step  $k$ . From the laws of probability,

$$\pi_j^{(k)} \geq 0, \quad (3.2.15)$$

$$\sum_{j=1}^d \pi_j^k = 1 \quad (3.2.16)$$

therefore

$$\pi_l^{(k+1)} = \sum_{j=1}^d P(X_{k+1} = l | X_k = j) \pi_j^{(k)} \quad (3.2.17)$$

$$= (\pi^{(k)} P)_l, \quad (3.2.18)$$

where  $P = (P(X_{k+1} = l | X_k = j))$  is a matrix of transition probabilities. Hence, for any state in general

$$\pi^{k+1} = \pi^k P, \quad (3.2.19)$$

which is known as the Chapman-Kolmogorov equation which relates the joint probability distributions of different sets of coordinates on a stochastic process.

The distribution  $\pi$  is called a stationary distribution if conditions  $\pi_j \geq 0$  and  $\sum_j^d \pi_j = 1$  remain unchanged. Stationary distributions are not necessarily unique, i.e.,

$$\pi = \pi P, \quad (3.2.20)$$

where  $\pi$  is a left eigenvector of  $P$ . Markov chains do not necessarily converge, however we are interested in chains that converge independently of initial distribution  $\pi^0$ . The distribution is stationary when convergence is achieved. For this we need to consider two conditions, viz: *irreducibility* and *periodicity*.

- **Irreducibility** -The state  $j$  is reachable from state  $i$  if there is an  $n > 0$  that satisfies  $P(X_{k+n} = j | X_k = i) > 0$ . A chain is then said to be irreducible if every  $j$  is reachable from every  $i$ .

- **Periodicity** -A state  $j$  is called periodic with period  $r \geq 1$  if  $r$  is the greatest integer such that  $P_{jj}^n = 0$  if  $n$  is not a multiple of  $r$ . A state is called aperiodic if it has  $r = 1$  and a chain is called aperiodic if every state is aperiodic.

A Markov chain is said to be aperiodic if it has one aperiodic state. For a finite state space, an irreducible Markov chain has a unique stationary distribution, and if it is aperiodic,

$$\lim_{k \rightarrow \infty} \pi^{(k)} = \pi, \quad (3.2.21)$$

for all initial distributions. Another property that certain Markov chains have is the property of *recurrence*. The recurrence time or step of state  $i$  given by  $T_i = \min\{n \geq 1 : X_n = i \text{ given } X_0 = i\}$  for random variable  $T_i$ . In a sense, recurrence time is the minimum time it takes to get back to a state. A state  $i$  is called recurrent if  $P(T_i < \infty) = 1$ ; otherwise  $i$  is transient. A recurrent state is called positive recurrent if  $E[T_i] < \infty$ , otherwise it is called null recurrent. A Markov chain is said to be positive recurrent if all states are positive recurrent. These properties lead us to the *ergodic* principle, which states that an ergodic Markov chain has a unique stationary distribution  $\pi$  such that

$$\pi = \pi P, \text{ and} \quad (3.2.22)$$

$$\lim_{k \rightarrow \infty} \pi^{(k)} = \pi \quad \forall \pi^{(0)}. \quad (3.2.23)$$

A chain that is irreducible, aperiodic and positive recurrent is called ergodic. In the next Section we will look at a specific MCMC algorithm that is commonly used in practice.

### 3.2.2 The Metropolis-Hastings Algorithm

The Metropolis algorithm was generalised into a Metropolis-Hastings (MH) algorithm [132], enabling complete Bayesian inference of all kinds. We consider the MH algorithm for a continuous state space. Letting  $t(y|x)$  represent the transition probability density of finding a system

in state  $y$  given it is in state  $x$ , we have a sufficient condition for existence of a unique stationary distribution given by

$$\pi^{(k+1)}(y) = \int_{\Omega} t(y|x)\pi^{(k)}(x)dx. \quad (3.2.24)$$

The goal of the MH algorithm is to construct a Markov chain such that its asymptotic distribution of states is identical to the target distribution, thereby providing a solution to equation (3.1.6). The target distribution  $\pi$  is therefore a stationary distribution. A useful condition for stationarity is the detailed balanced equation (DBC)

$$\pi(x)t(y|x) = \pi(y)t(x|y). \quad (3.2.25)$$

The proof follows as:

$$\int_{\Omega} \pi(x)t(y|x)dx = \int_{\Omega} \pi(y)t(x|y)dx \quad (3.2.26)$$

$$= \pi(y) \int_{\Omega} t(x|y)dx \quad (3.2.27)$$

$$= \pi(y) \quad (3.2.28)$$

with the last line due to  $\int_{\Omega} t(x|y)dx = 1$ . The MH algorithm satisfies this property thus making it a useful tool for sampling the posterior distribution of parameters. The steps of the algorithm is as follows:

- To move from state  $x_k$  to  $x_{k+1}$ , we draw a proposal  $y_k$  from the proposal distribution  $q(y_k|x_k)$ .
- We accept the proposal i.e. set  $x_{k+1} = y_k$  as our new state with probability

$$\rho(y_k, x_k) = \min \left[ \frac{\pi(y_k)}{\pi(x_k)} \times \frac{q(x_k|y_k)}{q(y_k|x_k)}, 1 \right] \quad (3.2.29)$$

- Otherwise, reject the proposal and set  $x_{k+1} = x_k$ .

Note that the Metropolis algorithm has  $q(x|y) = q(y|x) \forall x, y$ , therefore,

$$\rho(y_k, x_k) = \min \left[ \frac{\pi(y_k)}{\pi(x_k)}, 1 \right]. \quad (3.2.30)$$

This means that if the proposed step  $y_k$ , has a higher probability density than  $x_k$ , the proposal will always be accepted. If  $\pi(y_k) = \pi(x_k)$ , the step will be accepted only with probability  $\pi(y_k)/\pi(x_k)$ . The proposal is typically, but not restricted to, a Gaussian distribution with covariance  $C_T$ , that controls the step size for the following chain element. We say the step is accepted if  $\pi(x_*)/\pi(x_n) \geq 1$ ; otherwise it is rejected. The chain is correlated and in the infinitely long chain limit, the chain fully explores the target parameter space spanned by  $\pi$ , with the limiting distribution of the chain equivalent to  $\pi$ ; at which time we say the chain has converged. The proposal distribution should be chosen such that it can be fine-tuned to resemble  $\pi$ . The step size is important because the sampler could miss regions of the target distribution for too large steps. Conversely, for too small steps, the chain could remain stuck at regions for a long time, failing to fully explore the target parameter space within the restricted time limit. It is therefore necessary to carefully consider the covariance. There are monitors and diagnostic tools that assess convergence [35, 133], as well as recommendations for optimal scaling and calibration [134].

Even though it is widely used in cosmological parameter estimation, the MCMC sampler has been scrutinised because it cannot be easily parallelised and has correlations in its final sample set [31, 135]. In this thesis we study an alternate importance sampling algorithm called Population Monte Carlo (PMC), which overcomes these shortfalls.

### 3.3 Population Monte Carlo algorithm

PMC is an iterative sampling algorithm that aims to improve estimates of parameters using the sequence of samples generated after each iteration. It is based on the identity,

$$\begin{aligned}\langle f \rangle &= \int f(x)\pi(x)dx \\ &= \int f(x)\frac{\pi(x)}{q(x)}q(x)dx.\end{aligned}\tag{3.3.31}$$

We use equation (3.3.31) to independently draw a sample  $(x_1, x_2, \dots, x_{N_s})$  from  $q$ , where  $q$  is known as the *importance* function. This function is synonymous with the trial distribution in MCMC. It is an initial guess of the target distribution which adapts iteratively to  $\pi$ . The importance function therefore has the property of a probability density function such that  $q(x_i) \geq 0$  for  $i = 1, 2, \dots, N_s$ . A convergent estimator of equation (3.3.31) is,

$$\begin{aligned}\langle f \rangle &\approx \frac{1}{N_s} \sum_{i=1}^{N_s} f(x_n) \frac{\pi(x_n)}{q(x_n)} \\ &= \frac{1}{N_s} \sum_{i=1}^{N_s} f(x_n) w_n,\end{aligned}\tag{3.3.32}$$

where

$$w_n = \frac{\pi(x_n)}{q(x_n)},\tag{3.3.33}$$

are known as the importance weights. The normalised importance weights are defined as,

$$\bar{w}_n = \frac{w_n}{\sum_{i=1}^{N_s} w_i}.\tag{3.3.34}$$

Using this result we obtain the self-normalised importance ratio,

$$\langle f_N \rangle \approx \frac{1}{N_s} \sum_{i=1}^{N_s} f(x_n) \bar{w}_n,\tag{3.3.35}$$

which is independent of the normalisation of  $\pi(x)$ .

The estimation is done via an adaptive method using importance sampling. Sample sets are generated at each iteration  $t = 1, 2, \dots, T$ , from respective importance functions,  $(q^1, q^2, \dots, q^T)$ . The importance weights at each iteration,  $t$ , are computed by,

$$w_n^t = \frac{\pi(x_n^t)}{q^t(x_n^t)}; \quad n = 1, \dots, N_s, \quad (3.3.36)$$

with the normalised importance weights obtained from equation (3.3.34). These weights form the basis for the updating procedure as all parameters of the importance function are updated accordingly to  $\bar{w}_n$ . This process continues at each iteration, at which a measure is in place to signal to the sampler if it should stop or continue. This measurement is based on the closeness between  $q$  and  $\pi(x)$ .

We use the *Kullback Liebler* divergence [136] as a measure of the fit between  $q$  and  $\pi(x)$ , given by,

$$K(\pi||q^t) = \int \ln \left( \frac{\pi(x)}{q^t(x)} \right) \pi(x) dx. \quad (3.3.37)$$

This measure is adjusted accordingly with each iteration of  $q^t$ , so as to minimise  $K$ . To enable  $q$  to closely match  $\pi(x)$  and ensure the computation of  $K$  is feasible, a linear sum of a number of mixture densities or components for  $q$  are suggested [137] instead of an explicit form. That is,

$$q^t(x^t, \alpha^t, \theta^t) = \sum_{d=1}^{N_c} \alpha_d^t \phi_d(x^t, \theta_d^t), \quad (3.3.38)$$

where  $\alpha^t = (\alpha_1^t, \alpha_2^t, \dots, \alpha_{N_c}^t)$  are the component weights associated with the sample size chosen from each respective component,  $\phi_d$ , for  $N_c$  components. These are constrained such that  $\sum_{d=1}^{N_c} \alpha_d = 1$ . Each  $\phi_d$  is therefore dependent on a parameter set  $\theta_d$ . It is suggested [31], that the  $\phi_d$  distributions chosen are either multivariate Gaussian or Student-t distributions with the latter chosen in the case that  $\pi(x)$  is suspected to have heavy tails. In this thesis we find t-distributions

to be less efficient in general than Gaussian distributions which proved to be sufficient for our studies. We therefore have components of dimension  $D$  such that,

$$\begin{aligned} \phi_d(x, \mu_d, \Sigma_d) &= (2\pi)^{-D/2} |\Sigma_d|^{-1/2} \\ &\exp \left[ -\frac{1}{2} (x - \mu_d)^T \Sigma_d^{-1} (x - \mu_d) \right], \end{aligned} \quad (3.3.39)$$

with covariance matrix  $\Sigma_d$  and mean  $\mu_d$ . We summarise the algorithm based on [30] as follows:

- (a) First choose an importance function based on a linear combination of densities  $\phi_d$ .
- (b) Draw a sample set  $(x_1^1, x_2^1, \dots, x_{N_s}^1)$  from these densities, thereafter, computing the weights  $w_n^1$  for each sample point  $x_n^1$  for  $n = 1, \dots, N_s$ . At any iteration  $t$ , the importance weights are updated by

$$w_n^t = \frac{\pi(x_n^t)}{\sum_{d=1}^{N_c} \alpha_d \phi_d(x_n^t, \mu_d^t, \Sigma_d^t)}, \quad (3.3.40)$$

with corresponding  $\bar{w}_n^t$  given in equation (3.3.34).

- (c) Once this is done, we iterate and update component weights  $\alpha_d$ , and the parameters of the components, using the normalised importance weights  $\bar{w}_n^t$ , such that

$$\alpha_d^{t+1} = \sum_{n=1}^{N_s} \bar{w}_n^t \rho_d, \quad (3.3.41)$$

$$\mu_d^{t+1} = \sum_{n=1}^{N_s} \bar{w}_n^t x_n^t \rho_d / \alpha_d^{t+1}, \quad (3.3.42)$$

$$\Sigma_d^{t+1} = \frac{\sum_{n=1}^{N_s} \bar{w}_n^t (x_n^t - \mu_d^{t+1})(x_n^t - \mu_d^{t+1})^T \rho_d}{\alpha_d^{t+1}}, \quad (3.3.43)$$

where

$$\begin{aligned} \rho_d &= \rho_d(x_n, \alpha_d, \mu_d, \Sigma_d) \\ &= \frac{\alpha_d \phi_d(x_n, \mu_d, \Sigma_d)}{\sum_{d=1}^{N_c} \alpha_d \phi_d(x_n, \mu_d, \Sigma_d)}, \end{aligned} \quad (3.3.44)$$

and can be regarded as the weighted PDF of the components, with respect to a particular sample point. We then draw a new sample set  $(x_1^{t+1}, x_2^{t+1}, \dots, x_{N_s}^{t+1})$  from the updated components and compute  $w_n^{t+1}$  associated with each new sample point. This procedure is repeated until iteration  $T$ , when equation (3.3.37) is minimised sufficiently, hence satisfying the convergence criterion.

The degeneracy of the importance weights  $w_n^t$  as the dimension increases is a drawback of the importance sampling scheme. The problem arises here when the variance of the weights  $\text{Var}(\{w_n^t\})$  is large, with  $\max\{w_n^t\} \approx 1$ . This indicates a low number of representative samples and often occurs when  $\pi$  and  $q^t$  have disjoint support. The maximum of the importance weights converges to 1 if  $N_s$  is sub-exponential in the system dimension. For a likelihood  $P(D|\mathbf{x}^t)$  we have for sample point  $n$  at any iteration,

$$w_n \propto P(D|x_n) = \prod_{i=1}^M P(D_i|x_n), \quad n = 1, \dots, N_s, \quad (3.3.45)$$

where the importance weights are drawn from a likelihood consisting of a product of a possible large number of factors  $M$ , which can result in a large variance of the normalised importance weights, hence indicating a low number of effective samples. To overcome this problem for Gaussian mixture densities, a nonlinear transformation of the importance weights is suggested [138], creating a Nonlinear PMC (NPMC) algorithm. For a nonlinear positive function  $\varphi^t : (0, \infty) \rightarrow (0, \infty)$ , that may depend on iterations  $t$  and sample size  $N_s$ , we have the nonlinear transformation

$$w_n^{t*} = \varphi^t(w_n^t) \quad (3.3.46)$$

with corresponding normalised transformed importance weights

$$\bar{w}_n^{t*} = \frac{w_n^{t*}}{\sum_{i=1}^{N_s} w_i^{t*}}. \quad (3.3.47)$$

The algorithm is then updated as before with  $\bar{w}_n^{t*}$  replacing  $\bar{w}_n^t$ .

There are two types of nonlinear transformations with the first given by

$$w_n^{t*} = (w_n^t)^{\gamma^t}, \quad (3.3.48)$$

where  $\gamma^t$  adapts with every iteration and  $0 < \gamma^t \leq 1$ , with  $\gamma^t$  closer to 0 during earlier iterations and approaching 1 as the sampler gets closer to converging. This transformation method is known as *tempering* and recommended choices of  $\gamma^t$  are polynomial functions of the form  $\gamma^t = t^m$  for  $m \geq 0$  or a sigmoid function  $\gamma^t = \frac{1}{1+e^{-t}}$ .

The second type of transformation is called *clipping*, where we have

$$w_n^{t*} = \min(w_n^t, \Upsilon_{M_T}^t), \quad (3.3.49)$$

where  $\Upsilon_{M_T}^t$  is a specified threshold value that denotes the  $M_T^{\text{th}}$  highest importance weight, with  $1 < M_T \leq N_s$ . The threshold is chosen to guarantee that the number of samples that satisfy  $w_n^t \geq \Upsilon_{M_T}^t$ , is equal to  $M_T < N_s$ . A sufficient requirement is that  $M_T$  is selected to adequately represent the target distribution. An alternative to the transformation in equation (3.3.49) is to use

$$w_n^{t*} = \frac{2\beta^t}{1 + \exp\left(\frac{-2w_n^t}{\beta^t}\right)} - \beta^t, \quad (3.3.50)$$

where  $\beta^t > 0$  should increase along with iterations since there is a progressive reduction in the nonlinear distortion of the standard weights. The above transformation can be implemented if the ESS, calculated using the conventional importance weights, is below some minimum specified by the user. In the applications considered in Chapter 5, we found it necessary to implement these techniques, specifically the soft clipping method, in order to improve the sampling of large and degenerate parameter spaces.

The conventional PMC algorithm has already been applied successfully to cosmological parameter estimation and model selection using weak gravitational lensing in the Canada-France

Hawaii Telescope Lensing Survey (CFHTLenS) [45], strengthening support for the standard cosmological model. Apart from constraining the cosmological parameters of the standard model, tight constraints have also been obtained for dark energy and the primordial power spectrum parameters [139]. A recent Bayesian technique called *Model Averaging* [139] has been designed to overcome the problem of weak model discrimination with data, which can result in underestimating the uncertainty of parameters. The method aims to solve this problem by incorporating the uncertainty of the correct model into the calculation of the parameter probabilities, which requires estimating posterior probabilities of models and parameters using PMC. PMC has even had success in challenging inference for the multivariate skew-normal model parameters [140], through a Bayesian analysis approach. The NPMC algorithm has been applied to stochastic kinetic models (SKMs) [138], which are multivariate Markov jump processes that explain the interaction among species in biochemical systems according to a set of unknown rate parameters. To estimate the posterior distribution of these rate parameters requires Bayesian analysis, via Monte Carlo sampling techniques. A comparison between MCMC and PMC has been done in a cosmological setting [31] and in a SKM problem with NPMC and MCMC, with both studies concluding higher computational costs for MCMC. The increased availability of computational resources, such as high performance computing systems, motivates the use of algorithms that can easily run in parallel.

## CHAPTER 4

---

### Optimisation of PMC algorithm parameters

---

The PMC sampler must be initialised such that its algorithm parameters are specified according to the choice of the target distribution. The choice of the algorithm parameters will ultimately affect the performance of the algorithm. This Chapter focuses on selecting the parameters to optimise the performance of PMC. We define a measure of performance so as to determine the effect of the algorithm parameters. Furthermore, we will also address the convergence of the chain and compare convergence measures for the PMC algorithm to that used for MCMC. The optimisation study is extensively investigated for a wide range of target distributions that are commonly experienced in practice. We begin by examining the measure of convergence of the chains.

#### **4.1 A sufficient convergence statistic**

PMC is capable of detecting when the importance function has converged sufficiently to the target distribution. This process of stopping the sampler when convergence is achieved can therefore

be automated. In this Section, we establish the convergence criteria and discuss its compatibility with a convergence statistic used in MCMC algorithms to determine if it is sufficient.

Following the work of [31], instead of the Kullback-Liebler divergence,  $K$ , we use an estimator of equation (3.3.37) at each iteration  $t$ ,

$$\text{ESS}^t = \frac{1}{N_s} \left( \sum_{n=1}^{N_s} (\bar{w}_n^t)^2 \right)^{-1}, \quad (4.1.1)$$

which is called the normalised *Effective Sample Size*, which essentially measures the number of sample points accepted. The statistic lies in the range  $[0, 1]$ , with  $\text{ESS} \geq 95\%$  indicating that convergence has been achieved [31], which we confirm from our simulations. For complex distributions of high dimensions (greater than six) however, the ESS threshold decreases to 80% [31]. We compare this statistic to the convergence statistic  $r$ , used in [35] for the MCMC algorithm, since both measures require different criteria to be satisfied to indicate convergence. The statistic  $r$  is defined as,

$$r = \frac{\sigma_{\bar{x}}^2}{\sigma_{\pi}^2}, \quad (4.1.2)$$

where  $\sigma_{\bar{x}}^2$  is the variance from the sample mean and  $\sigma_{\pi}^2$  is the variance of the target distribution. Convergence is achieved when  $r < 0.01$ , which indicates the mean of the sample is close enough to the target mean.

We conduct a simulation to compare the effectiveness of both these statistics which involves sampling from a one dimensional Gaussian target distribution. In order to relate the  $r$  statistic to *ESS*, we look at which statistic requires more iterations to satisfy convergence criteria. This indicates which of the two statistics are more strict. We use 100 realisations at various sample sizes, viz:  $N_s = 30, 40, 50, 80, 100, 150, 200, 300, 400$ . Figure 4.1 shows an ESS cut-off greater than 99% is related to the  $r$  statistic cut-off of  $r < 0.01$ . A high ESS cut-off of 99% is used since the mixture densities are of the same form as the target distribution. Furthermore, we are working with a one dimensional distribution, so high ESS measures can be achieved. For small

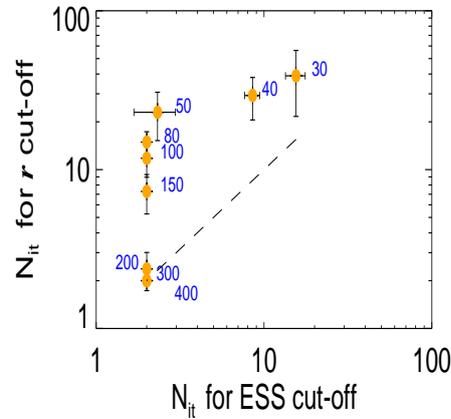


Figure 4.1: The minimum number of iterations,  $N_{it}$  required to attain  $r < 0.01$  against  $N_{it}$  required to attain an ESS greater than 99%. Data points for each sample size (in orange) are shown with error bars from the standard deviation estimated from 100 independent realisations. The corresponding sample sizes are indicated next to the points. The dashed curve represents a one-to-one relationship between the two statistics.  $N_{it}$  for  $N_s = 300$  and 400 are almost the same.

sample sizes less than 200 the statistic  $r$  is far stricter than ESS judging by the greater number of iterations required to achieve  $r < 0.01$  as can be seen from Figure 4.1. However for large enough  $N_s$  ( $N_s \geq 200$ ) results indicate that ESS is just as strong a constraint as  $r$ , since to achieve both statistics, the same number of iterations are required. The difference between ESS and  $r$  is that the latter is more dependent on the sample size for an accurate estimate of the target mean, whereas ESS requires that most sample points replicate the target distribution, irrespective of the coverage of high density regions. This means that for smaller sample sizes, convergence based on the  $r$  statistic requires many more iterations than convergence based on the ESS statistic. It is therefore advisable to choose the sample size large enough to increase the probability of covering all regions of the target parameter space. In what follows we explain how the algorithm can be optimised to find high density regions of an unknown parameter space. We will investigate the

choice of an adequate sample size in Section 4.4.3 but first we derive a quantity to measure the performance of PMC.

## 4.2 Defining efficiency

Computational cost and wall-clock time are affected by the number of sample points PMC must evaluate to converge. A weighted proportion ( $\alpha_d N_s$ ) of samples are drawn from each mixture density,  $\phi_d$ , that collectively make up the sample of size  $N_s$  at each iteration  $t$ . Hence in order to compute  $q^t$ , each component must evaluate every sample point  $x_1, x_2, \dots, x_{N_s}$ . All  $N_s$  points are thereafter evaluated  $N_c$  times. The number of components may vary per iteration since low sample weights ( $\alpha_d \approx 1/N_s$ ) are disregarded. The total computational cost for varying number of components at each iteration  $t$  is then given by,

$$N_{cost,t} = N_{c,t} N_s. \quad (4.2.3)$$

If no components become disregarded, then after  $T$  iterations we have the total number of evaluations,

$$N_{Total} = \sum_{t=1}^T N_{cost,t} = T N_c N_s, \quad (4.2.4)$$

where  $N_{Total}$  is the total number of evaluations and is the actual computational cost that we wish to minimise. Note that we are not restricted to a constant sample size at each iteration [141], therefore equation (4.2.3) can be modified as necessary. For the purpose of this study however, we keep the sample size constant at each iteration. We define the efficiency,  $E$ , of the PMC algorithm to be the inverse computational cost, which is then measured by,

$$E = (N_{Total})^{-1}. \quad (4.2.5)$$

In what follows we will determine the optimal choice of the PMC algorithm parameters by maximising  $E$  with respect to those parameters through the use of simulations.

### 4.3 Simulations

In this Section we describe the PMC algorithm parameters which affect efficiency of the sampler and briefly explain the simulation used to optimise efficiency with respect to these parameters. A large number of independent realisations are conducted in order to derive the relationship between  $E$  and the input variables, which we refer to as PMC algorithm parameters, thereby allowing us to study the optimal choices of these parameters. The mixture densities follow a Gaussian distribution for all simulations in this thesis. We are required to choose  $N_c$  and  $N_s$  initially, as well as the parameters of the mixture densities, arising from their mean,  $\mu_d$ , and covariance,  $\Sigma_d$ . We choose to draw the mean and diagonal covariance,  $\Sigma_d = \text{diag}(\sigma_d, \dots, \sigma_d)$ , from uniform distributions as follows. We choose  $\sigma_d$  such that,

$$\sigma_d \sim \text{UNIFORM}(\sigma_0 - \sigma_1, \sigma_0 + \sigma_1), \quad (4.3.6)$$

with  $\sigma_0$  and  $\sigma_1$  representing the center and width of the uniform distribution respectively and similarly for the mean,

$$\mu_d \sim \text{UNIFORM}(\mu_0 - \mu_1, \mu_0 + \mu_1). \quad (4.3.7)$$

The six PMC algorithm parameters are therefore  $(N_s, N_c, \sigma_0, \sigma_1, \mu_0, \mu_1)$ .

Each simulation involves varying all six algorithm parameters over a specific range, creating a discrete parameter space over which  $E$  is measured. To reduce noise in the simulated data we use 30 realisations for each point in the parameter space. To obtain the relationship of  $E$  against each algorithm parameter, we average over the remaining five algorithm parameters reducing the parameter space to the parameter of interest with corresponding  $E$ . Using this method enables us to optimise the performance of PMC for each of its algorithm parameters.

## 4.4 The Gaussian target distribution

We first run a simulation on a simple target distribution and use the ESS convergence statistic as well as the measure of efficiency to gauge the performance of PMC on a distribution that has the same functional form as its components. A multivariate Gaussian distribution for various dimensions is used as our target distribution. The dimensions considered are 2, 3, 4, 5 and 8 to give us a general understanding of the performance of PMC against dimension. The target distribution is

$$\pi(\mathbf{x}) \sim N(\boldsymbol{\mu}, \mathbf{I}), \quad (4.4.8)$$

with  $\boldsymbol{\mu}$  a vector of means set to 1 for all dimensions, and  $\mathbf{I}$  is the Identity matrix.

The components are able to obtain the exact functional form of this target distribution and PMC is therefore able to match  $q$  very closely to  $\pi$ , obtaining a high ESS. We therefore use an  $\text{ESS} \geq 98\%$  as our convergence statistic. Using this ESS cut-off, we now study the relationship of  $E$  and the algorithm parameters.

### 4.4.1 Sensitivity to mixture density parameters

We first examine the parameters of the mixture densities, specifically the effect of the algorithm parameters  $\sigma_0, \sigma_1, \mu_0$  and  $\mu_1$  on  $E$ . If the shape of the mixture densities is well matched to the target distribution we would expect much quicker convergence. The mixture density parameters will reflect our prior knowledge of the target distribution, therefore this section will illustrate the penalty for limited prior knowledge. We vary  $\sigma_1$  and  $\mu_1$  over a range that will sufficiently illustrate the efficiency of the algorithm against these parameters. This relationship is shown in the top panel of figure 4.2 for the dimensions considered.

For lower dimensions ( $\leq 3$ ), efficiency is more sensitive and penalised for larger variances which reflects more ignorance about the target distribution parameters. For larger dimensions

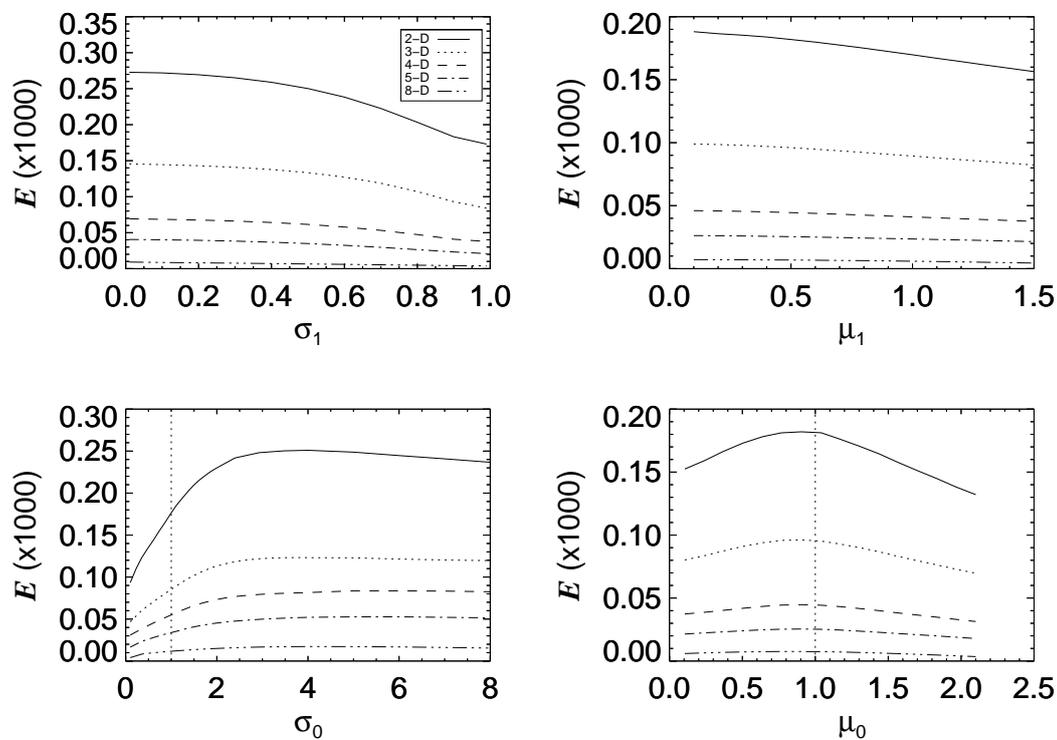


Figure 4.2: *Top left:* The efficiency against the scatter of the uniform distribution from which the variance of the mixture density is drawn. *Top right:* As in the top left panel, the efficiency against  $\mu_1$  for various dimensions. *Bottom left:* The efficiency against  $\sigma_0$ . The vertical line represents the magnitude of the variance of the target distribution. *Bottom right:* Efficiency against  $\mu_0$  which peaks at the target mean indicated by the vertical line. This indicates that PMC performs best when the components are positioned close to the target as would be expected.

however, efficiency is relatively flat as  $\sigma_1$  and  $\mu_1$  increase, indicating that the penalty for lack of knowledge of the target parameters is not substantial, compared to that for smaller dimensions. This would indicate that PMC is not very sensitive to the selection of  $\sigma_1$  and  $\mu_1$  when the dimension of the distribution considered is greater than three for a multivariate Gaussian distribution.

The positioning of components is also important when initialising the sampler. This is done by selecting  $\sigma_0$  and  $\mu_0$  to position the variance and mean of the mixture density respectively. The ideal situation is to position the components directly on the target i.e. setting  $\sigma_0$  and  $\mu_0$  to the

mean and variance of the target distribution. This is not possible in most situations due to the lack of information about the underlying distribution. We illustrate how efficiency is affected by the choice of  $\sigma_0$  and  $\mu_0$  in the bottom plots of Figure 4.2. Results from the simulation show that  $\mu_0$  should be chosen to position the components in high density regions of the target parameter space, preferably matching the means of the components with that of  $\pi$ . We find PMC favours a large width compared to the target distribution and performs better as compared to a smaller choice of  $\sigma_0$ . Efficiency increases even after  $\sigma_0$  exceeds the width of the target distribution, which is shown by the vertical line. PMC quickly adapts to sample the target distribution when there is a good match of distributions between the mixture densities and the target distribution. When the mixture densities encompass all relevant regions of the target distribution, it is able to quickly assume the target distribution shape, covering all relevant regions of low and high density. This occurs when  $\sigma_0$  is chosen to be larger than the width of the target distribution. On the other hand, when  $\sigma_0$  is chosen smaller than that, the sampler searches in a manner synonymous with ‘searching in the dark’, since it has little information about the parameter space spanned by  $\pi$ . The sampler must then find relevant regions of the parameter space with more trial and error, hence requiring more iterations. It is therefore advisable to choose a larger width when specifying the covariance matrix than what you assume the target covariance to be. This is in agreement with MCMC [35] but here the efficiency is not as strongly dependent on  $\sigma_0$ .

Prior information is important when initialising the sampler but the relative penalty due to ignorance of the mixture density parameters is less severe for higher dimensions.

#### 4.4.2 Optimal number of components

The sampler is initialised such that all components have an equal probability of succeeding in sampling  $\pi$ ; therefore, we assign equal weighting to all components, i.e.  $\alpha_1 = \alpha_2 = \dots = \alpha_{N_c}$ . Components that sample from the target, such that the variance of the normalised importance

weights is low, will have more samples drawn from their distribution function, whereas components that struggle to sample the target distribution and remain stuck in very low density regions will be discarded or *buried*. The sampler can be modified such that components that are buried, can be *revived* by positioning them close to the component with highest weight,  $\alpha_{j,\max}$ , and given the same covariance matrix as that component. The weights for all components can then be distributed equally. This is needed for more complex distributions and we implement this method in our application to cosmology. Other reasons that may cause the components to be buried are due to their covariance no longer satisfying the criteria for being positive-definite, i.e.  $\lambda_i \leq 0$  for eigenvalue  $\lambda$  corresponding to dimension  $i$ . This is most likely to occur in regions of the target parameter space that are degenerate and is difficult to overcome. For the Gaussian target distribution however, components rarely become buried as most often they are able to sufficiently sample  $\pi$ .

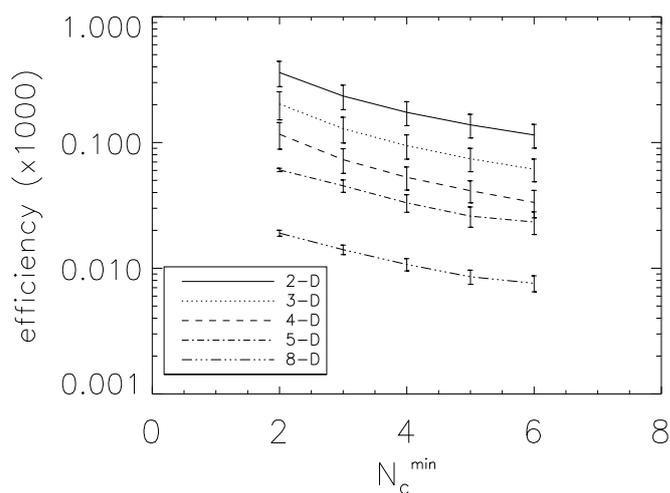


Figure 4.3: The efficiency against the minimum number of components,  $N_c^{\min}$ , needed for convergence. These are points from our simulation averaged over all five remaining PMC parameters. Data points including  $1\sigma$  errors are shown along with fitting curves for all dimensions.

The results of the simulation for various dimensions of the target distribution are shown in Figure 4.3. The plot illustrates fitting curves along with the data and  $1\sigma$  errors. We fitted a power law to each of these curves and found a power law index consistent with -1 in each case, that is  $E \sim 1/N_c$ .

As can be seen from the figure, the optimal choice for the number of components is  $N_c^{opt} = 2$ . Components, like the target distribution, are multivariate Gaussian distributions which enables the least amount of components, namely  $N_c = 2$ , to match the target distribution, which is required to satisfy the ESS convergence criterion. For more complex target distributions however, we are not guaranteed that this result will persist, as studied in section 4.5.

### 4.4.3 Optimal sample size

We next consider how the efficiency varies as a function of sample size and derive what the optimal choice of the sample size,  $N_s$ , is for a given dimension. The optimal sample size,  $N_s^{opt}$ , should be large enough to yield the required ESS but small enough to minimise the computational cost, hence maximising  $E$ . As can be seen in Figure 4.4, there is a particular  $N_s$ , corresponding to the number of iterations,  $T$ , that maximises  $E$ . Iterations, as efficiency, are averaged over the many independent realisations mentioned in Section 4.3. We therefore examine the relationship of  $T$  against  $N_s$  and find it is well described by the sum of two exponential functions given by,

$$T(N_s) = a_1 \exp\left[\frac{b_1}{N_s}\right] + a_2 \exp\left[\frac{b_2}{N_s}\right] \quad (4.4.9)$$

where  $a_1, a_2, b_1$  and  $b_2$  are the fitting parameters found using least squares. By assuming the number of components do not change significantly after each iteration (since this is evident in most realisations), we use equation (4.4.9) in conjunction with equation (4.2.4) to obtain a profile

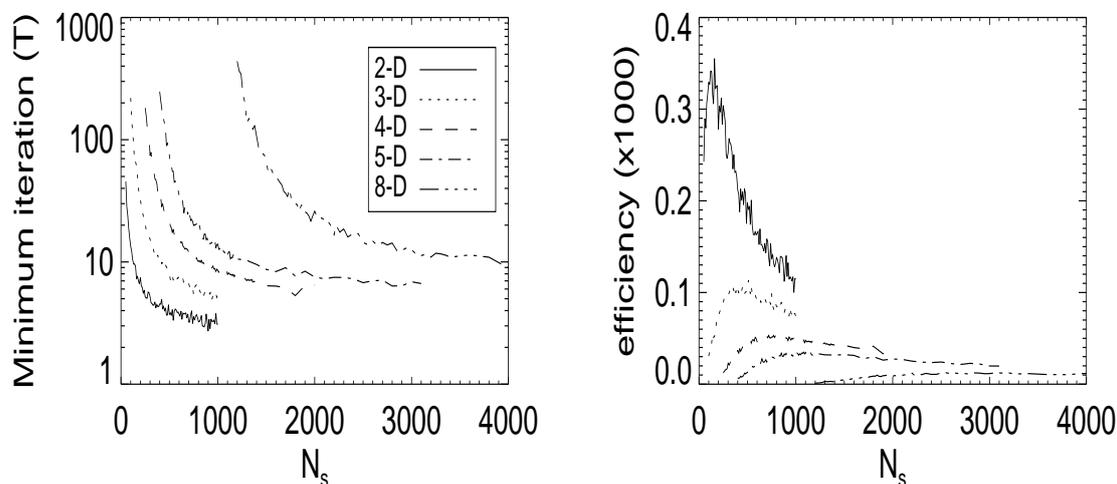


Figure 4.4: *Left*: The minimum number of iterations required for convergence according to the sample size chosen for various dimensions of the target distribution. *Right*: The corresponding efficiency against the sample size.

of  $E$  against  $N_s$  such that

$$\begin{aligned}
 E &= [N_{Total}]^{-1} \\
 &= \frac{1}{N_s N_c} \left[ a_1 \exp\left(\frac{b_1}{N_s}\right) + a_2 \exp\left(\frac{b_2}{N_s}\right) \right]^{-1}.
 \end{aligned} \tag{4.4.10}$$

In Figure 4.5 we show an example of the fit of  $T$  to the data using the 3-dimensional case, and the corresponding efficiency fit. In this case and in other dimensions, the adopted functional form provides a good fit to the simulation data. The optimal sample size is found by maximising equation (4.4.10) with respect to  $N_s$  to obtain  $N_s^{opt}$ , with  $N_c = N_c^{opt} = 2$  since these are the optimal number of components. The data for the simulation is shown in Figure 4.6, which illustrates the relation between the optimal sample sizes and dimension,  $D$  in the top left plot. To extend the range of the  $N_s - D$  relation, we have filled in other dimensions than those mentioned earlier,

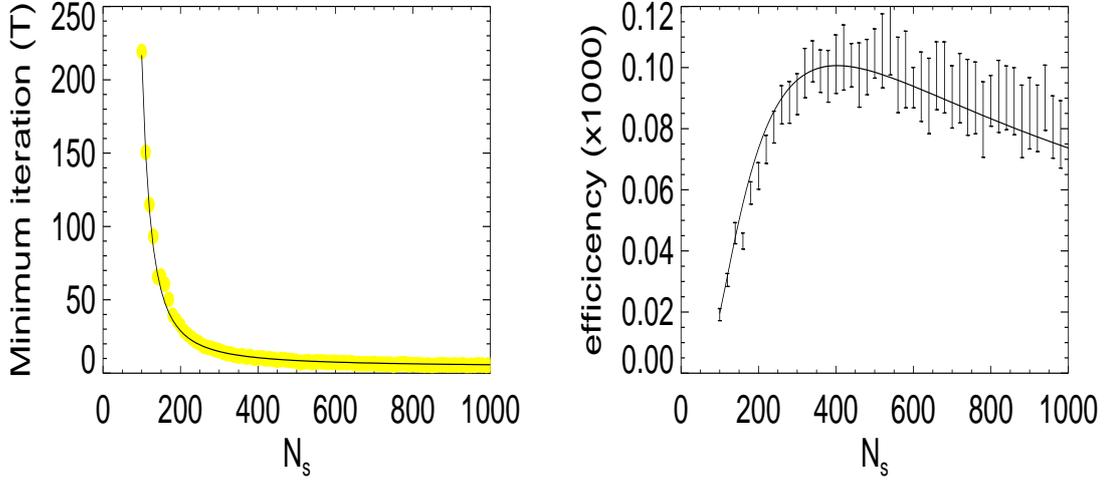


Figure 4.5: *Left*: Fitting function for  $T$  against the data shown for a 3-dimensional target distribution. The solid curve represents the fitting function with the closed circles representing data points. Including error bars would make the plot very crowded but here the emphasis is on the behaviour of  $T$  against  $N_s$  and the quality of fit to the data, which proves quite accurate. *Right*: The corresponding fit to data of efficiency against sample size including  $1\sigma$  errors bars obtained from the sample variance of the realisations.

but keeping  $\sigma_1$  and  $\mu_1$  fixed (sampling the full 6 dimensional PMC algorithm parameter space is very costly, particularly in higher dimensions), since the study in Section 4.4.1 indicated a weak dependence on those parameters. Closed circle points represent the full six PMC algorithm parameter simulation while open circles represent the restricted simulation. The figure suggests that there is a simple power law relation between  $N_s^{opt}$  and  $D$ , and we find that,

$$N_s^{opt} = \gamma D^\beta, \quad (4.4.11)$$

where  $\gamma \approx 49.1558 \pm 15.53$  and  $\beta \approx 1.964 \pm_{0.03}^{0.049}$  are the best fits to the free parameters. Equation (4.4.11) indicates a quadratic relation between  $N_s^{opt}$  and  $D$ , which is explained by the elements of the covariance matrix scaling as  $D^2$ .

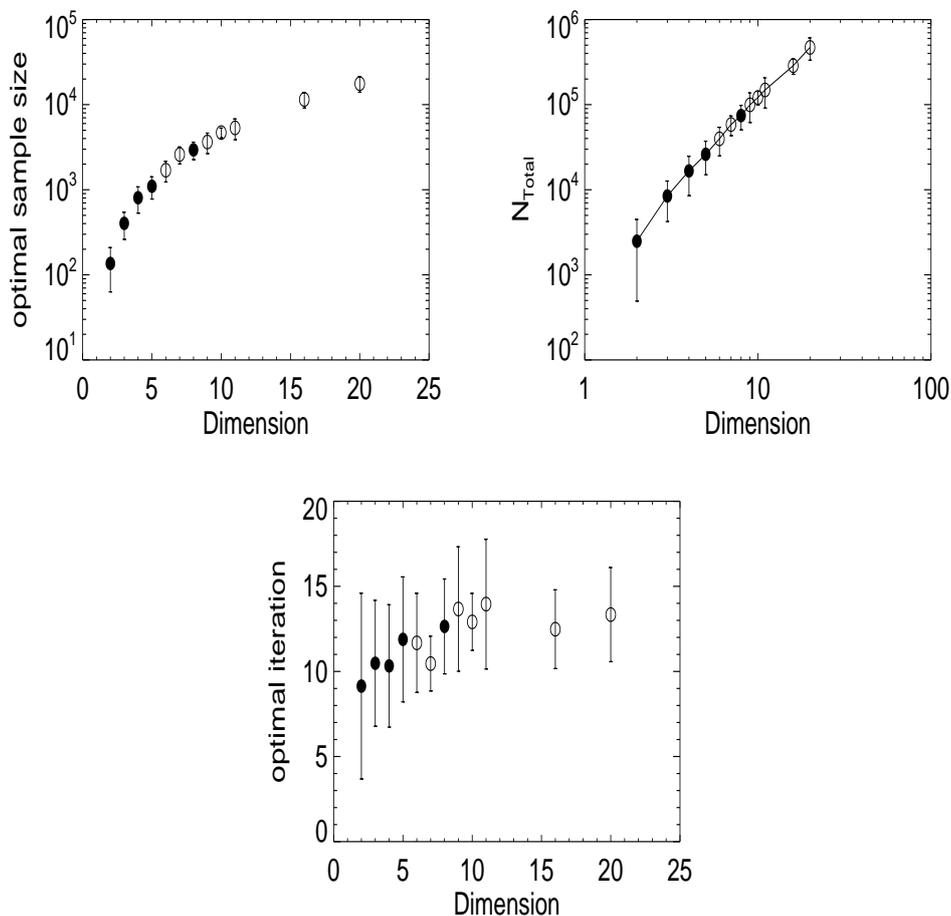


Figure 4.6: *Top Left*: The optimal sample size for the complete algorithm parameter set simulation shown by closed circles with the restricted set shown by open circles. Data is shown with  $1\sigma$  errors. *Top Right*: The cost of sampling multivariate Gaussian distributions of various dimensions using optimal sample size and number of components. Solid line represents the fitting curve for equation (4.2.4) using (4.4.11) and (4.4.9). *Bottom*: The optimal number of iterations  $T^{\text{opt}}$  for the corresponding optimal sample size in each dimension. The data indicates that the number of iterations required to converge for all dimensions are the same when using an optimal sample size.

By using the optimal sample size we can determine the optimal number of iterations for

Dimension	$N_{s,opt}$	$N_{Total}$ at $N_{s,opt}$
2	$136 \pm 73$	2485.61
3	$403 \pm 142$	8446.2
4	$806 \pm 275$	16641.4
5	$1098 \pm 321$	26096.8
6	[1697 $\pm$ 463]	39640.6
7	[2793 $\pm$ 587]	58430.2
8	$2941 \pm 682$	74389.2
9	[3641 $\pm$ 984]	99526.8
10	[4696 $\pm$ 617]	121269
11	[5330 $\pm$ 1463]	148763
16	[11410 $\pm$ 2852]	326761
20	[17637 $\pm$ 3636]	470523

Table 4.1: Optimal sample size including  $1\sigma$  errors. Entries in brackets represent the restricted algorithm parameter simulation. The last column illustrates the corresponding computational cost for  $N_{s,opt}$  at  $N_c = 2$  components.

each dimension and the associated computational cost. We find that using  $N_s^{opt}$  will result in convergence with the same number of iterations, that is, the optimal number of iterations,  $T^{opt}$ , is virtually the same for all dimensions as can be seen in the bottom panel of Figure 4.6. The relationship of the number of evaluations to dimensions is quadratic, as shown in the top right plot of Figure 4.6, which is due to the quadratic relation between  $N_s^{opt}$  and  $D$ . The optimal sample size and computational cost for each each dimension is shown in Table 4.1. We have thus been able to optimise the algorithm with respect to the sample size from equation (4.4.11) for an optimised number of components  $N_c^{opt} = 2$  and mixture density parameters based on prior knowledge of the target distribution.

## 4.5 Complex target distributions

So far we have considered the simple case of a Gaussian target distribution and used components that have the same probability distribution function as that of the target. It is instructive to investigate cases where the target distribution have a different functional form to that of the components and to determine whether or not these mixture densities can adapt in order to reconstruct more complex distributions. To this end we investigate the ability of PMC to efficiently sample from a banana shaped, bimodal and hypercube target distribution. We will start by looking at the banana shaped target distribution.

### 4.5.1 Banana distribution

The banana distribution has a strong degeneracy between parameters that are twisted. The target distribution for  $D$  random variables,  $x_i$ , for  $i = 1, \dots, D$ , is a multivariate Gaussian distribution in all dimensions, except  $x_D$  which is twisted in the remaining dimensions, in the following manner

$$(x_1, \dots, x_{D-1}, y) \sim \mathbf{N}(\mathbf{0}, \Sigma) \quad (4.5.12)$$

with the random variable  $y$  encoding the twist, such that

$$y = x_D + B \left( \frac{x_{D-1}^2}{\sigma_1^2} - 1 \right) + \dots + B \left( \frac{x_1^2}{\sigma_1^2} - 1 \right). \quad (4.5.13)$$

We set  $\sigma_1 = 10$  with  $D \times D$  covariance matrix  $\Sigma = \text{diag}(\sigma_1^2, \sigma_1^2, \dots, 1)$ . The parameter  $B$  controls the degree of complexity of the distribution, shown in Figure 4.7 in a 2-dimensional case, for illustrative purposes. We therefore, use  $B$  synonymously with complexity. We find that PMC is unable to sample dimensions higher than two when complexity exceeds  $B \approx 1.5$ . Consequently we examine only the two-dimensional case. We illustrate how PMC uses multiple components, shown by blue contours in Figure 4.7, to adequately cover the target parameter

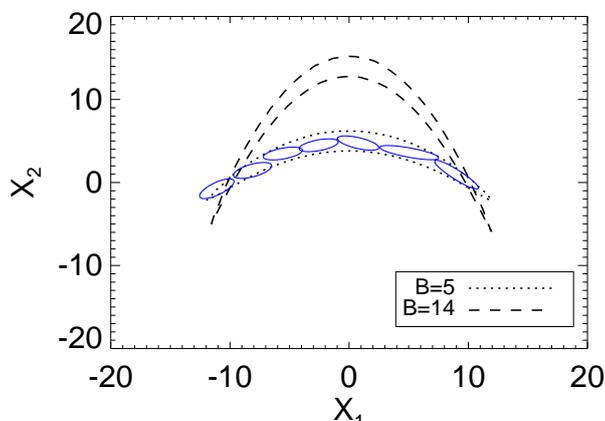


Figure 4.7: Contour plots of two dimensional distributions; with two different values of  $B$ , showing the effect of complexity on a banana distribution. Components of a converged chain are shown by solid blue contours covering the distribution with  $B = 5$ .

space for the case of  $B = 5$ . We wish to know how the degree of complexity affects the number of components required to reach convergence. In order to determine this, we conduct a simulation whereby we vary  $B$ , and for a particular  $N_c$ , we run a number of realisations until convergence is achieved. If a sufficiently large number of realisations have converged, e.g. more than 50%, then we are able to sample the target distribution of that complexity with at least that many components. We optimise all other algorithm parameters and position them in regions of high density ensuring we have ideal situations for each complexity chosen. This is done to ensure that failure to achieve convergence is due to insufficient components and not other factors. Results of the simulation are shown in Figure 4.8, which illustrates that the number of components must be increased along with increasing complexity. To investigate efficiency against complexity, we follow the same methodology as mentioned in Section 4.3, but with algorithm parameters  $N_s, N_c, \mu_0, \sigma_0$ , and  $B$ . We exclude  $\sigma_1$  and  $\mu_1$  because of their weak effect on  $E$ . Results are shown in the right panel of Figure 4.8. We find efficiency decreases exponentially at low  $B$ ,

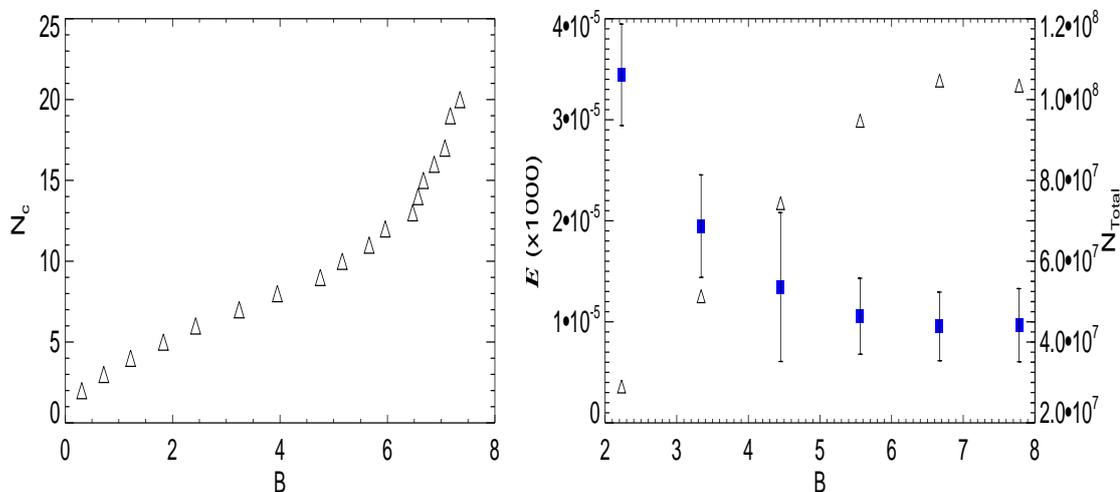


Figure 4.8: Banana target distribution results. *Left panel:* The minimum number of components, shown by triangles, required to sufficiently sample the target parameter space of a banana distribution as a function of complexity,  $B$ . *Right panel:* Efficiency decreases exponentially with complexity as shown by square points with  $1\sigma$  error bars obtained from the standard deviation of realisations. The corresponding computational cost,  $N_{Total}$ , is shown by triangles.

whereas at high complexity,  $6 \leq B \leq 8$ , efficiency becomes constant due to the compensation with the sample size. PMC requires more iterations for the components to adapt to the degenerate parameter space. Therefore increasing complexity causes the number of iterations to decrease exponentially. This effect causes the exponential decrease of efficiency at low  $B$ .

The banana target distribution forces PMC to use more components than in the case of the Gaussian target distribution. There is a minimum number of components  $N_c^{\min}$ , that the sampler requires in order to sufficiently sample the target distribution. If we use less components than  $N_c^{\min}$ , convergence will not be achieved. The computational cost is proportional to the number of components, as can be seen in equation 4.2.4, and we find that  $T$  stabilises for  $N_c \geq N_c^{\min}$ ,

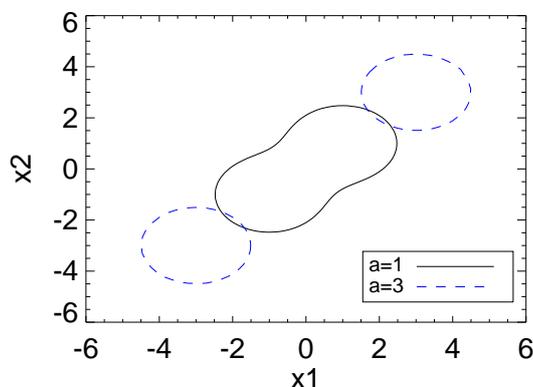


Figure 4.9: A symmetric bimodal distribution with two peaks shown by a contour plot at  $1\sigma$  levels. Peaks that are close together result in a mixture Gaussian distribution shown by  $a = 1$  (solid) and peaks that are separated by a region of low probability are shown with  $a = 3$  (dashed).

therefore the optimal number of components is  $N_c^{\text{opt}} = N_c^{\text{min}}$ .

We will now look at a more complicated distribution, specifically, the bimodal distribution.

### 4.5.2 Bimodal distribution

Distributions with more than one peak can pose a problem for PMC, specifically those with narrow peaks connected by regions of very low probability. We consider the symmetric bimodal distribution

$$\begin{aligned}
 p(\mathbf{x}) \propto & \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{a})\Sigma^{-1}(\mathbf{x} - \mathbf{a})^T \right] \\
 & + \exp \left[ -\frac{1}{2}(\mathbf{x} + \mathbf{a})\Sigma^{-1}(\mathbf{x} + \mathbf{a})^T \right], \quad (4.5.14)
 \end{aligned}$$

with modes situated at  $\pm\mathbf{a}$  and covariance  $\Sigma = I$ , the identity matrix. A visual representation of a two-dimensional bimodal distribution for different modes is shown in Figure 4.9. If  $\mathbf{a} = (a_1, a_2, \dots, a_D)$  for  $D$  dimensions, then the greater the value of  $|a_i|$  for  $i = 1, 2, \dots, D$ , the

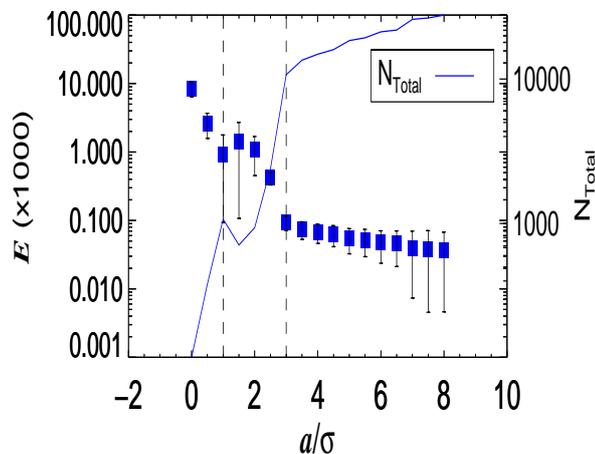


Figure 4.10: Plot of efficiency against  $a$  scaled by the width of each peak (blue squares). The corresponding cost  $N_{Total}$  of the simulation is shown by the solid blue line. Regions of interest are separated by vertical dashed lines.

further are the modes from each other and the more distinct the peaks are. We consider the case of  $a_i = a$  for all dimensions for simplicity. Furthermore we look at only the two dimensional case since our goal in this Section is to analyse the dependency of efficiency on  $a$ . By increasing  $a$ , we increase the distance between the two peaks and increase the length of the low probability region that separates them.

In keeping with the methodology of previous Sections, we conduct a simulation to investigate the relation between  $E$  and  $a$ , including  $a$  as an algorithm parameter. The relation between  $E$  and other algorithm parameters remain unchanged for the bimodal target distribution. As shown in Figure 4.10, efficiency behaves differently for different values of  $a$ . There are essentially three regions of  $a$  that affect the performance of PMC —these are  $[0, 1)$ ,  $[1, 3)$ , and  $[3, \infty)$ ; scaled to the standard deviation of each peak. Distributions that fall into each range are shown in Figure

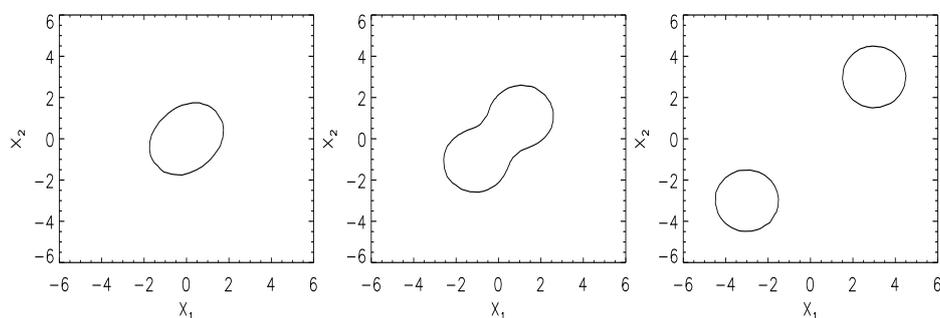


Figure 4.11: Shape of bimodal distributions for  $a$  in ranges  $[0, 1)$  (*left*),  $[1, 3)$  (*middle*) and  $[3, \infty)$  (*right*). Contours represent 68% confidence regions.

4.11. In the first range,  $[0, 1)$ , there is no region of low probability that separates the peaks and we are left with an almost perfect Gaussian distribution (fig 4.11 left panel), hence with Gaussian components, the sampler adapts easily and quickly to the target distribution.

If  $a$  increases to within the second interval,  $[1, 3)$ , we are faced with a mixture Gaussian distribution of two peaks, which are still significantly close together. At  $a$  close to 1 (fig 4.11 middle panel), there are two regions of high density where each mode is situated. Components tend to intersect with both these regions, until eventually adapting to one mode. This occurs after many iterations, which takes a long time. Approaching the upper limit of the second interval  $[1, 3)$ , the two peaks become distinct, i.e., they are separated by a significant low probability region, thus enabling each component to represent each peak well. This is indicative of the behaviour in the third interval (fig 4.11 right panel). In the third interval,  $[3, \infty)$ , the dependence of  $E$  on  $a$  can be explained by the sum of two power laws. PMC exhibits the lowest efficiency and we are faced with the problem of false convergence. This occurs when components adapt to only one peak, because the other is too far away. This can result in a satisfactory ESS that signals to PMC that convergence has been reached; when in fact it has not. This is a shortcoming of PMC for these types of distributions, i.e., the sampler will stop searching the target parameter space if

convergence has been reached even though the full distribution has not been properly sampled. The convergence statistic and criteria must be modified in order to overcome this problem.

In the next Section we consider a distribution that is often difficult to sample in practice, i.e., the hypercube distribution.

### 4.5.3 Hypercube distribution

In this Section we consider the hypercube target distribution, which in one dimension is simply the uniform distribution

$$\text{prob}(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

It is a square grid in a two dimensional parameter space, i.e., the plane. For higher dimensions the parameter space can be thought of as the volume within the hypercube with the probability of the vector of parameters being the inverse of the volume if within the multidimensional parameter space, and 0 otherwise. The hypercube target distribution will require many mixture densities, specifically Gaussian densities, to adapt to the shape of the flat target distribution. The mismatch of the distributions between the target and components will require  $N_c$  to increase significantly, especially in higher dimensions. In our study we consider the unit interval length such that  $a = 0, b = 1$ . We first examine the simplest of the hypercube distribution which is one dimension or univariate hypercube distribution.

#### One dimension

We consider the simplest case of one dimension and position the components at various locations within the target parameter space so as to cover the boundaries and central regions. We illustrate

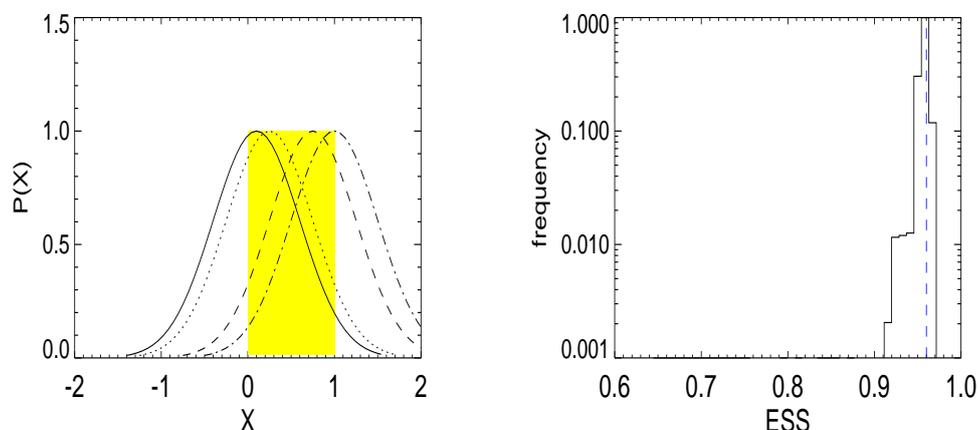


Figure 4.12: *Left panel:* Using 4 Gaussian mixture densities situated within the target parameter space with widths large enough to cover most of the parameter space. The yellow region is the parameter space of the target distribution. *Right panel:* Histogram of the ESS we achieved from the simulation with 8 Gaussian mixture densities. Vertical dashed line represents ESS cut-off of 96%.

the positioning of four components with respect to the target distribution in Figure 4.12. This is an ideal situation such that components sufficiently overlap almost all regions of the target parameter space.

The algorithm is able to achieve at least an ESS of 96% since the parameter space is small and PMC requires at least 8 components to achieve this. However, all components must effectively cover the target parameter space as in Figure 4.12. We repeat this simulation a number of times to determine the amount of trials that are able to converge. We plot the histogram of the ESS statistic in the right panel of Figure 4.12, which shows that a significant amount of trials permit convergence. We find the means of the components play an important role as the more central the positioning of the components relative to the target parameter space, the greater the fraction of points accepted. Intuitively this is reasonable because the Gaussian mixture densities overlap

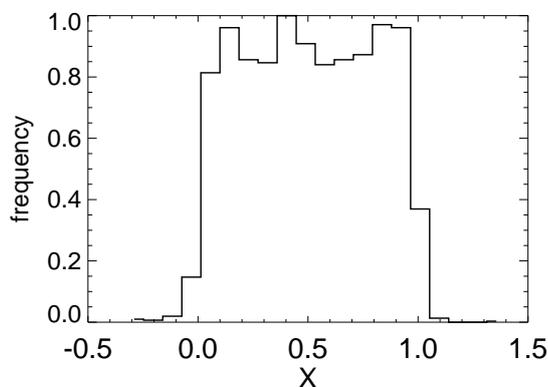


Figure 4.13: The distribution of the sample points from an iteration with corresponding  $\text{ESS} \approx 0.96$ . This illustrates the sample can reasonably adapt to the complex target distribution.

the parameter space of the target more than if they were closer to the end-points. There is no significant change in the performance when the widths are adjusted. Note that with 4 components, we are able to attain an ESS of 96%. A converged sample is shown in Figure 4.13.

### Higher dimensions

We increase the target distribution dimension to 2 and configure the means and widths of the components as in the univariate case. We assume from the number of vertices scaling with dimension,  $D$ , that we would need at least  $8^D$  components. We therefore choose to use 64 components and when the simulation is left to run for 3000 iterations, it is able to achieve an ESS of 96%. The distribution of ESS achieved is shown in the right panel of Figure 4.14, and the distribution of the converged chain is illustrated in the left panel of Figure 4.14. The distribution is very close to a square resembling the shape of the target distribution. We note the ESS on average exceeds 96% after 700 iterations, which is very time consuming and the cost of achieving an ESS of 98%, as in the Gaussian target case, is much higher. The weights of all components

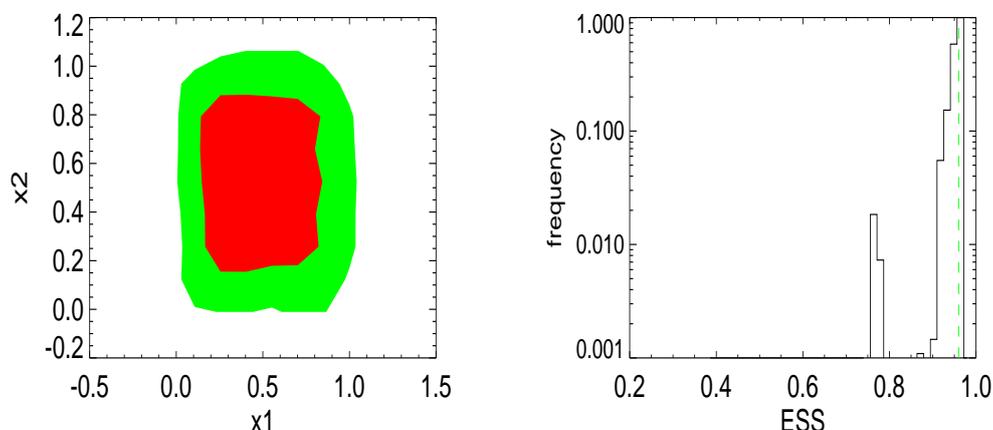


Figure 4.14: *Left panel:* The joint distribution of a converged sample from a two dimensional hypercube. Green regions represent the 95% confidence level with red representing the 68% level. *Right panel:* Histogram of ESS for a two dimensional simulation using 64 components (solid line). Dashed vertical line is the 96% ESS marker.

stabilise at the same weight during the simulation which indicates that each mixture density does not progress to higher density regions in order to draw more samples. A progression of two components is shown as an example after every 500<sup>th</sup> iteration in Figure 4.15.

For three dimensions we use  $8^3$  components with the same initialisation as for the previous lower dimensional studies. Using such a large number of components increases the computational cost significantly. Much time is spent inverting the covariance matrices of all components. We experience numerical problems when such a large number of components is used, due to some components venturing into regions of low density, which results in the component weight,  $\alpha_j$  for component  $j$ , approaching zero. When this happens, adjustments must be made to reduce the number of components or redistribute the samples drawn from all components. While this can overcome the problem of small component weights, another problem arises when the covariances of components become singular or are no longer positive-definite: components be-

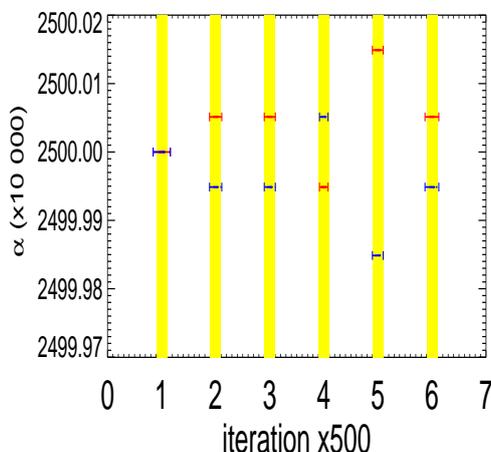


Figure 4.15: Progression of two components during the simulation after every 500 iterations shown by dashed blue and solid red horizontal bars in the shaded yellow region. The range of the y-axes has been scaled so that each component is distinguished from each other. Both components overlap in the first iteration and are essentially the same for the remaining iterations judging by the negligible difference in weights. The shaded region represents the positioning of the target mean with respect to the component mean and the width of the region is proportional to the target width.

come situated in regions of the target parameter space that are degenerate with respect to certain parameters. In this instance we disregard these components thus reducing the overall number of components.

For the three dimensional case the ESS oscillates around 80%, far from the 96% goal, and although it may be possible to reach this ESS level with more components, this does not seem practical in terms of the computational time and resources required. Overall the hypercube distribution seems to present a challenge for the PMC algorithm to sample, a challenge which becomes particularly significant in higher dimensions.

---

## Application of PMC methods to cosmology

---

### 5.1 The likelihood function for CMB data

We briefly discussed the likelihood in Section 3.1 and now address its role in real cosmological problems incorporating data. To derive the likelihood function we closely follow the approach and notation of Dodelson [108]. Several effects are accounted for by the likelihood code. These are the mode coupling due to sky masking and non-uniform pixel weighting; beam window function uncertainty, which is correlated across the entire spectrum; and residual point source subtraction uncertainty. All of these effects are addressed in [36, 142] and [143]. We let  $s$  represent the true temperature anisotropy of the sky in a particular pixel, and the average of the temperature anisotropies over the entire sky by  $\Delta$ , which is an estimate of  $s$ . The true anisotropy is believed to follow a Gaussian distribution with zero mean and variance due to the signal denoted by  $C_S$  (i.e.,  $C_S$  is the usual cosmic variance). The probability distribution for  $s$  given  $C_S$

is

$$P(s|C_S) = \frac{1}{\sqrt{2\pi C_S}} \exp\left[-\frac{s^2}{2C_S}\right], \quad (5.1.1)$$

and similarly for  $\Delta$  given  $s$ ,

$$P(\Delta|s) = \frac{1}{\sqrt{2\pi C_N}} \exp\left[-\frac{(\Delta - s)^2}{2C_N}\right], \quad (5.1.2)$$

where  $C_N$  is the variance due to the noise including all source noise covariance in the maps. We ultimately want two statistics: the best-fit estimate and its uncertainty, which summarise the information contained in the data.

The likelihood function is an integral over all possible  $s$  such that

$$\mathcal{L} = \int_{-\infty}^{\infty} P(s|C_S)P(\Delta|s)ds. \quad (5.1.3)$$

Therefore, substituting (5.1.1) and (5.1.2) into (5.1.3), we obtain

$$\mathcal{L} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi C_S}} \exp\left[-\frac{s^2}{2C_S}\right] \frac{1}{\sqrt{2\pi C_N}} \exp\left[-\frac{(\Delta - s)^2}{2C_N}\right] ds. \quad (5.1.4)$$

The exponent in (5.1.4) depends on both the variance of the noise and of the signal. Therefore,

$$C = C_N + C_S, \quad (5.1.5)$$

which is the full covariance matrix and incorporates contributions from both the noise and the signal. We can now rewrite the likelihood for a single pixel as

$$\mathcal{L} = \frac{1}{\sqrt{2\pi C}} \exp\left[-\frac{\Delta^2}{2C}\right]. \quad (5.1.6)$$

The likelihood function is a product over all  $N_p$  measurements over all pixels, given by

$$\mathcal{L} = \frac{1}{(2\pi)^{N_p/2} |C|^{1/2}} \exp\left[-\frac{1}{2} \Delta C^{-1} \Delta^T\right] \quad (5.1.7)$$

where  $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_{N_p})$  is now a vector of measurements over all  $N_p$  pixels. Finding determinants of matrices of large dimension is computationally expensive. This is why much time is spent computing the likelihood for a given model. The covariance is symmetric with  $N_p(N_p + 1)/2$  independent elements. Considering the special case when  $C$  is diagonal with all elements equal, we have the likelihood over all  $N_p$  pixels for a single parameter,

$$\mathcal{L} \propto \left( \frac{1}{C_S + C_N} \right)^{N_p/2} \exp \left[ -\frac{1}{2} \frac{\sum_{i=1}^{N_p} \Delta_i^2}{(C_S + C_N)} \right]. \quad (5.1.8)$$

The signal covariance matrix represents the theoretical signal in a CMB experiment and can be used to forecast the performance of an experiment. This is done by finding the best-fit  $C_S$  that maximises the likelihood function, which is accomplished by simply differentiating (5.1.8) with respect to  $C_S$  and equating to zero to solve for  $C_S$ . Doing this produces the best-fit

$$\hat{C}_S = \frac{1}{N_p} \sum_{i=1}^{N_p} \Delta_i^2 - C_N, \quad (5.1.9)$$

which states that the theoretical signal is the difference between the variance of the data points and the variance of the noise. By assuming the likelihood is quadratic in the parameters, we obtain the error associated with (5.1.9) by taking twice the coefficient of the quadratic term in a Taylor expansion, so that

$$\ln \mathcal{L}(C_S) = \ln \mathcal{L}(\hat{C}_S) + \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}}{\partial C_S^2} \Big|_{C_S=\hat{C}_S} (C_S - \hat{C}_S)^2. \quad (5.1.10)$$

Representing the uncertainty by  $\sigma_{C_S}$ , we have

$$\begin{aligned} \sigma_{C_S} &= \left( -\frac{\partial^2 \ln \mathcal{L}}{\partial C_S^2} \right)^{-1/2} \\ &= \sqrt{\frac{2}{N_p} (C_S + C_N)}. \end{aligned} \quad (5.1.11)$$

Equation (5.1.11) shows that in order to reduce the uncertainty tenfold, we must increase the number of pixels a hundredfold. In the event that the free parameters are the  $C_\ell$ 's instead, we have

$$\sigma_{C_S} = \sqrt{\frac{2}{(2\ell + 1)f_s}}(C_S + C_{N,\ell}), \quad (5.1.12)$$

where the number of pixels have been replaced with  $(2\ell + 1)f_s$  for  $f_s$  representing the fraction of the sky covered. This is an accuracy threshold for the  $C_\ell$  measurements and is due to cosmic variance. In practice the likelihood analysis is more complicated and because we are using the WMAP nine-year data in this chapter, we briefly describe how the likelihood works for WMAP. The likelihood code uses the Fisher matrix,  $\ln \mathcal{F}$ , to assign errors to points in the CMB angular spectra, i.e., the code calculates the inverse Fisher matrix for the given CMB model spectrum, which is used to calculate  $-2 \ln \mathcal{L}$ . The central likelihood routine uses a pixel-based likelihood for low  $\ell$ , specifically  $2 \leq \ell \leq 30$  for the TT spectrum and  $2 \leq \ell \leq 24$  for the TE, EE and BB spectra. The value  $-2 \ln \mathcal{L}$  returned by the likelihood code is renormalised by subtracting a constant offset  $\ln |C_f|$ ,

$$-2 \ln \mathcal{L} = \chi^2 + \ln |\mathcal{F}^{-1}| - \ln |C_f|, \quad (5.1.13)$$

where  $\ln |C_f|$  is the sum of the determinant contributions to  $-2 \ln \mathcal{L}$  computed for the CMB spectrum and  $\chi^2$  measures the fit of the model spectrum to the data.

## 5.2 Parameter estimation: Adiabatic model

We apply the optimised PMC algorithm to the posterior distribution of the flat  $\Lambda$ CDM model with six parameters, mentioned in Section 2.10. We do this with the latest CMB data from the WMAP nine-year release [41]. We emphasize that our aim is not to make updated precision constraints of the cosmological parameters, but to optimise the performance of the sampler for a cosmological parameter estimation problem. Nevertheless the results obtained are interesting

and original in their own right, as they provide the most recent constraints on mixed adiabatic models correlated with two or more isocurvature modes.

### 5.2.1 Cosmological data

The WMAP mission objective is to determine the geometry, content, and evolution of the universe using a map of the temperature and polarization anisotropy of the CMB radiation. The likelihood of any model is calculated using the publicly available Version 5 WMAP 9 code [36, 60, 91, 142, 144]. In most parts, the likelihood for the nine-year data release is the same as the seven-year release except for minor adjustments [41].

In order for the likelihood to return a useful statistic, it must be given as input, the theoretical CMB power spectra for that model. We use the publicly available package CAMB [28] to compute the spectra. The WMAP nine-year likelihood code then takes as input the spectra for temperature (TT), B-mode (BB) and E-mode (EE) polarization as well as cross temperature with E-mode (TE), and returns the log likelihood for that model as computed in equation 5.1.13. This statistic is normalised to the log likelihood of the fiducial model and used as the numerator in equation 3.3.33. The data described in this section is used in our parameter estimation study for the purely adiabatic model and the models with adiabatic correlated with isocurvature modes.

### 5.2.2 Cosmological parameters

The parameters of interest are the standard six parameters for the flat  $\Lambda$ CDM model which are: the baryon density ( $\Omega_b h^2$ ); the cold dark matter density ( $\Omega_c h^2$ ); the dark energy density ( $\Omega_\Lambda$ ); optical depth to re-ionization ( $\tau$ ); the scalar spectral index ( $n_s$ ); and the amplitude of the primordial power spectrum ( $A_s$ ). We set flat uniform priors over a large enough range of interest for all parameters except  $A_s$ , with priors for the remaining parameters designed to exclude points that would cause numerical errors. The range of these uniform distributions are shown in Table 5.1.

Symbol	Description	Minimum	Maximum
$\Omega_b h^2$	Baryon density	0.01	0.04
$\Omega_c h^2$	Cold dark matter density	0	0.4
$\Omega_\Lambda$	Dark energy density	0.5	0.9
$\tau$	Optical depth to re-ionization	0.01	0.2
$n_s$	Scalar spectral index	0.8	1.3
$A_s$	Amplitude of the primordial power spectrum	0	none

Table 5.1: Prior constraints for cosmological parameters.

### 5.2.3 Optimizing PMC for cosmological parameter estimation: Adiabatic model

We assume that the joint distribution for the cosmological parameters is well approximated by a multivariate Gaussian distribution. There are, however, degeneracies expected between some of the parameters [145], which we expect to affect the performance of the sampler. In keeping with [31], we use Gaussian mixture densities, with covariance matrices and vector of means as its parameters. A covariance matrix must therefore be specified for each component initially. We use the inverse Fisher matrix ( $\mathcal{F}$ ) as the covariance, which is calculated by taking the derivatives of the  $C_\ell$ s with respect to the cosmological parameters at the fiducial model values  $(\Omega_b h^2, \Omega_c h^2, \Omega_\Lambda, \tau, n_s, A_s) = (0.0226, 0.11, 0.734, 0.088, 0.963, 15.7)$ . Each component is provided a covariance

$$\Sigma_d = c_d \mathcal{F}^{-1} \quad , \text{ for } d = 1, \dots, N_c, \quad (5.2.14)$$

where  $c_d$  is a scalar randomly chosen in the range  $[1, 3]$ . We choose this range following our result in Section 4.4.1 for  $\sigma_0$  and in keeping with the result for  $\mu_0$ , we choose the means of the components such that they are shifted to within 1% of the fiducial model. We initially use a non-conservative number of components,  $N_c = 15$ , to ensure that the algorithm can in fact reach

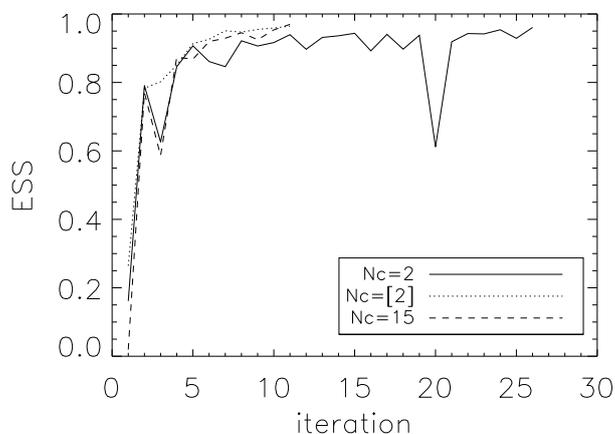


Figure 5.1: Adiabatic target distribution. Progression of ESS for runs that use different amount of components. The run with  $N_c = 2$  requiring 26 iterations to exceed a 96% ESS. The run using prior updates is indicated by  $N_c = [2]$ .

convergence and keep the sample size at  $N_s \approx 2000$ , which is the optimal sample size for a six dimensional Gaussian target distribution from Table 4.1. Results from [31] indicate the target distribution closely follows a Gaussian distribution. The likelihood evaluation of all models has been parallelised. Therefore  $N_s$  must be selected so that the work can be divided equally amongst all the processors. For this simulation, convergence is reached with the ESS exceeding 96% after 11 iterations.

We wish to investigate the dependency of cost as a function of the number of components, also if it is possible to sample the entire parameter space with as few components as possible. We therefore use two components but keeping the sample size the same and initialise the algorithm as in the 15 component case. Results are shown in Figure 5.1, which shows that more iterations are required to reach convergence ( $\approx 26$  iterations). Nonetheless the  $N_c = 2$  run remains less costly with an efficiency almost three times more than the 15 component run as can be seen in Table 5.2.

$N_c$	$N_{Total}$	Efficiency ( $\times 1000$ )
15	325 500	$3.07 \times 10^{-3}$
2	109 200	$9.16 \times 10^{-3}$
[2]	42 000	$23.81 \times 10^{-3}$

Table 5.2: Summary of simulation with different components using a sample size of  $N_s = 2100$ . The  $N_c = [2]$  case used updates from the previous run.

It is worth investigating if prior knowledge affects the performance of the sampler. Therefore we use information from the converged chain of the 15 component run to initialise another simulation using 2 components. From the 15 components we chose 2 that have the covariance with highest determinants. Their means and covariances are now used as the initial parameters of the mixture densities for the updated run. With the same sample size as before, we find this simulation takes the same number of iterations to converge as the 15 component run ( $\approx 11$ ). This suggests PMC is sensitive to the initial choice of the algorithm parameters when the target distribution has parameters that display mild degeneracies. A key strategy for the PMC algorithm that we infer from this study is that it is worthwhile starting with a larger number of components using a less conservative sample size initially and updating the next run based on these results. The initial run need not converge but requires that an ESS exceeds 20% for some iteration which will yield sufficient distribution parameters to update the next run in which we can reduce the sample size and number of components thereby reducing the computational cost. This ESS limit must be adjusted to a smaller value, if the dimension of the target distribution increases [31]. There are some key points in this Section worth pointing out. Our results in this Section were more sensitive to the initialisation of the covariance matrices of the components compared to Section 4.4.1, as the sampler would fail to converge if samples were drawn from a very low density region. This is due to the degeneracies between parameters making them more complex

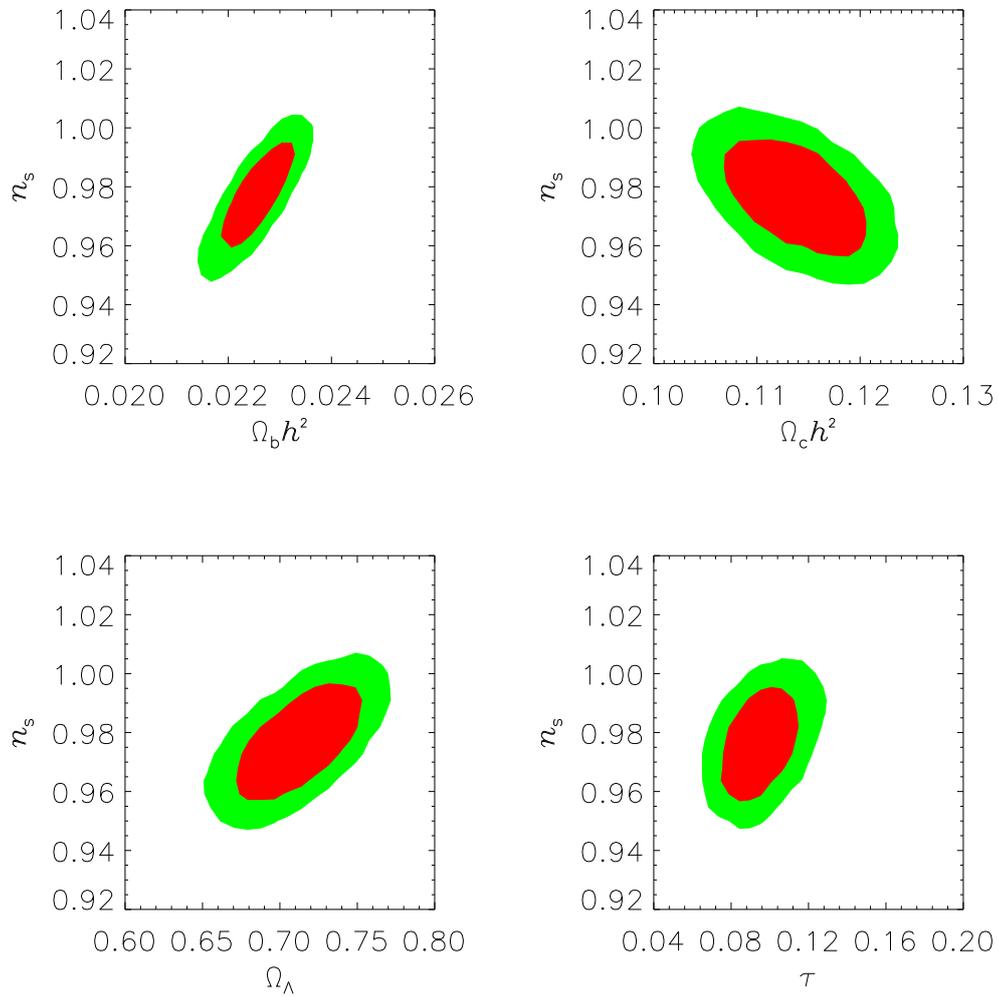


Figure 5.2: Pure adiabatic model. Marginal distributions between pairs of parameters illustrating the mild degeneracies. Confidence regions of 95% and 68% are shown by the green and red contours respectively.

than the Gaussian target distribution considered. The sampler would also take a long time to find models satisfying the prior conditions if a poor choice of the covariances and means is chosen. It

Parameter	mean $\pm 1\sigma$	MLE from WMAP nine-year $\pm 1\sigma$
$\Omega_b h^2$	$0.0226 \pm 0.0005$	$0.0226 \pm 0.0005$
$\Omega_c h^2$	$0.1136 \pm 0.0046$	$0.1138 \pm 0.0045$
$\Omega_\Lambda$	$0.713 \pm 0.027$	$0.721 \pm 0.025$
$\tau$	$0.096 \pm 0.013$	$0.089 \pm 0.014$
$n_s$	$0.977 \pm 0.013$	$0.972 \pm 0.013$

Table 5.3: Estimates from samples along with Maximum Likelihood Estimates (MLE) from WMAP nine-year data only [41].

is because of this that an updating procedure with various component sizes must be adopted. As mentioned in Section 5.1, the most time consuming step in computing the likelihood for the specified model is inverting the covariance matrix, but this process is parallelised, which drastically reduces the wall-clock time for PMC to sample the target parameter space.

Included for interest are some of the marginal distributions between the six cosmological parameters along with the marginal means and errors obtained from the run using two components with prior information from previous runs. These are shown in Figure 5.2 and Table 5.3 respectively. Statistics obtained from all runs with different components are in good agreement with each other. We find that parameter estimates are consistent with estimates obtained from the WMAP team’s analysis, which we list for comparison.

### 5.3 Sampling models with the adiabatic mode correlated with isocurvature modes

In this study we consider the adiabatic mode correlated with a single, two, and finally three isocurvature modes. For convenience we denote the modes as AD (adiabatic), CI (cold dark matter), NID (neutrino density), and NIV (neutrino velocity). The likelihood function described

in Section 5.1 is used with the difference being the input spectrum, which is now a sum of the spectra for adiabatic and isocurvature contributions. The target distributions studied are very complex with a higher dimensional and degenerate parameter space. It is for this reason that the performance of the sampler is extremely sensitive to the initial parameters chosen for the Gaussian mixture densities. For instance, if a component begins sampling in a region of low density, the covariance of that component will be updated to an invalid matrix (singular or negative definite) or the component weight  $\alpha$  will be updated to a value close to zero. To overcome these numerical problems we use the ‘revive’ method discussed in Section 4.4.2, which increases the chance of achieving convergence by moving components to higher density regions.

Prior constraints are now extended to the auto-correlations such that  $z_{II} > 0$  for  $I = \text{CI, NIV, NID}$ . It is therefore necessary to position components in a region of high density. It is for this reason that we start the algorithm with components positioned close to the adiabatic model. The means of the components are slightly perturbed around means obtained from the converged samples listed in Table 5.3 with the means of the auto- and cross-correlations set very close to zero. The covariance matrix is made up of a submatrix from the adiabatic results in Section 5.2.3 with the remaining off-diagonals set to zero and diagonal elements corresponding to the auto-correlations set to  $\approx 10^{-3}$ .

In our model the adiabatic auto-correlation is omitted since it is calculated from  $A_s$ , therefore the number of parameters to be estimated are  $D = 5 + \frac{N(N+1)}{2}$  for  $N$  being the total amount of modes considered. Our initial covariance  $\Sigma_{ij}^{\text{ISO}}$  is thus,

$$\Sigma_{ij}^{\text{ISO}} = \Sigma_{ij}^{\text{AD}}, \quad \text{for } 1 \leq i \leq 6, 1 \leq j \leq 6, \quad (5.3.15)$$

$$\Sigma_{ij}^{\text{ISO}} = 0, \quad \text{for } i \neq j \text{ and } 6 < i \leq D, 6 < j \leq D, \quad (5.3.16)$$

$$\Sigma_{ii}^{\text{ISO}} = 10^{-3}, \quad \text{for } i > 6, \quad (5.3.17)$$

where  $\Sigma_{ij}^{\text{AD}}$  is the adiabatic covariance matrix obtained from the converged samples in Section 5.2.3. This enables a gradual movement along the degenerate target parameter space so that components can adapt accordingly to those regions of the target distribution.

The Gaussian target distribution requires a minimum of two components to achieve convergence. However, the complexity of these target distributions motivates the use of many more components to explore the parameter space. These distributions are banana-like, and the complexity, defined in Section 4.5.1, is equivalent to  $B \approx 6$ . While we do not know the shape of the posterior in advance, the inherent degeneracies motivate choosing a larger  $N_c$  than 2. We therefore use 10 components and optimal sample size from Table 4.1. Using more than 10 components does not improve the efficiency of the sampler.

The complex parameter space of the target distribution in addition to the high dimension proves that it is more costly to sample from the the pure adiabatic case than would be expected. The cost of sampling these distributions is shown in Table 5.4. We use the optimal sample size of  $N_s = 3000, 6000$  and  $10000$  for the adiabatic plus one, two, and three isocurvature modes, respectively. The same results are found with larger sample sizes with the same number of iterations. It is therefore the number of components and their initial positioning that affect the performance of the sampler more significantly than the choice of  $N_s$ . For all mixed models, the exploration of the target parameter space is gradual and converges at different rates for each model. With respect to the single isocurvature mode models, the AD+CI model requires the most iterations ( $\approx 30$ ) followed by the AD+NID model ( $\approx 25$ ). The ESS statistic is expected to decrease as the dimension and complexity of the target distribution increases, but as long as the importance weights do not become degenerate, the ESS statistic can still achieve a large enough level, greater than 85%, which is sufficient for convergence.

The mixed models including two and more isocurvature modes are more complex. The higher dimension increases the parameter space causing the conventional ESS to stabilise at very low

Model	Iterations	$N_{Total}$	Efficiency ( $\times 1000$ )
AD+CI	30	900 000	$1.11 \times 10^{-3}$
AD+NID	25	750 000	$1.33 \times 10^{-3}$
AD+NIV	17	510 000	$1.96 \times 10^{-3}$
AD+CI+NIV	34	2 040 000	$0.49 \times 10^{-3}$
AD+CI+NID	32	1 920 000	$0.52 \times 10^{-3}$
AD+NIV+NID	34	2 040 000	$0.49 \times 10^{-3}$
AD+CI+NIV+NID	41	4 100 000	$0.24 \times 10^{-3}$

Table 5.4: The cost and efficiency of sampling models with correlated adiabatic and isocurvature modes.

values because there is significant degeneracy of the importance weights,  $\{w_n^t\}$  at iteration  $t$ , as discussed in Section 3.3. This is due to the high dimension and complexity of the target distribution. Nonetheless we explore the parameter space using the conventional PMC algorithm until there is no change in marginal distributions and statistics from the sample which occurs after 30 iterations. To overcome the degeneracy problem of the importance weights, we adopt the soft clipping transformation method in equation (3.3.50) with  $\beta_t = \frac{t}{t+1}$ , for  $t > 0$ . This satisfies the requirement that  $\beta_t > 0$  and increases with iterations since  $\beta_t \rightarrow 1$  as  $t \rightarrow \infty$ . The transformation of the importance weights occurs only when the conventional ESS calculated is less than 50%. The transformation enables the chain to converge with an ESS greater than 90%. We use information obtained from the run using conventional PMC. The converged chain from that run is used to construct a covariance matrix and position components around models with various  $f_{ISO}$  and a good  $\chi^2$  statistic, i.e. within a  $2\sigma$  confidence level. The updated run using the soft clipping transformation thus uses prior information from the conventional PMC run, and is able to converge quickly afterwards. The costs of these simulations are shown in Table 5.4, taking into account the expense of using conventional PMC prior to the transformation method.

On a side note, the simpler distribution with one isocurvature mode achieves a reasonably high ESS without the transformation of the importance weights since there is negligible degeneracy of  $\{w_n^t\}$  for the lower dimensional target distribution. Nonetheless we implemented the soft clipping transformation to confirm that the results are consistent with conventional PMC.

The complete mixture model including all isocurvature modes proves most difficult to sample. The parameter space is challenging due to the high dimension and degeneracy amongst the parameters. The conventional PMC algorithm is unable to achieve a converged chain because of the degeneracy of  $w_n$ . Therefore we use the NPMC algorithm, in addition with updates from previous runs. The soft clipping transformation is used for the two isocurvature mode mixed models. The algorithm struggles to probe regions of higher isocurvature fractions. We therefore ensured that each update positioned components into regions with increasing  $f_{\text{ISO}}$  and high likelihood. The sampler often becomes stuck in regions of parameter space for a long time with no change in statistics from the sample. When this occurs we stop the algorithm and update it for the next run. This process is continued until the ESS threshold for convergence is exceeded. The cost of the simulation is given in Table 5.4. As expected, the mixed model including all isocurvature modes is computationally the most expensive case.

Having ensured that the PMC algorithm has converged for all adiabatic and isocurvature mode combinations, we turn to the results from the chains. The results that follow for the correlated two and three mode cases are the most recent. We begin by looking at the the adiabatic mode correlated with a single isocurvature mode first.

### 5.3.1 Adiabatic mode correlated with a single isocurvature mode

In this section we examine the adiabatic mode correlated with one of the three isocurvature modes (i.e., the CI, NID, and NIV mode). For each of the above combinations we illustrate the one-dimensional marginal posterior distributions from samples derived using PMC in Figures

5.3 (AD+CI), 5.4 (AD+NID), and 5.5 (AD+NIV). The top two rows in the Figures illustrate the cosmological parameters with the last row showing the distributions of the auto- and cross-correlations. For convenience we denote  $z_{\langle I, J \rangle}$  as  $\langle I, J \rangle$ . The distributions of mixed models are shown by the solid curve against the pure adiabatic model represented by the dashed curve. The dashed curve are only for the six cosmological parameters considered, which were obtained from the converged samples in Section 5.2. The marginalised posterior distributions for all models are in all cases as broad or broader than the pure adiabatic model. For all three mixed models, we find that the mean optical depth is lower than the pure adiabatic results but the uncertainty is large enough to indicate consistency between the mixed and adiabatic model. The marginal distribution of  $\Omega_\Lambda$  has a more pronounced skewed normal distribution for all mixed models. The introduction of isocurvature contributions shifts the cosmological constant density to the higher values. The cosmological constant density has the same effect on the CMB spectrum as in the adiabatic case. An increase in the density shifts the spectrum to lower multipoles, thereby modifying the locations of the peaks. The scalar amplitude,  $A_s$ , is directly proportional to  $\langle AD, AD \rangle$ , which is less constrained compared to the pure adiabatic case. Therefore the distributions in the mixed models are much broader compared to the pure adiabatic model.

The fractions permitted for the pure isocurvature modes and cross-correlations are small, and all cross-correlations are consistent with zero at the  $2\sigma$  level, with all three models dominated by the adiabatic mode. Models with high isocurvature fractions do not fit the data better than the best-fit adiabatic model. The marginal distributions for the relative powers  $\langle I, J \rangle$  follow a prominent skewed normal distribution for the auto-correlations, making it challenging for Gaussian components to replicate. Our mean estimates are for all model parameters are illustrated in Table 5.5 along with their 68% confidence levels. These results for all cosmological parameters are consistent with previous analysis with WMAP seven-year [79] and nine-year data [41]. The spectral index is in good agreement with results using Planck+WMAP nine-year data

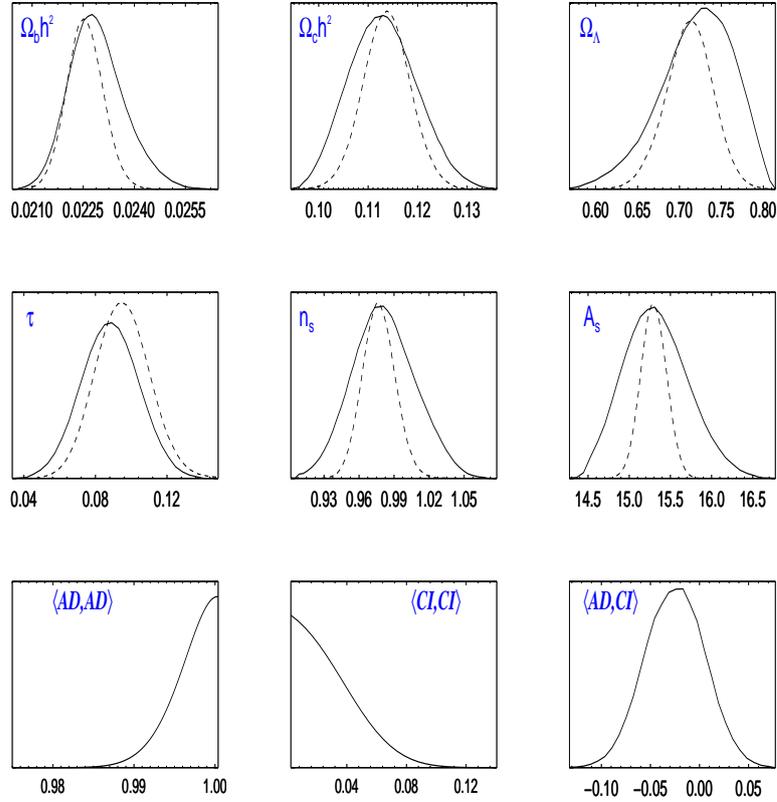


Figure 5.3: Adiabatic plus CI. Row 1 and 2 —the marginal posterior distributions for the six cosmological parameters. Row 3 —the marginal posterior distributions of the auto- and cross-correlations of the modes. The dashed line represents the pure adiabatic model with the adiabatic and isocurvature mixed model shown by the solid line.

$n_s = 0.9603 \pm 0.0073$  [81], with emphasis on tight bounds on  $n_s$  being crucial on constraining inflation. Unlike the adiabatic model, all mixed models include a scale-invariant spectrum with  $n_s = 1$  within the  $1\sigma$  error. This exceeds the limit for the simplest inflation models with  $n_s \leq 0.98$  [146, 147].

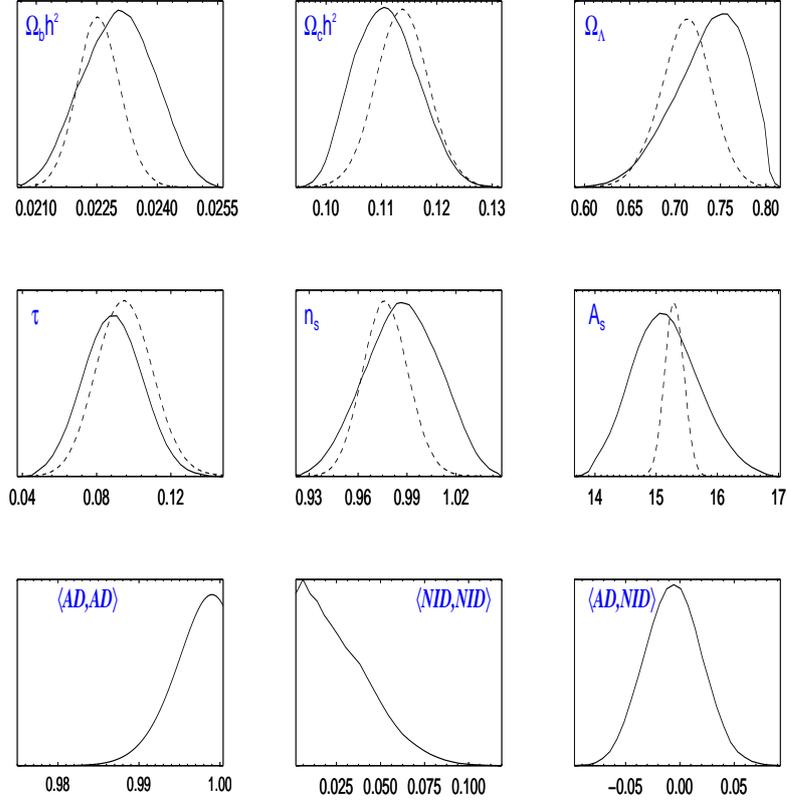


Figure 5.4: Adiabatic plus NID. Distributions as described in Figure 5.3 including only the neutrino density isocurvature mode.

These results indicate a strong preference for the adiabatic model, as most models are sampled with  $Z_{\langle AD, AD \rangle} \geq 99\%$ , although there is evidence to indicate that models with isocurvature contributions are sufficient to explain the data with maximum auto-correlations  $\leq 5\%$ . The fractional isocurvature contributions by all mixed models are consistent with previous findings using WMAP nine-year data [80]. Our results show that  $f_{\text{ISO}}$  is virtually the same for all models, being less than 9% with the least permitted fraction for the AD+NID model ( $f_{\text{ISO}} \leq 6\%$ ). The frac-

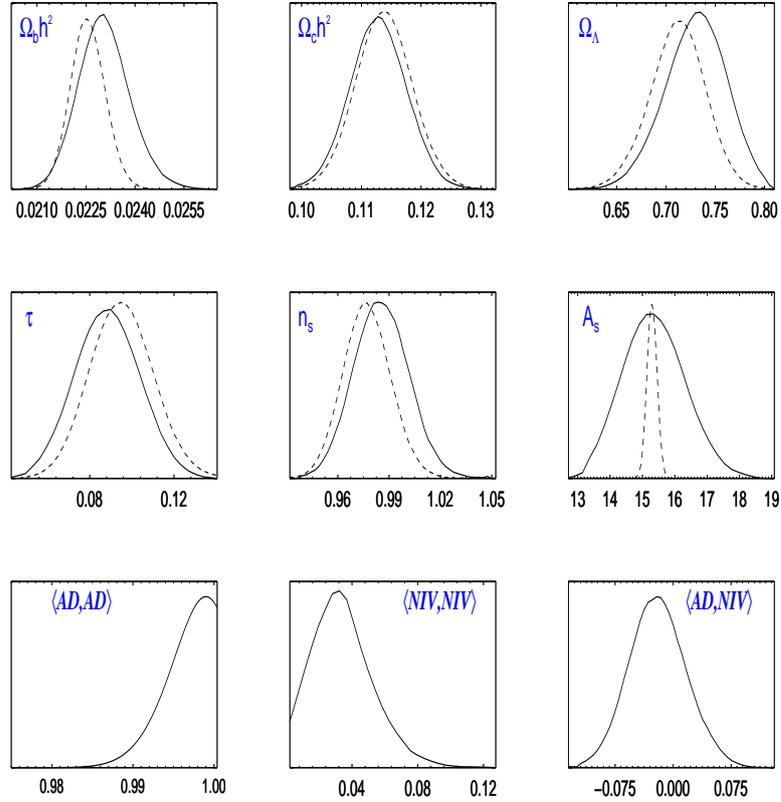


Figure 5.5: Adiabatic plus NIV. Distributions as described in Figure 5.3 including only the neutrino velocity isocurvature mode.

tional isocurvature contribution allowed is lower than results using WMAP first-year [71] and three-year [148] data, with a reduction by almost 50% from the first-year data, and far fewer mixed models preferred by the data. We find all estimates for the  $\langle I, J \rangle$  parameters to be in agreement with previous work using WMAP nine-year, three-year and first-year data with pure isocurvature contributions reduced by more than 50% compared to the first-year and three-year data.

Parameter	AD+CI	AD+NID	AD+NIV
$\Omega_b h^2$	$0.023 \pm 0.001$	$0.023 \pm 0.001$	$0.023 \pm 0.001$
$\Omega_c h^2$	$0.113 \pm 0.007$	$0.111 \pm 0.005$	$0.113 \pm 0.005$
$\Omega_\Lambda$	$0.724 \pm 0.042$	$0.739 \pm 0.037$	$0.728 \pm 0.032$
$\tau$	$0.088 \pm 0.013$	$0.089 \pm 0.014$	$0.088 \pm 0.014$
$n_s$	$0.981 \pm 0.026$	$0.987 \pm 0.021$	$0.986 \pm 0.015$
$A_s$	$15.33 \pm 0.36$	$15.16 \pm 0.50$	$15.27 \pm 0.84$
$Z_{\langle AD, AD \rangle}$	$0.998 \pm 0.003$	$0.999 \pm 0.001$	$0.998 \pm 0.002$
$Z_{\langle CI, CI \rangle}$	$0.029 \pm 0.022$	...	...
$Z_{\langle NID, NID \rangle}$	...	$0.03 \pm 0.02$	...
$Z_{\langle NIV, NIV \rangle}$	...	...	$0.04 \pm 0.02$
$Z_{\langle AD, CI \rangle}$	$-0.03 \pm 0.03$	...	...
$Z_{\langle AD, NID \rangle}$	...	$-0.006 \pm 0.025$	...
$Z_{\langle AD, NIV \rangle}$	...	...	$-0.017 \pm 0.035$
$Z_{\text{ISO}}$	$0.06 \pm 0.03$	$0.04 \pm 0.02$	$0.06 \pm 0.03$
$f_{\text{ISO}}$	$0.05 \pm 0.03$	$0.04 \pm 0.02$	$0.06 \pm 0.03$

Table 5.5: Adiabatic plus one isocurvature mode. We illustrate the mean with 68% confidence intervals for mixture models of adiabatic (AD) and one of three isocurvature modes, i.e. CI, NIV, NID.

We find it useful to quantify the posterior distributions in terms of its principal components which are the underlying structure in the model fit to the data. Principal components indicate the direction in which there is the most variance in the data. Principal component analyses can be constructed into eigenvalues and eigenvectors of the covariance for the data. We therefore calculate the eigenvalues,  $\lambda_\alpha$ , and eigenvectors,  $y_\alpha$ . The largest eigenvalue,  $\lambda_0$ , indicates the largest variance in the data and corresponding eigenvector,  $y_0$ , indicating the direction of the degeneracy. The components of the eigenvector (eigen-directions) are indicative of the contributed weight towards the degeneracy. We consider the eigen-directions that collectively make up 90%

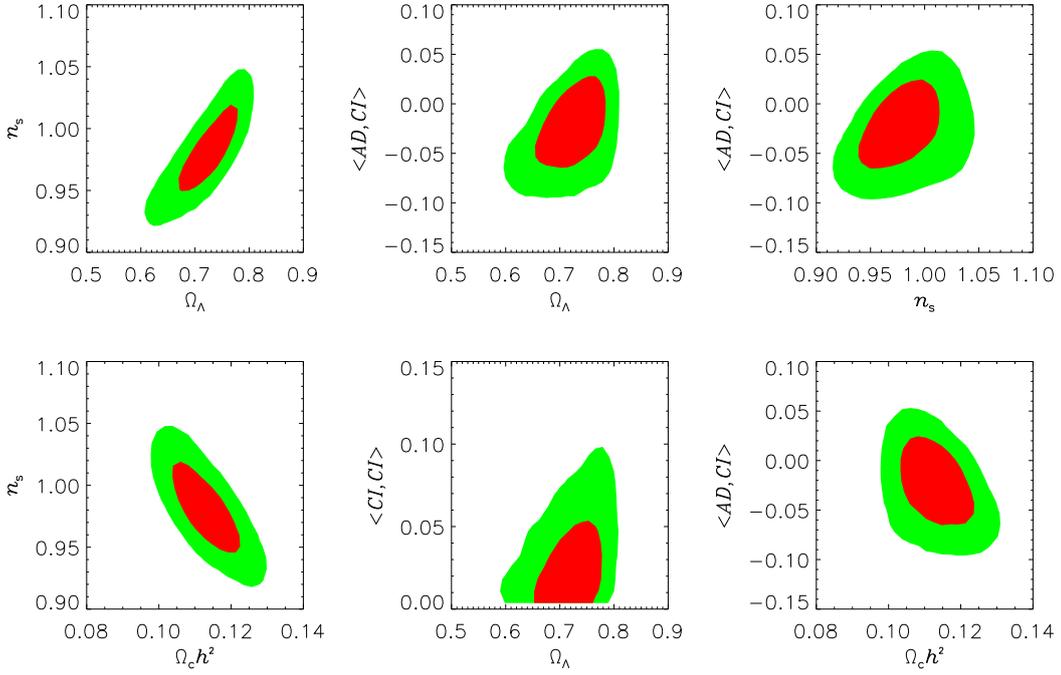


Figure 5.6: Adiabatic plus CI. Selected joint distributions between parameters depicting the most degenerate direction. Contour colours follow the same convention as Figure 5.2.

of the degeneracy.

For all mixed models we find that the error is dominated by  $\Omega_\Lambda$ ,  $n_s$  and cross-correlation  $\langle AD, I \rangle$  for  $I = CI, NIV, NID$ . Amongst the cosmological parameters we find significant degeneracy between  $n_s$  and  $\Omega_c h^2$  because the principal effect of the isocurvature modes is to modify the CMB temperature anisotropies at low  $\ell$ . We plot some of these joint distributions in Figures 5.6 (CI), 5.7 (NIV) and 5.6 (NID). All models have a strongly positive correlation between  $\Omega_\Lambda$  and  $n_s$ . The cosmological constant density is increased when the isocurvature contribution increases, which also shifts the spectral index to higher values. The NIV model has the smallest shift in  $\Omega_\Lambda$  since it is negatively correlated to  $\langle AD, NIV \rangle$  as seen in Figure 5.7. The joint distribution of the cosmological parameters and the auto- and cross-correlation parameters are degenerate for

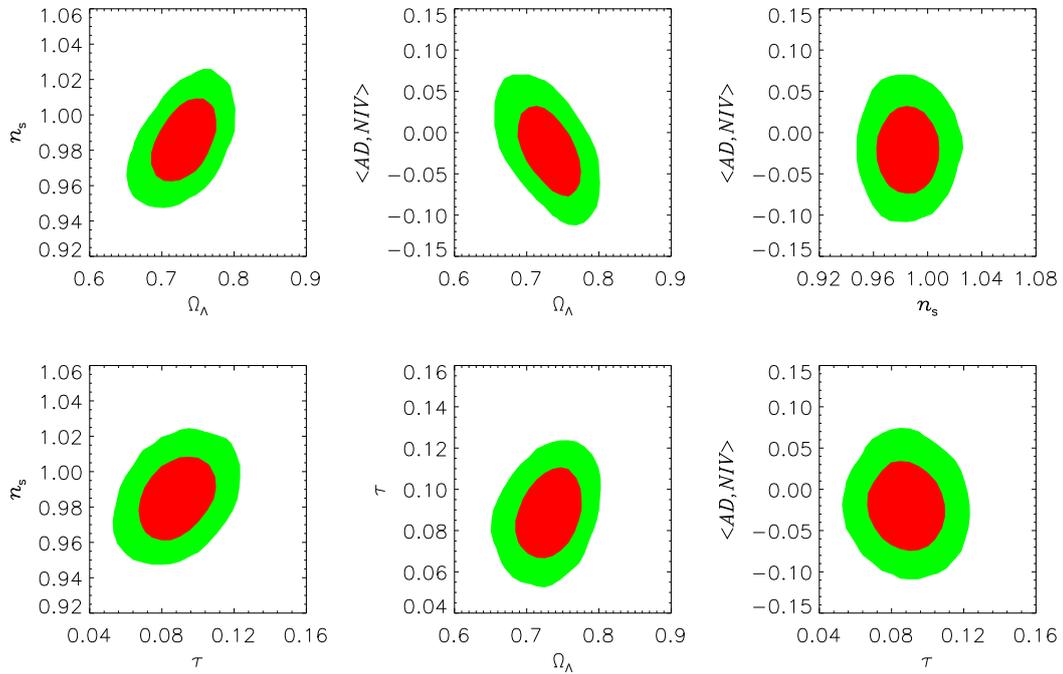


Figure 5.7: Adiabatic plus NIV. Selected joint distributions between parameters depicting the most degenerate direction. Contour colours follow the same convention as Figure 5.2.

all modes, making it challenging to achieve convergence, and with more isocurvature modes the distribution becomes more complex, as we will see in the next Section.

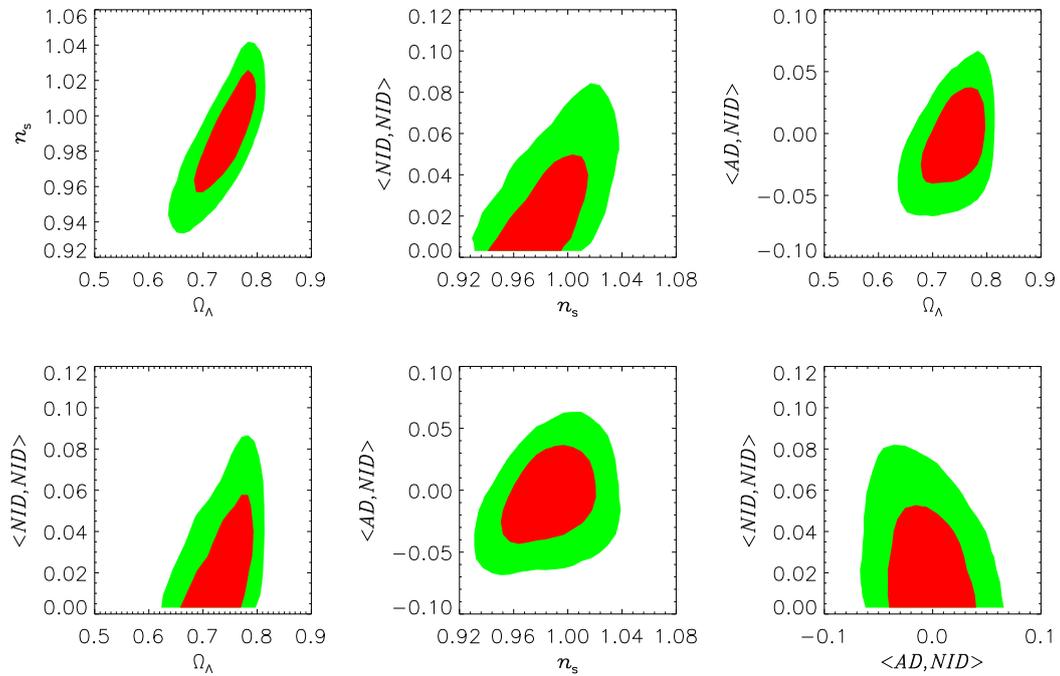


Figure 5.8: Adiabatic plus NID. Selected joint distributions between parameters depicting the most degenerate direction. Contour colours follow the same convention as Figure 5.2.

### 5.3.2 Adiabatic mode correlated with two isocurvature modes

In this Section we present the results of the adiabatic mode correlated with two isocurvature modes. The three mixed models are AD+CI+NID, AD+CI+NIV and AD+NID+NIV. In all three cases the fraction of isocurvature has now increased to 24.6% (AD+CI+NID), 21% (AD+CI+NIV) and 18% (AD+NID+NIV) as seen in Table 5.6, with a reduction of 38 – 59% from that found using WMAP first-year [71] and three-year [148] data. The baryon density however, is excluded from the range allowed using three-year data which was found to be  $0.025 < \Omega_b h^2 < 0.027$ . Compared to the adiabatic results, we find all cosmological parameters to be consistent. The constraint on  $n_s$  is poorer compared to the adiabatic model but still excludes the scale-invariant model. The adiabatic contribution decreases significantly as expected, with the pure isocurvature contribution much higher than in the single mode case.

Marginal posterior distributions for all mixture models are illustrated in Figures 5.9 (AD+CI+NID), 5.10 (AD+CI+NIV), and 5.11 (AD+NIV+NID). The cosmological parameters have not shifted significantly from the mean values in the adiabatic case, but we can see the deviation from a Gaussian distribution specifically for  $\Omega_\Lambda$ . The width of the  $A_s$  distribution has increased as expected since  $\langle AD, AD \rangle$  is less constrained compared to the single isocurvature mode case. The posterior distributions of the auto- and cross-correlations are shown in the last two rows of each Figure illustrating the shift of  $\langle AD, AD \rangle$  to lower power.

Parameter	AD+CI+NID	AD+CI+NIV	AD+NID+NIV
$\Omega_b h^2$	$0.024 \pm 0.001$	$0.021 \pm 0.001$	$0.022 \pm 0.001$
$\Omega_c h^2$	$0.119 \pm 0.006$	$0.123 \pm 0.008$	$0.119 \pm 0.006$
$\Omega_\Lambda$	$0.71 \pm 0.04$	$0.67 \pm 0.05$	$0.68 \pm 0.04$
$\tau$	$0.091 \pm 0.015$	$0.079 \pm 0.014$	$0.079 \pm 0.013$
$n_s$	$0.952 \pm 0.024$	$0.923 \pm 0.045$	$0.957 \pm 0.025$
$A_s$	$15.69 \pm 0.66$	$19.51 \pm 2.58$	$18.33 \pm 2.50$
$Z_{\langle AD, AD \rangle}$	$0.941 \pm 0.045$	$0.961 \pm 0.027$	$0.972 \pm 0.019$
$Z_{\langle CI, CI \rangle}$	$0.128 \pm 0.075$	$0.040 \pm 0.025$	...
$Z_{\langle NID, NID \rangle}$	$0.110 \pm 0.063$	...	$0.027 \pm 0.017$
$Z_{\langle NIV, NIV \rangle}$	...	$0.065 \pm 0.032$	$0.083 \pm 0.036$
$Z_{\langle AD, CI \rangle}$	$-0.075 \pm 0.080$	$-0.115 \pm 0.046$	...
$Z_{\langle AD, NID \rangle}$	$0.061 \pm 0.058$	...	$-0.082 \pm 0.037$
$Z_{\langle AD, NIV \rangle}$	...	$-0.112 \pm 0.048$	$-0.091 \pm 0.048$
$Z_{\langle CI, NID \rangle}$	$-0.119 \pm 0.057$	...	...
$Z_{\langle CI, NIV \rangle}$	...	$0.057 \pm 0.032$	...
$Z_{\langle NID, NIV \rangle}$	...	...	$0.059 \pm 0.026$
$Z_{\text{ISO}}$	$0.32 \pm 0.12$	$0.26 \pm 0.09$	$0.22 \pm 0.08$
$f_{\text{ISO}}$	$0.25 \pm 0.08$	$0.21 \pm 0.06$	$0.18 \pm 0.06$

Table 5.6: Adiabatic plus two isocurvature modes. The same conventions apply as in Table 5.5.

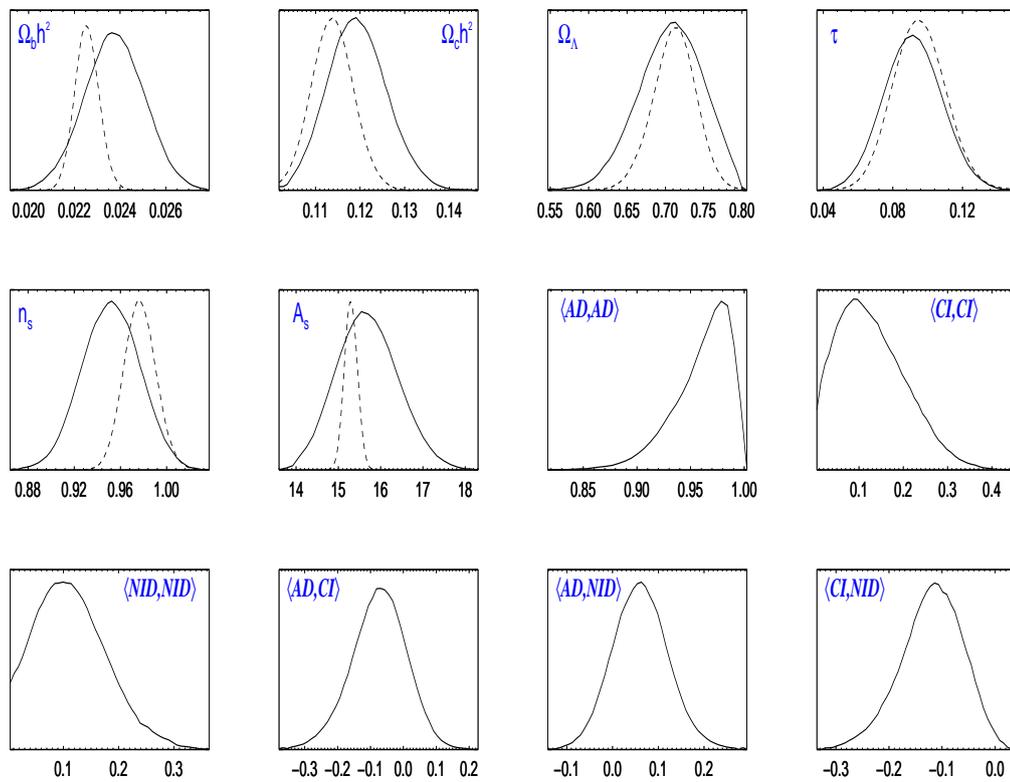


Figure 5.9: Adiabatic plus CI+NID. As in Figure 5.3, with the last six plots in the bottom rows showing the marginal posterior distributions of the auto- and cross-correlations for the amplitude of the modes.

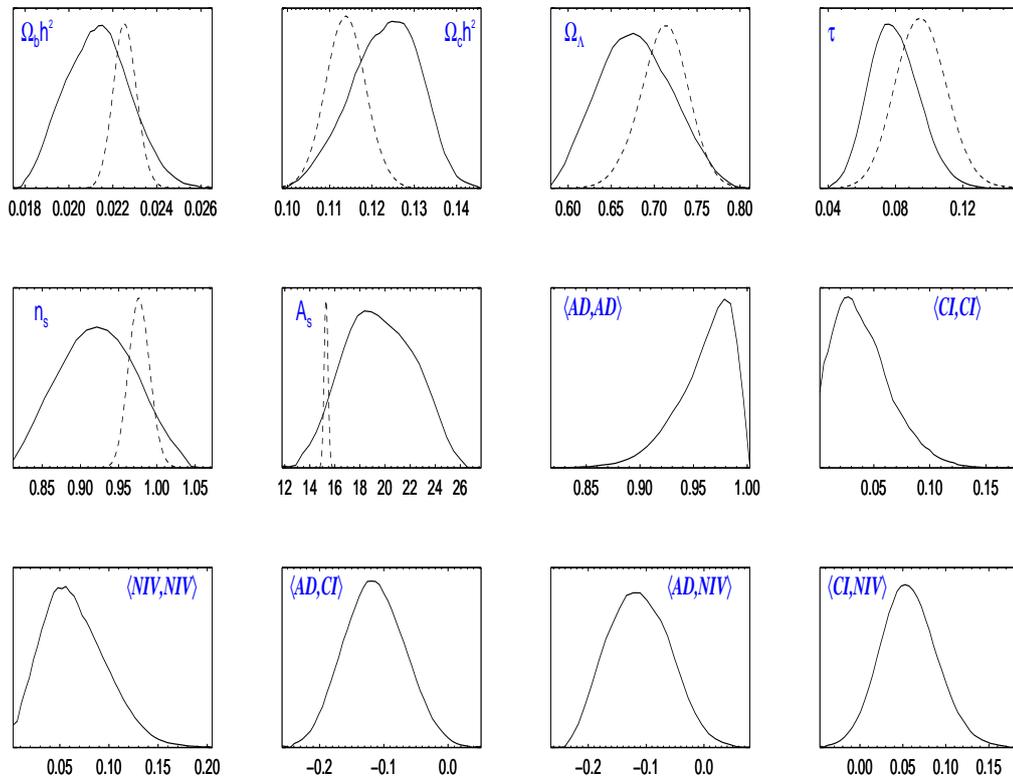


Figure 5.10: Adiabatic plus CI+NIV. Distributions as in Figure 5.9, with CDM density and neutrino velocity isocurvature modes included.

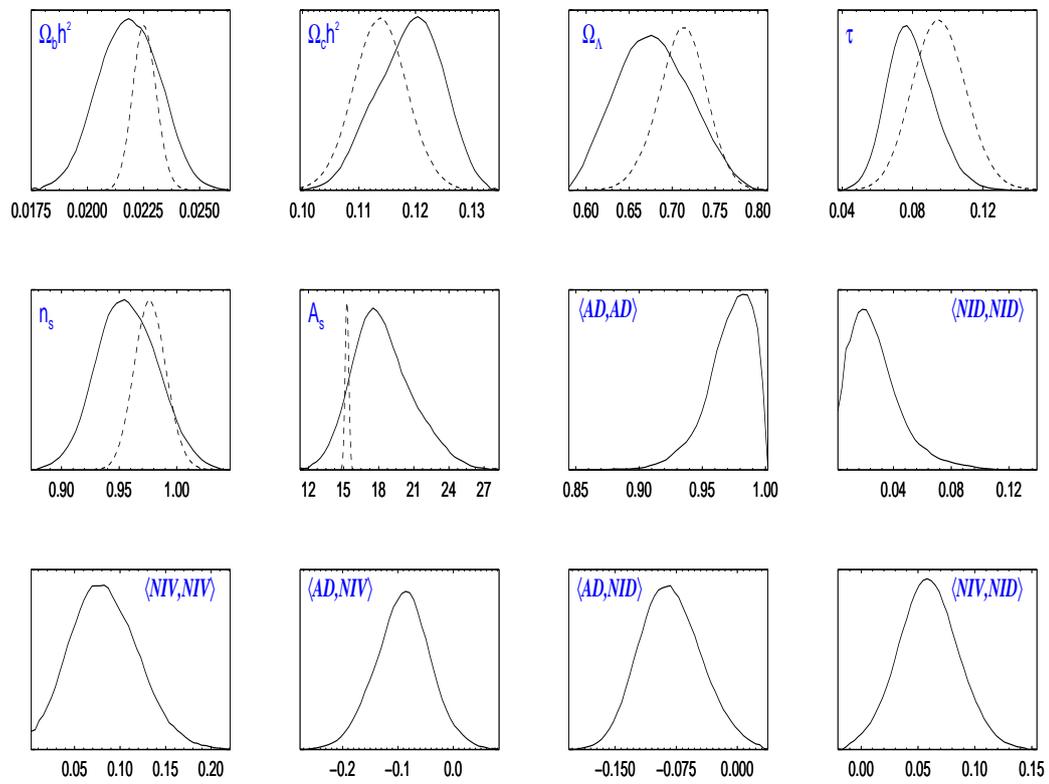


Figure 5.11: Adiabatic plus NID+NIV. Distributions as in Figure 5.9, with the both neutrino density and neutrino velocity isocurvature modes.

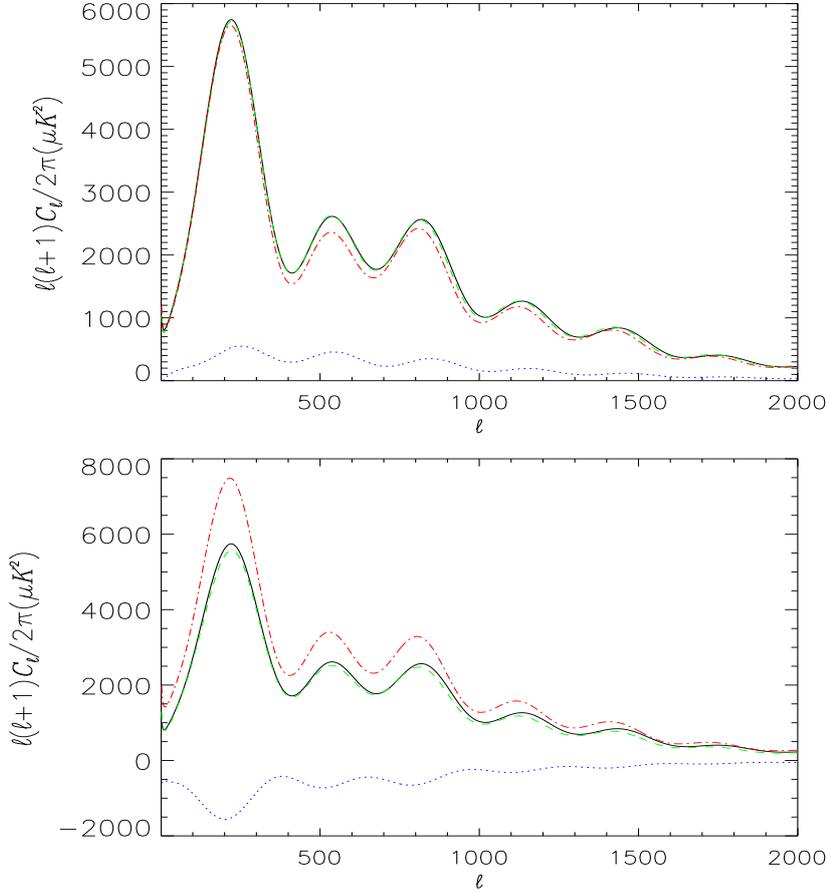


Figure 5.12: *Top*: The CMB angular power spectrum,  $C_\ell^{TT}$ , for the best-fit mixed model (green dashed) for AD+CI+NID overlapping the fiducial adiabatic model spectra (solid black). This model has an  $f_{\text{ISO}} = 27\%$ . The adiabatic contribution is shown by the red dash-dot curve with the non-adiabatic contribution by the blue dotted curve. *Bottom*: The same for models with AD+CI+NIV with a best-fit mixed model with  $f_{\text{ISO}} = 25\%$ .

We plot the  $C_\ell^{TT}$  spectra for the various mode contributions compared to the fiducial adiabatic spectrum in Figure 5.12 to illustrate how well models with high isocurvature fractions can explain the data. The AD+CI+NID model has the greatest isocurvature contribution, with the best-fit models having both positive and negative mode correlations, which cancel each other to produce a small non-adiabatic power, as a result leaving an adiabatic contribution with a power closely matched to the total spectrum. We illustrate a mixed model with parameters  $(\Omega_b h^2, \Omega_c h^2, \Omega_\Lambda, \tau, n_s, A_s, f_{\text{ISO}}) = (0.023, 0.123, 0.685, 0.087, 0.93, 15.5, 0.27)$  for AD+CI+NID (top panel) and for AD+CI+NIV with parameters  $(0.02, 0.132, 0.643, 0.088, 0.89, 22.37, 0.25)$

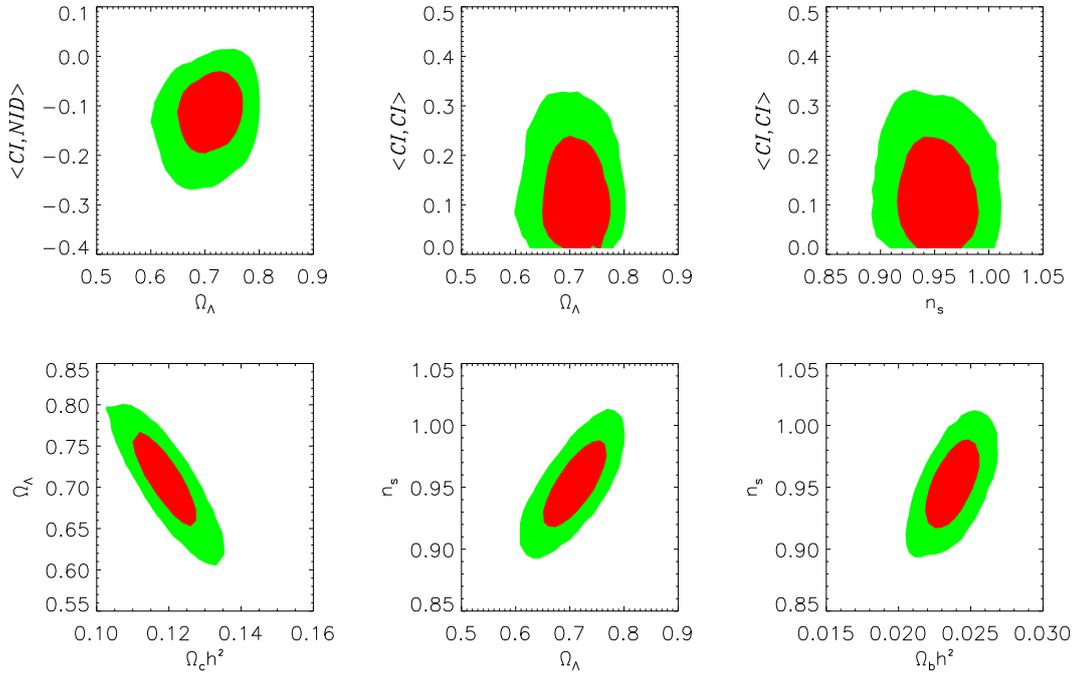


Figure 5.13: Adiabatic plus CI+NID. Selected joint distributions between parameters depicting the most degenerate direction. Contour colours follow the same convention as Figure 5.2.

(bottom panel). For the AD+CI+NIV model, the non-adiabatic contribution is negative which reduces the excess power provided by the adiabatic contribution. This model has negative correlations with the adiabatic mode and illustrates the significance of correlated adiabatic and isocurvature models.

Using principal component analysis, we find that most degeneracies occur between the auto- and cross-correlations for all mixed models. This is in agreement with the degeneracies found using WMAP first-year data [71]. The degree of degeneracy between the  $\langle I, J \rangle$  parameters is shown for selected combinations in figures 5.13 (AD+CI+NID), 5.14 (AD+CI+NIV) and 5.15 (AD+NID+NIV). As in the single mode case, there is also degeneracy between  $n_s$  and  $\Omega_\Lambda$  for all mixed models with a strong positive correlation between the two parameters.

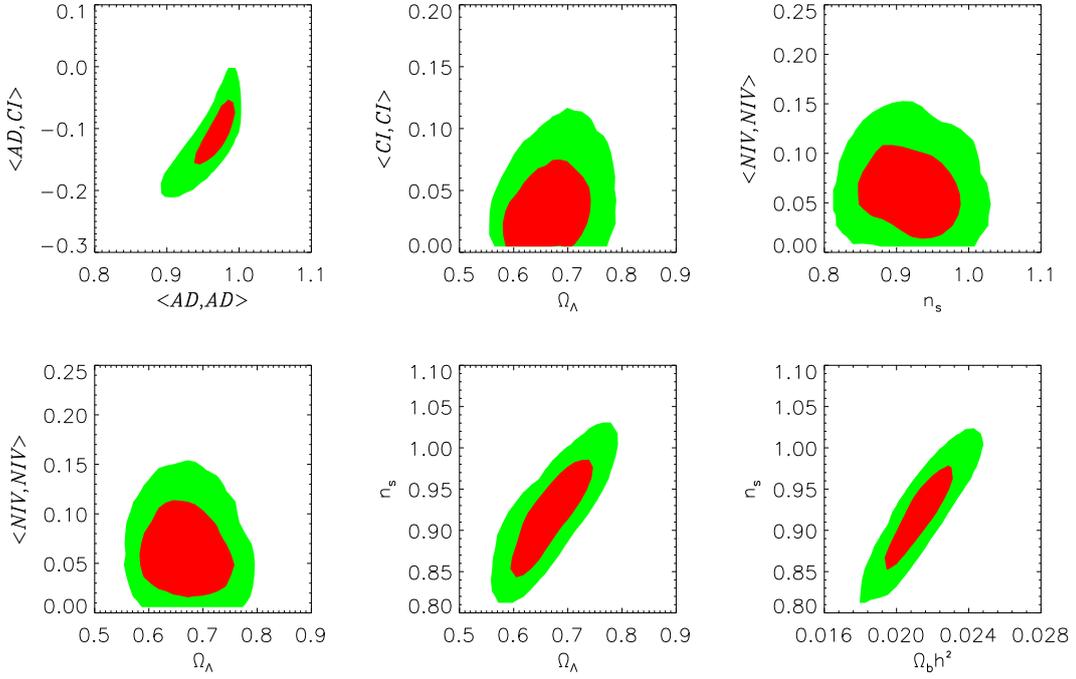


Figure 5.14: Adiabatic plus CI+NIV. Selected joint distributions between parameters depicting the most degenerate direction. Contour colours follow the same convention as Figure 5.2.

As with all mixed models the increase in pure isocurvature contributions induces an increase in  $\Omega_\Lambda$  and  $n_s$ . For the AD+CI+NIV and AD+NIV+NID models, the cosmological constant density and spectral index are positively correlated with  $\Omega_b h^2$  and negatively correlated with  $\Omega_c h^2$ . This effect is evident in the contour plots. For these models,  $n_s$  and  $\Omega_\Lambda$  are shifted to lower values compared to the adiabatic model, thus increasing  $\Omega_c h^2$  and decreasing  $\Omega_b h^2$ . For the AD+CI+NID model however, the baryon density has increased despite no increase in  $\Omega_\Lambda$ . This is due to the relation between  $\langle \text{CI}, \text{NID} \rangle$  and  $\Omega_\Lambda$ , where a negative cross-correlation between CI and NID induces a decrease in the cosmological constant density.

A good example of the complexity of the target distribution is shown in the relation between  $\langle \text{AD}, \text{CI} \rangle$  and  $\langle \text{AD}, \text{AD} \rangle$  in the AD+CI+NIV model. The distribution closely follows a banana

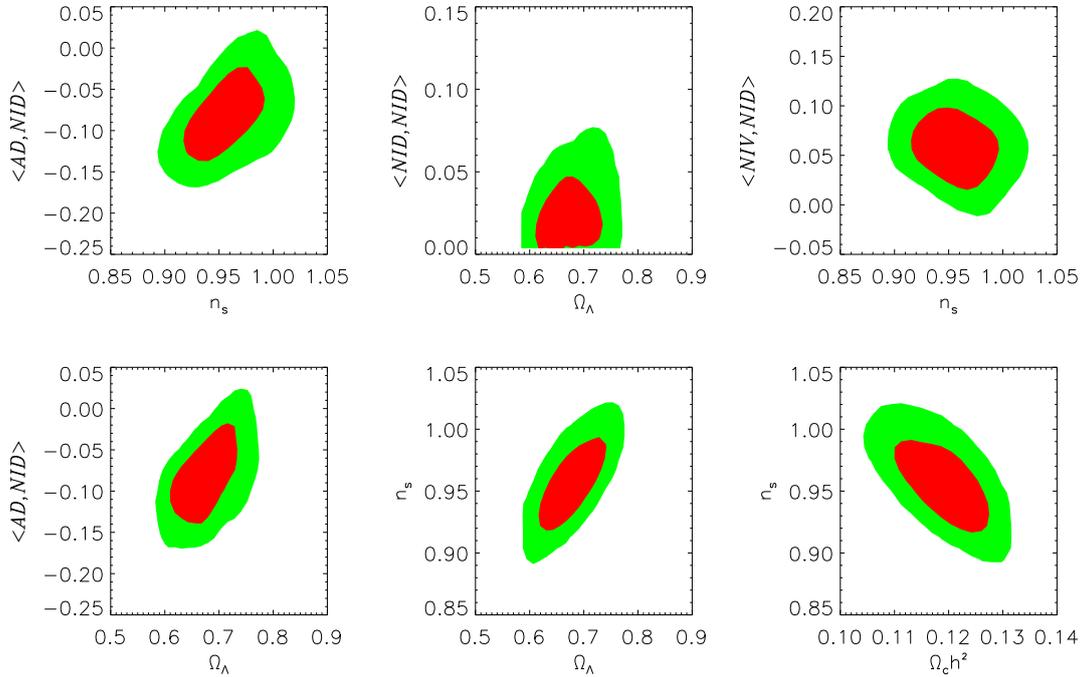


Figure 5.15: Adiabatic plus NIV+NID. Selected joint distributions between parameters depicting the most degenerate direction. Contour colours follow the same convention as Figure 5.2.

shaped distribution as studied in Section 4.5.1. The joint distribution of parameters is very complex, illustrating the need for many components to achieve convergence. The degeneracy amongst parameters increases as we introduce more isocurvature modes. This is most evident in our general mixture model including all isocurvature modes, which we study next.

### 5.3.3 Adiabatic mode correlated with all three isocurvature modes

In this Section we consider the general model with an adiabatic mode correlated with all three isocurvature modes (i.e., CI, NIV and NID). The statistics for the cosmological parameters and relative mode amplitudes are given in Table 5.7. The one-dimensional marginal distributions for all parameters are shown in Figure 5.16.

The adiabatic mode is much weaker with a smaller mean of 0.88 and there are much larger

Parameter	AD+CI+NID+NIV
$\Omega_b h^2$	$0.0252 \pm 0.0011$
$\Omega_c h^2$	$0.117 \pm 0.005$
$\Omega_\Lambda$	$0.73 \pm 0.03$
$\tau$	$0.09 \pm 0.01$
$n_s$	$0.99 \pm 0.02$
$A_s$	$17.29 \pm 0.87$
$Z_{\langle AD, AD \rangle}$	$0.88 \pm 0.02$
$Z_{\langle CI, CI \rangle}$	$0.172 \pm 0.041$
$Z_{\langle NID, NID \rangle}$	$0.185 \pm 0.045$
$Z_{\langle NIV, NIV \rangle}$	$0.08 \pm 0.03$
$Z_{\langle AD, CI \rangle}$	$0.034 \pm 0.075$
$Z_{\langle AD, NID \rangle}$	$-0.022 \pm 0.043$
$Z_{\langle AD, NIV \rangle}$	$-0.005 \pm 0.059$
$Z_{\langle CI, NID \rangle}$	$-0.216 \pm 0.041$
$Z_{\langle CI, NIV \rangle}$	$-0.102 \pm 0.024$
$Z_{\langle NID, NIV \rangle}$	$0.102 \pm 0.036$
$Z_{\text{ISO}}$	$0.495 \pm 0.045$
$f_{\text{ISO}}$	$0.363 \pm 0.028$

Table 5.7: All three isocurvature modes. Statistics of converged samples from PMC. The same conventions as in Table 5.5

isocurvature contributions allowed, with the isocurvature fraction increasing to 36%. The data from WMAP-9 permit a much smaller isocurvature fraction compared to the first-year (60% [71]) and third-year (44% [148]) data. There is a noticeable increase in  $\Omega_b h^2$  although it is less than the lower limit found using WMAP three-year data [148] and WMAP first-year data [71]. The 1-d histogram shows that the CDM density is consistent between adiabatic and mixed model.

The baryon density is no longer in agreement with estimates found in the adiabatic model at the  $1\sigma$  level. Apart from these parameters, all other cosmological parameters are consistent with their values for the adiabatic model. The departure of  $\Omega_b h^2$  from the purely adiabatic model is strongly correlated with larger isocurvature contributions.

In Figure 5.17 we plot the CMB spectrum for a mixed model with a high likelihood and isocurvature fraction. The parameters of the mixed model are  $(\Omega_b h^2, \Omega_c h^2, \Omega_\Lambda, \tau, n_s, A_s, f_{\text{ISO}}) = (0.026, 0.125, 0.679, 0.099, 0.958, 18.22, 0.39)$ . The mixed model is indistinguishable from the adiabatic model. Therefore we plot the error bars of the data instead. We also decompose the spectrum of the mixed model into its adiabatic and isocurvature contributions. We notice  $\langle AD, AD \rangle$  and  $\langle AD, NIV \rangle$  contributions account for most of the total spectrum.

We use principal component analysis to determine which parameters are most degenerate. The parameters that contribute the most to the degenerate direction are (in descending order)  $\langle AD, CI \rangle$ ,  $\langle AD, AD \rangle$ ,  $\langle NID, NID \rangle$ ,  $\langle CI, NID \rangle$ ,  $\langle AD, NID \rangle$  and  $\Omega_\Lambda$ . These parameters collectively contribute to over 90% of the degenerate direction. The WMAP first-year analysis found similar parameters contributing to the degeneracy. We also search for degenerate directions that arise from smaller subsets in order to determine what physical effects are responsible for flat directions. We find the spectral index to be strongly correlated with the baryon and cold dark matter density as in the previous mixed models studied. The two-dimensional marginal distributions illustrating the degeneracies are shown in Figure 5.18.

The introduction of isocurvature modes induces degeneracy amongst the cosmological parameters. Mixed models are capable of generating CMB spectra compatible with a purely adiabatic model, and the data are consistent with these models. This demonstrates that the data permit large amounts of isocurvature. Statistically there is no evidence to prefer this model over the adiabatic model, as there is no further reduction in  $\chi^2$  for these models and they have more free

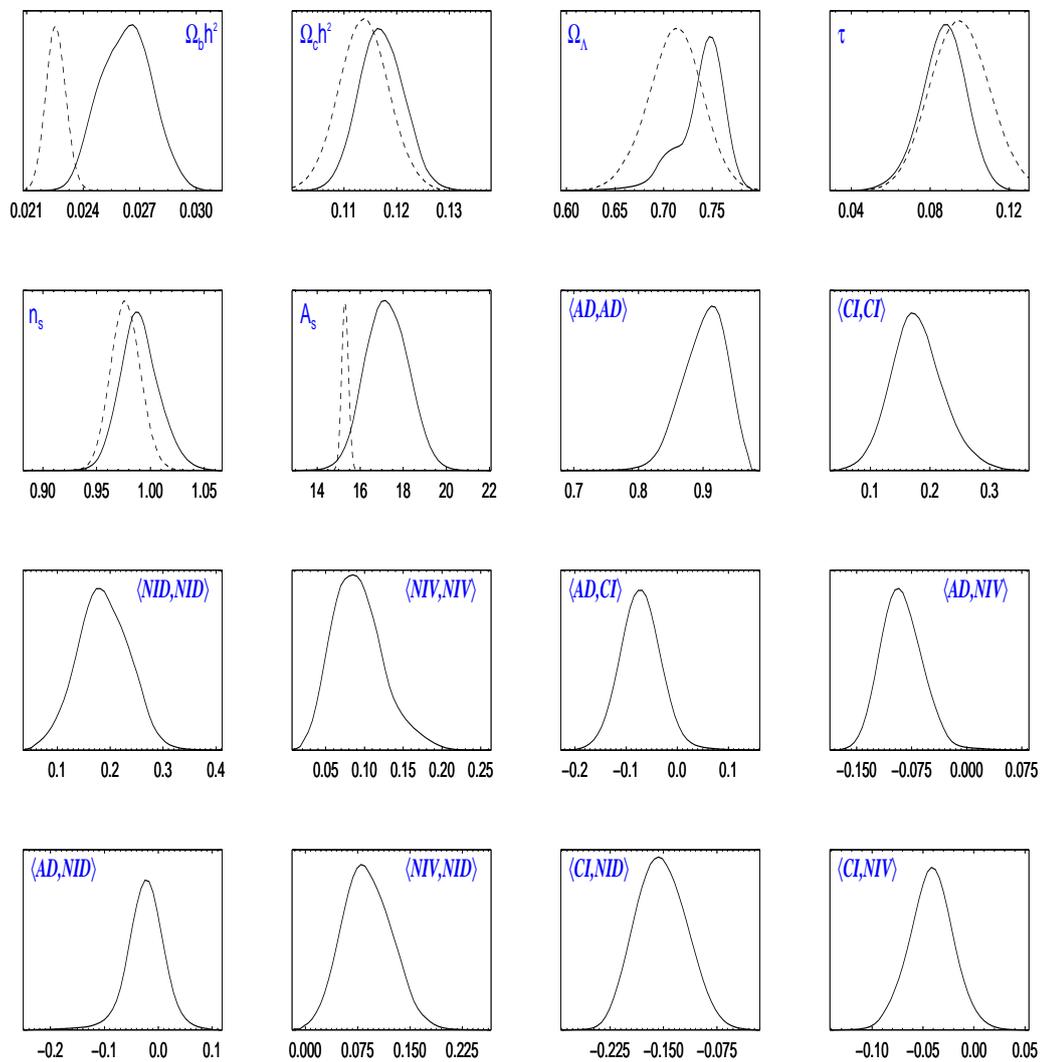


Figure 5.16: Adiabatic plus CI+NID+NIV. Distributions as in Figure 5.9, for models with correlated adiabatic and three isocurvature modes.

parameters. In keeping with the Occam’s razor principle, one would be inclined to choose the simpler adiabatic model that explains the data as well as the more general mixture model.

The constraints on mixed adiabatic and isocurvature models presented here are the most updated

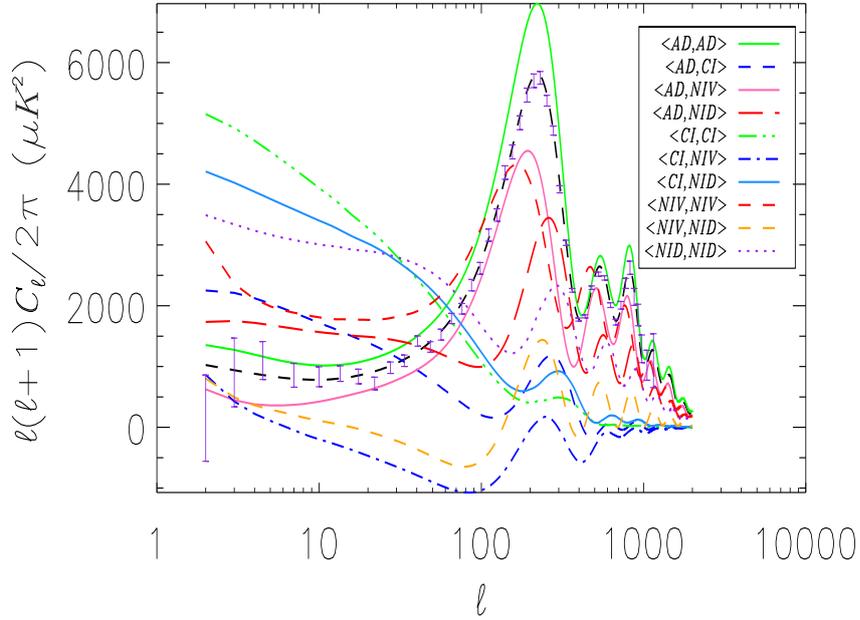


Figure 5.17: Adiabatic plus three isocurvature modes. The angular power spectrum of CMB temperature anisotropies,  $C_\ell$ , for a mixed model with high isocurvature fraction and likelihood (black dashed). The spectrum is a good fit to the data shown by the purple error bars. Errors include instrument noise and uncertainties due to correlated beam and point source subtraction. The spectra for all isocurvature and adiabatic contributions are also shown.

for the adiabatic mode correlated with two or three isocurvature modes. Considering the complex nature of these distributions and the necessity of good choice of the covariance matrix for the PMC, or MCMC sampler, these covariance matrices provide useful starting points for a future study using the most recent Planck data [3, 4].

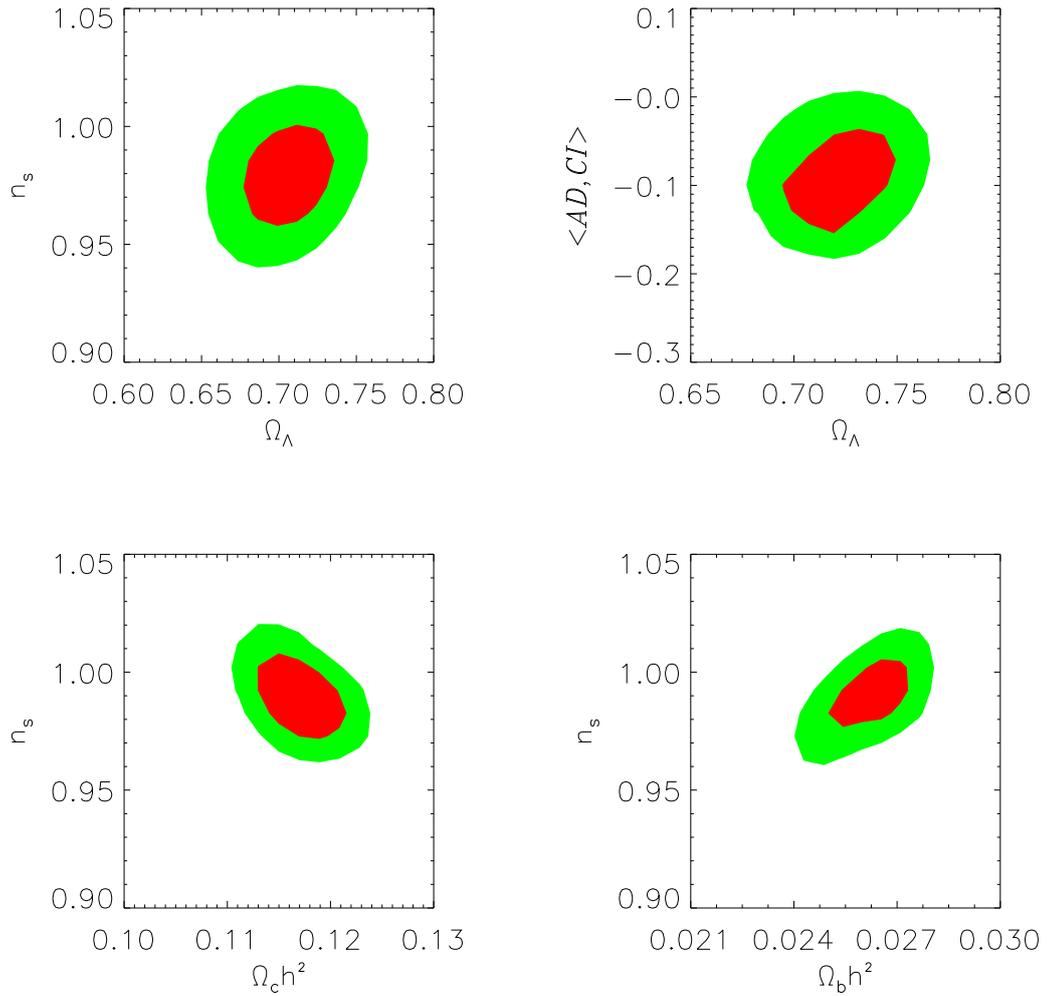


Figure 5.18: Adiabatic plus CI+NID+NIV. Selected two dimensional marginal distributions of degenerate parameters. Contours follow the same convention as Figure 5.2.

## CHAPTER 6

---

### Conclusion

---

The amount of available cosmological data is increasing at a very fast pace and the precision of this data is unprecedented. The WMAP team has made excellent data available for cosmology with the recent Planck experiment building on this progress. The new data require faster and more efficient methods of analysis. Therefore it is worth investigating alternate sampling algorithms that can be easily run in parallel with straight-forward diagnostics for convergence. Large clusters of slower CPUs are easier and cheaper to attain than a single up-to-date supercomputer. Therefore, parallel processing is the way forward in sampling data. The PMC sampler is easily parallelised and has convenient methods to assess convergence. This assessment can be automated, but the algorithm has to be initialised according to a specific target distribution to optimise its performance. In this thesis we established various methods for optimising the performance of this algorithm for various target distributions. For the case of a Gaussian target distribution we found that the sampler was optimised with as little as two components in all dimensions and the optimal sample size scaled quadratically with dimension. Prior information

about the target distribution was useful to maximise efficiency, since the mixture densities could be initialised to sufficiently cover all regions of parameter space within a few iterations.

We also investigated the performance of the PMC algorithm on more complex distributions, namely, the banana-shaped, bimodal, and hypercube distribution. We defined a measure of complexity,  $B$ , for the banana-shaped distribution that enabled us to maximise efficiency with respect to the parameter through simulations. The complex nature of the banana-shaped distribution forced PMC to use more components to adequately sample the degenerate parameter space. The number of components required for convergence increased with complexity, and we found efficiency decreased exponentially with increased  $B$ . The bimodal distribution provided a different challenge to the PMC algorithm in that the distance between modes had a significant effect on efficiency, especially when these modes were separated by large regions of low density. For this case, the algorithm only sampled well the region around one mode if components were not carefully positioned to cover the entire parameter space. However, if components were well positioned, or the modes were closely situated, then PMC achieved convergence within a few iterations. For the challenging hypercube distribution the algorithm had to use a relatively large number of components to achieve convergence, with the number of components becoming impractically large as the dimensionality of the hypercube distribution increased.

We thereafter applied our results from the simulations to a cosmological parameter estimation problem, namely, using CMB data from the WMAP nine-year release to constrain admixtures of adiabatic and isocurvature perturbations. We showed that the six parameters of the pure adiabatic model can be sampled with as little as two components and a sample size equivalent to the optimal sample size found for a Gaussian target distribution of six dimensions. The mixed models studied included an adiabatic mode correlated with one, two, and three isocurvature modes.

These models are significantly complex with degenerate parameter spaces, with the underlying distributions requiring a larger number of components to achieve convergence. Using the PMC algorithm on admixture models containing two or more isocurvature modes caused significant degeneracies amongst the importance weights, which required us to implement the Non-linear Population Monte Carlo algorithm (NPMC) instead. The admixture model with the adiabatic mode and all three isocurvature modes was the most challenging of all models and numerous updates were required before convergence was reached.

Our constraints derived for the single isocurvature mode model were consistent with a similar study performed using WMAP nine-year data. The results for admixture models with two or three isocurvature modes are new and indicate a smaller allowed isocurvature fraction compared to previous studies using earlier WMAP data releases. The covariance matrices derived in this thesis for the various admixture models using the WMAP nine-year dataset will serve as useful prior information for updated constraints on isocurvature perturbations with future datasets. Specifically we expect that the higher precision CMB temperature and polarisation power spectra measured by Planck will yield stronger constraints on isocurvature perturbations. The results derived here will be a useful starting point for such a future study.

---

## Bibliography

---

- [1] C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, C. Barnes, M. R. Greason, R. S. Hill, E. Komatsu, M. R. Nolta, N. Odegard, H. V. Peiris, L. Verde, and J. L. Weiland, “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results,” *Astrophys. J. Supp.*, vol. 148, pp. 1–27, Sept. 2003, astro-ph/0302207.
- [2] “Wilkinson Microwave Anisotropy Probe site.” <http://map.gsfc.nasa.gov/site/>.
- [3] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., “Planck 2013 results. XVI. Cosmological parameters,” *ArXiv e-prints*, Mar. 2013, 1303.5076.
- [4] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al.,

- “Planck 2013 results. I. Overview of products and scientific results,” *ArXiv e-prints*, Mar. 2013, 1303.5062.
- [5] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry, “Observational evidence from supernovae for an accelerating universe and a cosmological constant,” *The Astronomical Journal*, vol. 116, no. 3, p. 1009, 1998.
- [6] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project, “Measurements of  $\Omega_m$  and  $\Omega_\Lambda$  from 42 high-redshift supernovae,” *The Astrophysical Journal*, vol. 517, no. 2, p. 565, 1999.
- [7] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon, T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny, and D. G. York, “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies,” *The Astrophysical Journal*, vol. 633, pp. 560–574, Nov. 2005, astro-ph/0501171.

- [8] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, “A Direct Empirical Proof of the Existence of Dark Matter,” *The Astrophysical Journal Letters*, vol. 648, pp. L109–L113, Sept. 2006, astro-ph/0608407.
- [9] P. J. Peebles and B. Ratra, “The cosmological constant and dark energy,” *Reviews of Modern Physics*, vol. 75, pp. 559–606, Apr. 2003, astro-ph/0207347.
- [10] D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, “Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology,” *The Astrophysical Journal Supplement*, vol. 170, pp. 377–408, June 2007, astro-ph/0603449.
- [11] J. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multi-scale Approach*. Cambridge University Press, 1998.
- [12] J. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. Astronomy and Astrophysics Library, Springer, 2007.
- [13] V. Martinez and E. Saar, *Statistics of the Galaxy Distribution*. Taylor & Francis, 2001.
- [14] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge University Press.
- [15] G. Babu and E. Feigelson, *Astrostatistics*. Chapman & Hall/CRC Interdisciplinary Statistics, Taylor & Francis, 1996.
- [16] R. Trotta, “Bayes in the sky: Bayesian inference and model selection in cosmology,” *Contemporary Physics*, vol. 49, pp. 71–104, Mar. 2008, 0803.4089.

- [17] M. Bayes and M. Price, “An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.,” *Philosophical Transactions*, vol. 53, pp. 370–418, 1763, <http://rstl.royalsocietypublishing.org/content/53/370.full.pdf+html>.
- [18] D. Lindley, “Kendall’s advanced theory of statistics, volume 2b, bayesian inference, 2nd edn,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 168, no. 1, pp. 259–260, 2005.
- [19] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*. Oxford science publications, OUP Oxford, 2006.
- [20] C. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer texts in statistics, Springer, 2007.
- [21] J. Skilling, “Nested Sampling,” in *American Institute of Physics Conference Series* (R. Fischer, R. Preuss, and U. V. Toussaint, eds.), vol. 735 of *American Institute of Physics Conference Series*, pp. 395–405, Nov. 2004.
- [22] P. Mukherjee, D. Parkinson, and A. R. Liddle, “A Nested Sampling Algorithm for Cosmological Model Selection,” *Astrophysical Journal Letters*, vol. 638, pp. L51–L54, 2006, arXiv:astro-ph/0508461.
- [23] D. Parkinson, P. Mukherjee, and A. R. Liddle, “Bayesian model selection analysis of WMAP3,” *Physical review D*, vol. 73, no. 12, p. 123523, 2006, arXiv:astro-ph/0605003.
- [24] B. Nikolic, “Fitting and Comparison of Models of Radio Spectra,” *ArXiv e-prints*, 2009, arXiv:0912.2317.

- [25] J. Väiviita and T. Giannantonio, “Constraints on primordial isocurvature perturbations and spatial curvature by Bayesian model selection,” *Physical review D*, vol. 80, no. 12, p. 123516, 2009, arXiv:0909.5190.
- [26] F. Feroz, M. P. Hobson, and M. Bridges, “MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics,” *Monthly Notices of the Royal Astrophysical Society*, vol. 398, pp. 1601–1614, 2009, arXiv:0809.3437.
- [27] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, June 1953.
- [28] A. Lewis and S. Bridle, “Cosmological parameters from cmb and other data: A monte carlo approach,” *Phys. Rev. D*, vol. 66, p. 103511, Nov 2002.
- [29] M. S. Oh and J. O. Berger, “Adaptive importance sampling in Monte Carlo integration,” *Journal of Statistical Computation and Simulation*, vol. 41, pp. 143–168, 1992.
- [30] O. Capp, A. Guillin, J. M. Marin, and C. P. Robert, “Population monte carlo,” *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. pp. 907–929, 2004.
- [31] D. Wraith, M. Kilbinger, K. Benabed, O. Cappé, J. Cardoso, G. Fort, S. Prunet, and C. P. Robert, “Estimation of cosmological parameters using adaptive importance sampling,” *Physical Review D*, vol. 80, p. 023507, July 2009, 0903.0837.
- [32] J. P. Dietrich and J. Hartlap, “Cosmology with the shear-peak statistics,” *Monthly Notices of the Royal Astrophysical Society*, vol. 402, pp. 1049–1058, Feb. 2010, 0906.3512.
- [33] M. Kilbinger, D. Wraith, C. P. Robert, K. Benabed, O. Cappé, J. Cardoso, G. Fort, S. Prunet, and F. R. Bouchet, “Bayesian model comparison in cosmology with Population

- Monte Carlo,” *Monthly Notices of the Royal Astrophysical Society*, vol. 405, pp. 2381–2390, July 2010, 0912.1614.
- [34] M. Kilbinger, K. Benabed, O. Cappe, J.-F. Cardoso, G. Fort, S. Prunet, C. P. Robert, and D. Wraith, “CosmoPMC: Cosmology Population Monte Carlo,” *ArXiv e-prints*, Jan. 2011, 1101.0950.
- [35] J. Dunkley, M. Bucher, P. G. Ferreira, K. Moodley, and C. Skordis, “Fast and reliable Markov chain Monte Carlo technique for cosmological parameter estimation,” *Monthly Notices of the Royal Astrophysical Society*, vol. 356, pp. 925–936, Jan. 2005, arXiv:astro-ph/0405462.
- [36] J. Dunkley, E. Komatsu, M. R. Nolta, D. N. Spergel, D. Larson, G. Hinshaw, L. Page, C. L. Bennett, B. Gold, N. Jarosik, J. L. Weiland, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, “Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters from the WMAP Data,” *The Astrophysical Journal Supplement Series*, vol. 180, no. 2, p. 306, 2009.
- [37] “The Sloan Digital Sky Survey Home Page.” <http://www.sdss.org/>.
- [38] M. Tegmark, M. A. Strauss, M. R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D. H. Weinberg, I. Zehavi, N. A. Bahcall, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, A. Berlind, T. Budavari, A. Connolly, D. J. Eisenstein, D. Finkbeiner, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R. C. Nichol, J. P. Ostriker, A. Pope, R. Scranton, U. Seljak, R. K. Sheth, A. Stebbins, A. S. Szalay, I. Szapudi, Y. Xu, J. Annis, J. Brinkmann, S. Burles, F. J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, B. Gillespie, G. Hennessey, D. W. Hogg, Ž. Ivezić, G. R. Knapp, D. Q. Lamb, B. C. Lee, R. H. Lupton, T. A. McKay, P. Kunszt, J. A.

- Munn, L. O’Connell, J. Peoples, J. R. Pier, M. Richmond, C. Rockosi, D. P. Schneider, C. Stoughton, D. L. Tucker, D. E. vanden Berk, B. Yanny, and D. G. York, “Cosmological parameters from SDSS and WMAP,” *Physical Review D.*, vol. 69, p. 103501, May 2004, astro-ph/0310723.
- [39] A. Veropalumbo, F. Marulli, L. Moscardini, M. Moresco, and A. Cimatti, “An improved measurement of baryon acoustic oscillations from the correlation function of galaxy clusters at  $z \sim 0.3$ ,” *Mon. Not. R. Astron. Soc.*, vol. 442, pp. 3275–3283, Aug. 2014, 1311.5895.
- [40] H. Alavirad and A. Sheykhi, “Cosmological constraints on ghost dark energy in the Brans-Dicke theory by using MCMC approach,” *Physics Letters B*, vol. 734, pp. 148–156, June 2014, 1405.2515.
- [41] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R.olta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results,” *The Astrophysical Journal Supplement Series*, vol. 208, no. 2, p. 19, 2013.
- [42] P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, E. Battaner, K. Benabed, A. Benoit-Lévy, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. R. Bond, J. Borrill, F. R. Bouchet, C. Burigana, R. C. Butler, E. Calabrese, A. Chamballu, H. C. Chiang, P. R. Christensen, D. L. Clements, L. P. L. Colombo, F. Couchot, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Diego, H. Dole,

O. Doré, X. Dupac, T. A. Enßlin, H. K. Eriksen, O. Fabre, F. Finelli, O. Forni, M. Frailis, E. Franceschi, S. Galeotta, S. Galli, K. Ganga, M. Giard, J. González-Nuevo, K. M. Górski, A. Gregorio, A. Gruppuso, F. K. Hansen, D. Hanson, D. L. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, K. M. Huffenberger, A. H. Jaffe, W. C. Jones, E. Keihänen, R. Keskitalo, R. Kneissl, J. Knoche, M. Kunz, H. Kurki-Suonio, J.-M. Lamarre, A. Lasenby, C. R. Lawrence, R. Leonardi, J. Lesgourgues, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, N. Mandolesi, M. Maris, P. G. Martin, E. Martínez-González, S. Masi, S. Matarrese, P. Mazzotta, P. R. Meinhold, A. Melchiorri, L. Mendes, E. Menegoni, A. Mennella, M. Migliaccio, M.-A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, L. Pagano, F. Pajot, D. Paoletti, F. Pasian, G. Patanchon, O. Perdereau, L. Perotto, F. Perrotta, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, G. W. Pratt, S. Prunet, J. P. Rachen, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, S. Ricciardi, I. Ristorcelli, G. Rocha, G. Roudier, B. Rusholme, M. Sandri, G. Savini, D. Scott, L. D. Spencer, V. Stolyarov, R. Sudiwala, D. Sutton, A.-S. Suur-Uski, J.-F. Sygnet, J. A. Tauber, D. Tavagnacco, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J.-P. Uzan, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, L. A. Wade, D. Yvon, A. Zacchei, and A. Zonca, “Planck intermediate results. XXIV. Constraints on variation of fundamental constants,” *ArXiv e-prints*, June 2014, 1406.7482.

- [43] B. Novosyadlyj, O. Sergijenko, R. Durrer, and V. Pelykh, “Constraining the dynamical dark energy parameters: Planck-2013 vs WMAP9,” *Journal of Cosmology and Astropar-*

- ticle Physics.*, vol. 5, p. 30, May 2014, 1312.6579.
- [44] C. Heymans, E. Grocutt, A. Heavens, M. Kilbinger, T. D. Kitching, F. Simpson, J. Benjamin, T. Erben, H. Hildebrandt, H. Hoekstra, Y. Mellier, L. Miller, L. Van Waerbeke, M. L. Brown, J. Coupon, L. Fu, J. Harnois-Déraps, M. J. Hudson, K. Kuijken, B. Rowe, T. Schrabback, E. Semboloni, S. Vafaei, and M. Velander, “CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments,” *Mon. Not. R. Astron. Soc.*, vol. 432, pp. 2433–2453, July 2013, 1303.1808.
- [45] M. Kilbinger, L. Fu, C. Heymans, F. Simpson, J. Benjamin, T. Erben, J. Harnois-Déraps, H. Hoekstra, H. Hildebrandt, T. D. Kitching, Y. Mellier, L. Miller, L. Van Waerbeke, K. Benabed, C. Bonnett, J. Coupon, M. J. Hudson, K. Kuijken, B. Rowe, T. Schrabback, E. Semboloni, S. Vafaei, and M. Velander, “CFHTLenS: combined probe cosmological model comparison using 2D weak gravitational lensing,” *Mon. Not. R. Astron. Soc.*, vol. 430, pp. 2200–2220, Apr. 2013, 1212.3338.
- [46] K. Enqvist, H. Kurki-Suonio, and J. Väliiviita, “Limits on isocurvature fluctuations from Boomerang and MAXIMA,” *Physical Review D.*, vol. 62, p. 103003, Nov. 2000, astro-ph/0006429.
- [47] R. Stompor, A. J. Banday, and K. M. Gorski, “Flat Dark Matter–dominated Models with Hybrid Adiabatic Plus Isocurvature Initial Conditions,” *Astrophys. J.*, vol. 463, p. 8, May 1996, astro-ph/9511087.
- [48] W. Hu, E. F. Bunn, and N. Sugiyama, “COBE Constraints on Baryon Isocurvature Models,” *Astrophys. J. Lett.*, vol. 447, p. L59, July 1995, astro-ph/9501034.
- [49] W. Hu and N. Sugiyama, “Thermal history constraints on the isocurvature baryon model,” *Astrophys. J.*, vol. 436, pp. 456–466, Dec. 1994, astro-ph/9403031.

- [50] T. Chiba, N. Sugiyama, and Y. Suto, “Microwave background anisotropies in primeval isocurvature baryon models: Constraints on the cosmological parameters,” *Astrophys. J.*, vol. 429, pp. 427–433, July 1994, astro-ph/9311012.
- [51] E. L. Wright, S. S. Meyer, C. L. Bennett, N. W. Boggess, E. S. Cheng, M. G. Hauser, A. Kogut, C. Lineweaver, J. C. Mather, G. F. Smoot, R. Weiss, S. Gulkis, G. Hinshaw, M. Janssen, T. Kelsall, P. M. Lubin, S. H. Moseley, Jr., T. L. Murdock, R. A. Shafer, R. F. Silverberg, and D. T. Wilkinson, “Interpretation of the cosmic microwave background radiation anisotropy detected by the COBE Differential Microwave Radiometer,” *Astrophys. J. Lett.*, vol. 396, pp. L13–L18, Sept. 1992.
- [52] P. de Bernardis, P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, K. Coble, B. P. Crill, G. De Gasperis, P. C. Farese, P. G. Ferreira, K. Ganga, M. Giacometti, E. Hivon, V. V. Hristov, A. Iacoangeli, A. H. Jaffe, A. E. Lange, L. Martinis, S. Masi, P. V. Mason, P. D. Mauskopf, A. Melchiorri, L. Miglio, T. Montroy, C. B. Netterfield, E. Pascale, F. Piacentini, D. Pogosyan, S. Prunet, S. Rao, G. Romeo, J. E. Ruhl, F. Scaramuzzi, D. Sforna, and N. Vittorio, “A flat Universe from high-resolution maps of the cosmic microwave background radiation,” *Nature.*, vol. 404, pp. 955–959, Apr. 2000, astro-ph/0004404.
- [53] S. Hanany, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. G. Ferreira, V. V. Hristov, A. H. Jaffe, A. E. Lange, A. T. Lee, P. D. Mauskopf, C. B. Netterfield, S. Oh, E. Pascale, B. Rabbii, P. L. Richards, G. F. Smoot, R. Stompor, C. D. Winant, and J. H. P. Wu, “MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of  $10^\circ - 5^\circ$ ,” *Astrophys. J. Lett.*, vol. 545, pp. L5–L9, Dec. 2000, astro-ph/0005123.
- [54] L. Amendola, C. Gordon, D. Wands, and M. Sasaki, “Correlated Perturbations from Infla-

- tion and the Cosmic Microwave Background,” *Physical Review Letters*, vol. 88, p. 211302, May 2002, astro-ph/0107089.
- [55] E. Pierpaoli, J. García-Bellido, and S. Borgani, “Microwave background anisotropies and large scale structure constraints on isocurvature modes in a two-field model of inflation,” *Journal of High Energy Physics*, vol. 10, p. 15, Oct. 1999.
- [56] C. Gordon and A. Lewis, “Observational constraints on the curvaton model of inflation,” *Physical Review D.*, vol. 67, p. 123513, June 2003, astro-ph/0212248.
- [57] D. Langlois and A. Riazuelo, “Correlated mixtures of adiabatic and isocurvature cosmological perturbations,” *Physical Review D.*, vol. 62, p. 043504, Aug. 2000, astro-ph/9912497.
- [58] R. Trotta, A. Riazuelo, and R. Durrer, “Cosmological constant and general isocurvature initial conditions,” *Physical Review D.*, vol. 67, p. 063520, Mar. 2003, astro-ph/0211600.
- [59] R. Trotta, “The cosmological constant and the paradigm of adiabaticity,” *New Astronomy Reviews.*, vol. 47, pp. 769–774, Nov. 2003, astro-ph/0304525.
- [60] G. Hinshaw, D. N. Spergel, L. Verde, R. S. Hill, S. S. Meyer, C. Barnes, C. L. Bennett, M. Halpern, N. Jarosik, A. Kogut, E. Komatsu, M. Limon, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: The Angular Power Spectrum,” *The Astrophysical Journal Supplement Series*, vol. 148, no. 1, p. 135, 2003.
- [61] M. Beltrán, J. García-Bellido, J. Lesgourgues, A. R. Liddle, and A. Slosar, “Bayesian model selection and isocurvature perturbations,” *Physical Review D.*, vol. 71, p. 063532, Mar. 2005, astro-ph/0501477.

- [62] M. Beltrán, J. García-Bellido, J. Lesgourgues, and M. Viel, “Squeezing the window on isocurvature modes with the Lyman- $\alpha$  forest,” *Physical Review D.*, vol. 72, p. 103515, Nov. 2005, astro-ph/0509209.
- [63] H. Kurki-Suonio, V. Muhonen, and J. Väliviita, “Correlated primordial perturbations in light of CMB and large scale structure data,” *Physical Review D.*, vol. 71, p. 063005, Mar. 2005, astro-ph/0412439.
- [64] D. Parkinson, S. Tsujikawa, B. A. Bassett, and L. Amendola, “Testing for double inflation with WMAP,” *Physical Review D.*, vol. 71, p. 063524, Mar. 2005, astro-ph/0409071.
- [65] G. Lazarides, R. Ruiz de Austri, and R. Trotta, “Constraints on a mixed inflaton and curvaton scenario for the generation of the curvature perturbation,” *Physical Review D.*, vol. 70, p. 123527, Dec. 2004, hep-ph/0409335.
- [66] F. Ferrer, S. Räsänen, and J. Väliviita, “Correlated isocurvature perturbations from mixed inflaton curvaton decay,” *Journal of Cosmology and Astroparticle Physics.*, vol. 10, p. 10, Oct. 2004, astro-ph/0407300.
- [67] J. Väliviita and V. Muhonen, “Correlated Adiabatic and Isocurvature Cosmic Microwave Background Fluctuations in the Wake of the Results from the Wilkinson Microwave Anisotropy Probe,” *Physical Review Letters*, vol. 91, p. 131302, Sept. 2003, astro-ph/0304175.
- [68] A. P. Andrade, C. A. Wuensche, and A. L. Ribeiro, “Correlated mixture between adiabatic and isocurvature fluctuations and recent CMB observations,” *Physical Review D.*, vol. 71, p. 043501, Feb. 2005, astro-ph/0501399.

- [69] M. Beltrán, J. García-Bellido, J. Lesgourgues, and A. Riazuelo, “Bounds on cold dark matter and neutrino isocurvature perturbations from CMB and LSS data,” *Physical Review D*, vol. 70, p. 103530, Nov. 2004, astro-ph/0409326.
- [70] J. Valiviita, “Correlated adiabatic and isocurvature CMB fluctuations in the light of the WMAP data,” *ArXiv Astrophysics e-prints*, Oct. 2003, astro-ph/0310206.
- [71] K. Moodley, M. Bucher, J. Dunkley, P. G. Ferreira, and C. Skordis, “Constraints on isocurvature models from the WMAP first-year data,” *Physical Review D*, vol. 70, p. 103520, Nov. 2004, astro-ph/0407304.
- [72] A. P. A. Andrade, C. A. Wuensche, and A. L. B. Ribeiro, “Predictions of Mixed Non-Gaussian Cosmological Density Fields for the Cosmic Microwave Background Radiation,” *Astrophys. J.*, vol. 602, pp. 555–564, Feb. 2004, astro-ph/0312028.
- [73] C. Gordon and K. A. Malik, “WMAP, neutrino degeneracy, and non-Gaussianity constraints on isocurvature perturbations in the curvaton model of inflation,” *Physical Review D*, vol. 69, p. 063508, Mar. 2004, astro-ph/0311102.
- [74] P. Crotty, J. García-Bellido, J. Lesgourgues, and A. Riazuelo, “Bounds on Isocurvature Perturbations from Cosmic Microwave Background and Large Scale Structure Data,” *Physical Review Letters*, vol. 91, p. 171301, Oct. 2003, astro-ph/0306286.
- [75] H. V. Peiris, E. Komatsu, L. Verde, D. N. Spergel, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, E. Wollack, and E. L. Wright, “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications For Inflation,” *Astrophys. J. Supp.*, vol. 148, pp. 213–231, Sept. 2003, astro-ph/0302225.

- [76] M. R. Nolta, J. Dunkley, R. S. Hill, G. Hinshaw, E. Komatsu, D. Larson, L. Page, D. N. Spergel, C. L. Bennett, B. Gold, N. Jarosik, N. Odegard, J. L. Weiland, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright, “Five-Year Wilkinson Microwave Anisotropy Probe Observations: Angular Power Spectra,” *Astrophys. J. Supp.*, vol. 180, pp. 296–305, Feb. 2009, 0803.0593.
- [77] “SNLS - Supernovae Legacy Survey Home Page.” <http://cfht.hawaii.edu/SNLS/>.
- [78] I. Sollom, A. Challinor, and M. P. Hobson, “Cold dark matter isocurvature perturbations: Constraints and model selection,” *Physical Review D.*, vol. 79, p. 123521, June 2009, 0903.5257.
- [79] J. Väliviita, M. Savelainen, M. Talvitie, H. Kurki-Suonio, and S. Rusak, “Constraints on Scalar and Tensor Perturbations in Phenomenological and Two-field Inflation Models: Bayesian Evidences for Primordial Isocurvature and Tensor Modes,” *Astrophys. J.*, vol. 753, p. 151, July 2012, 1202.2852.
- [80] M. Savelainen, J. Väliviita, P. Walia, S. Rusak, and H. Kurki-Suonio, “Constraints on neutrino density and velocity isocurvature modes from WMAP-9 data,” *Physical Review D*, vol. 88, p. 063010, Sept. 2013, 1307.4398.
- [81] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., “Planck 2013 results. XXII. Constraints on inflation,” *ArXiv e-prints*, Mar. 2013, 1303.5082.
- [82] E. Hubble, “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae,” *Proceedings of the National Academy of Science*, vol. 15, pp. 168–173, Mar. 1929.

- [83] A. Raychaudhuri, “Relativistic Cosmology. I,” *Physical Review*, vol. 98, pp. 1123–1126, May 1955.
- [84] E. W. Kolb, D. J. H. Chung, and A. Riotto, “WIMPZILLAS!,” in *Dark matter in Astrophysics and Particle Physics* (H. V. Klapdor-Kleingrothaus and L. Baudis, eds.), p. 592, 1999, hep-ph/9810361.
- [85] S. Scopel, “Particle dark matter candidates,” *Journal of Physics Conference Series*, vol. 120, p. 042003, July 2008, 0711.2870.
- [86] F. Zwicky, “On the Masses of Nebulae and of Clusters of Nebulae,” *The Astrophysical Journal*, vol. 86, p. 217, Oct. 1937.
- [87] F. Zwicky, “Die Rotverschiebung von extragalaktischen Nebeln,” *Helvetica Physica Acta*, vol. 6, pp. 110–127, 1933.
- [88] A. A. Penzias and R. W. Wilson, “A Measurement of Excess Antenna Temperature at 4080 Mc/s.,” *Astrophysical Journal*, vol. 142, pp. 419–421, July 1965.
- [89] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson, “Structure in the COBE differential microwave radiometer first-year maps,” *Astrophysical Journal Letters*, vol. 396, pp. L1–L5, Sept. 1992.
- [90] H. Bondi and T. Gold, “The Steady-State Theory of the Expanding Universe,” *Mon. Not. R. Astron. Soc.*, vol. 108, p. 252, 1948.

- [91] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright, “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results,” *The Astrophysical Journal Supplement Series*, vol. 208, no. 2, p. 20, 2013.
- [92] J. C. Mather, E. S. Cheng, D. A. Cottingham, R. E. Eplee, Jr., D. J. Fixsen, T. Hewagama, R. B. Isaacman, K. A. Jensen, S. S. Meyer, P. D. Noerdlinger, S. M. Read, L. P. Rosen, R. A. Shafer, E. L. Wright, C. L. Bennett, N. W. Boggess, M. G. Hauser, T. Kelsall, S. H. Moseley, Jr., R. F. Silverberg, G. F. Smoot, R. Weiss, and D. T. Wilkinson, “Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument,” *Astrophys. J.*, vol. 420, pp. 439–444, Jan. 1994.
- [93] J. Huchra, M. Davis, D. Latham, and J. Tonry, “A survey of galaxy redshifts. IV - The data,” *Astrophys. J. Supp.*, vol. 52, pp. 89–119, June 1983.
- [94] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot, and et al., “SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems,” *Astron. J.*, vol. 142, p. 72, Sept. 2011, 1101.1529.
- [95] R. Opher and A. Pelinson, “Constraints on dark energy from the observed density fluctuations spectrum and supernova data,” *ArXiv Astrophysics e-prints*, 2005, arXiv:astro-ph/0505476.
- [96] G. M. Voit, “Tracing cosmic evolution with clusters of galaxies,” *Review of Modern Physics*, vol. 77, pp. 207–258, Apr. 2005, arXiv:astro-ph/0410173.

- [97] A. Mantz, S. W. Allen, H. Ebeling, and D. Rapetti, “New constraints on dark energy from the observed growth of the most X-ray luminous galaxy clusters,” *Mon. Not. R. Astron. Soc.*, vol. 387, no. 3, pp. 1179–1192, 2008.
- [98] S. Perlmutter, S. Gabi, G. Goldhaber, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, R. Pain, C. R. Pennypacker, I. A. Small, R. S. Ellis, R. G. McMahon, B. J. Boyle, P. S. Bunclark, D. Carter, M. J. Irwin, K. Glazebrook, H. J. M. Newberg, A. V. Filippenko, T. Matheson, M. Dopita, and W. J. Couch, “Measurements of the Cosmological Parameters  $\Omega$  and  $\Lambda$  from the First Seven Supernovae at  $z \geq 0.35$ ,” *Astrophys. J.*, vol. 483, pp. 565–581, July 1997, astro-ph/9608192.
- [99] S. Perlmutter, G. Aldering, M. della Valle, S. Deustua, R. S. Ellis, S. Fabbro, A. Fruchter, G. Goldhaber, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, R. A. Knop, C. Lidman, R. G. McMahon, P. Nugent, R. Pain, N. Panagia, C. R. Pennypacker, P. Ruiz-Lapuente, B. Schaefer, and N. Walton, “Discovery of a supernova explosion at half the age of the universe,” *Nature.*, vol. 391, p. 51, Jan. 1998, astro-ph/9712212.
- [100] M. Sullivan, J. Guy, A. Conley, N. Regnault, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, N. Palanque-Delabrouille, K. M. Perrett, C. J. Pritchett, J. Rich, V. Ruhlmann-Kleider, D. Balam, S. Baumont, R. S. Ellis, S. Fabbro, H. K. Fakhouri, N. Fourmanoit, S. González-Gaitán, M. L. Graham, M. J. Hudson, E. Hsiao, T. Kronborg, C. Lidman, A. M. Mourao, J. D. Neill, S. Perlmutter, P. Ripoche, N. Suzuki, and E. S. Walker, “SNLS3: Constraints on Dark Energy Combining the Supernova Legacy Survey Three-year Data with Other Probes,” *Astrophys. J.*, vol. 737, p. 102, Aug. 2011, 1104.1444.
- [101] K. Glazebrook, D. Eisenstein, A. Dey, B. Nichol, and The WFMOS Feasibility Study Dark

- Energy Team, “Dark Energy and Cosmic Sound:  $w(z)$  Surveys with the Gemini/Subaru Wide-Field Multi-Object Spectrograph,” *ArXiv Astrophysics e-prints*, July 2005, astro-ph/0507457.
- [102] C. Blake and K. Glazebrook, “Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler,” *The Astrophysical Journal*, vol. 594, no. 2, p. 665, 2003.
- [103] W. Hu and Z. Haiman, “Redshifting rings of power,” *Physical Review D.*, vol. 68, p. 063004, Sept. 2003, astro-ph/0306053.
- [104] H.-J. Seo and D. J. Eisenstein, “Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys,” *Astrophys. J.*, vol. 598, pp. 720–740, Dec. 2003, astro-ph/0307460.
- [105] S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and K. Taylor, “The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications,” *Mon. Not. R. Astron. Soc.*, vol. 362, pp. 505–534, Sept. 2005, astro-ph/0501174.
- [106] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind, V. Bhardwaj, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, A. S. Bolton, A. Borde, J. Bovy, W. N. Brandt, H. Brewington, J. Brinkmann, P. J. Brown, J. R. Brownstein, K. Bundy, N. G. Busca, W. Carithers, A. R. Carnero, M. A. Carr, Y. Chen, J. Comparat, N. Connolly, F. Cope,

R. A. C. Croft, A. J. Cuesta, L. N. da Costa, J. R. A. Davenport, T. Delubac, R. de Put-  
ter, S. Dhital, A. Ealet, G. L. Ebelke, D. J. Eisenstein, S. Escoffier, X. Fan, N. Filiz Ak,  
H. Finley, A. Font-Ribera, R. Génova-Santos, J. E. Gunn, H. Guo, D. Haggard, P. B. Hall,  
J.-C. Hamilton, B. Harris, D. W. Harris, S. Ho, D. W. Hogg, D. Holder, K. Honscheid,  
J. Huehnerhoff, B. Jordan, W. P. Jordan, G. Kauffmann, E. A. Kazin, D. Kirkby, M. A.  
Klaene, J.-P. Kneib, J.-M. Le Goff, K.-G. Lee, D. C. Long, C. P. Loomis, B. Lundgren,  
R. H. Lupton, M. A. G. Maia, M. Makler, E. Malanushenko, V. Malanushenko, R. Man-  
delbaum, M. Manera, C. Maraston, D. Margala, K. L. Masters, C. K. McBride, P. McDon-  
ald, I. D. McGreer, R. G. McMahon, O. Mena, J. Miralda-Escudé, A. D. Montero-Dorta,  
F. Montesano, D. Muna, A. D. Myers, T. Naugle, R. C. Nichol, P. Noterdaeme, S. E. Nuza,  
M. D. Olmstead, A. Oravetz, D. J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-  
Delabrouille, K. Pan, J. K. Parejko, I. Pâris, W. J. Percival, I. Pérez-Fournon, I. Pérez-  
Ràfols, P. Petitjean, R. Pfaffenberger, J. Pforr, M. M. Pieri, F. Prada, A. M. Price-Whelan,  
M. J. Raddick, R. Rebolo, J. Rich, G. T. Richards, C. M. Rockosi, N. A. Roe, A. J. Ross,  
N. P. Ross, G. Rossi, J. A. Rubiño-Martin, L. Samushia, A. G. Sánchez, C. Sayres, S. J.  
Schmidt, D. P. Schneider, C. G. Scóccola, H.-J. Seo, A. Shelden, E. Sheldon, Y. Shen,  
Y. Shu, A. Slosar, S. A. Smee, S. A. Snedden, F. Stauffer, O. Steele, M. A. Strauss,  
A. Streblyanska, N. Suzuki, M. E. C. Swanson, T. Tal, M. Tanaka, D. Thomas, J. L.  
Tinker, R. Tojeiro, C. A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, D. A. Wake,  
M. Watson, B. A. Weaver, D. H. Weinberg, B. J. Weiner, A. A. West, M. White, W. M.  
Wood-Vasey, C. Yèche, I. Zehavi, G.-B. Zhao, and Z. Zheng, “The Baryon Oscillation  
Spectroscopic Survey of SDSS-III,” *Astron. J.*, vol. 145, p. 10, Jan. 2013, 1208.0022.

- [107] A. R. Liddle and D. H. Lyth, *Cosmological inflation and large-scale structure*. Cambridge University Press., 1999.

- [108] S. Dodelson, *Modern Cosmology*. Academic Press, Elsevier Science, 2003.
- [109] C.-P. Ma and E. Bertschinger, “Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges,” *The Astrophysical Journal*, vol. 455, p. 7, Dec. 1995, astro-ph/9506072.
- [110] A. Guth, *The Inflationary Universe: The Quest for a New Theory of Cosmic Origins*. Helix books, Perseus Books, 1997.
- [111] S. W. Hawking, “The development of irregularities in a single bubble inflationary universe,” *Physics Letters B*, vol. 115, pp. 295–297, Sept. 1982.
- [112] G. Efstathiou and J. R. Bond, “Isocurvature cold dark matter fluctuations,” *Mon. Not. R. Astron. Soc.*, vol. 218, pp. 103–121, Jan. 1986.
- [113] P. J. E. Peebles, “Cosmic background temperature anisotropy in a minimal isocurvature model for galaxy formation,” *Astrophys. J. Lett.*, vol. 315, pp. L73–L76, Apr. 1987.
- [114] M. Bucher, K. Moodley, and N. Turok, “General primordial cosmic perturbation,” *Physical Review D*, vol. 62, p. 083508, Oct. 2000, astro-ph/9904231.
- [115] M. Zaldarriaga and U. Seljak, “CMBFAST for Spatially Closed Universes,” *Astrophys. J. Supp.*, vol. 129, pp. 431–434, Aug. 2000, astro-ph/9911219.
- [116] A. Melchiorri, P. A. R. Ade, P. de Bernardis, J. J. Bock, J. Borrill, A. Boscaleri, B. P. Crill, G. De Troia, P. Farese, P. G. Ferreira, K. Ganga, G. de Gasperis, M. Giacometti, V. V. Hristov, A. H. Jaffe, A. E. Lange, S. Masi, P. D. Mauskopf, L. Miglio, C. B. Netterfield, E. Pascale, F. Piacentini, G. Romeo, J. E. Ruhl, and N. Vittorio, “A Measurement of  $\Omega$  from the North American Test Flight of Boomerang,” *Astrophys. J. Lett.*, vol. 536, pp. L63–L66, June 2000, astro-ph/9911445.

- [117] A. Lasenby and A. Challinor *Proceedings of the XL1st Rencontre de Moriond*, eds. C. Magneville, R. Ansari, J. Dumarchez and J.T.T. Vˆan, *La Thuile*, vol. 40, pp. 15–23, 2006.
- [118] O. Lahav and A. R. Liddle, “The Cosmological Parameters,” *ArXiv Astrophysics e-prints*, June 2004, astro-ph/0406681.
- [119] Particle Data Group, S. Eidelman, K. G. Hayes, K. A. Olive, M. Aguilar-Benitez, C. Am- sler, D. Asner, K. S. Babu, R. M. Barnett, J. Beringer, P. R. Burchat, C. D. Carone, S. Caso, G. Conforto, O. Dahl, G. D’Ambrosio, M. Doser, J. L. Feng, T. Gherghetta, L. Gibbons, M. Goodman, C. Grab, D. E. Groom, A. Gurtu, K. Hagiwara, J. J. Hernandez- Rey, K. Hikasa, K. Honscheid, H. Jawahery, C. Kolda, Y. Kwon, M. L. Mangano, A. V. Manohar, J. March-Russell, A. Masoni, R. Miquel, K. Monig, H. Murayama, K. Naka- mura, S. Navas, L. Pape, C. Patrignani, A. Piepke, G. Raffelt, M. Roos, M. Tanabashi, J. Terning, N. A. Tornqvist, T. G. Trippe, P. Vogel, C. G. Wohl, R. L. Workman, W.-M. Yao, P. A. Zyla, B. Armstrong, P. S. Gee, G. Harper, K. S. Lugovsky, S. B. Lugovsky, V. S. Lugovsky, A. Rom, M. Artuso, E. Barberio, M. Battaglia, H. Bichsel, O. Biebel, P. Bloch, R. N. Cahn, D. Casper, A. Cattai, R. S. Chivukula, G. Cowan, T. Damour, K. Desler, M. A. Dobbs, M. Drees, A. Edwards, D. A. Edwards, V. D. Elvira, J. Erler, V. V. Ezhela, W. Fetscher, B. D. Fields, B. Foster, D. Froidevaux, M. Fukugita, T. K. Gaisser, L. Garren, H.-J. Gerber, G. Gerbier, F. J. Gilman, H. E. Haber, C. Hagmann, J. Hewett, I. Hinchliffe, C. J. Hogan, G. Hohler, P. Igo-Kemenes, J. D. Jackson, K. F. Johnson, D. Karlen, B. Kayser, D. Kirkby, S. R. Klein, K. Kleinknecht, I. G. Knowles, P. Kreitz, Y. V. Kuyanov, O. Lahav, P. Langacker, A. Liddle, L. Littenberg, D. M. Manley, A. D. Martin, M. Narain, P. Nason, Y. Nir, J. A. Peacock, H. R. Quinn, S. Raby, B. N. Rat- cliff, E. A. Razuvaev, B. Renk, G. Rolandi, M. T. Ronan, L. J. Rosenberg, C. T. Sachrajda, Y. Sakai, A. I. Sanda, S. Sarkar, M. Schmitt, O. Schneider, D. Scott, W. G. Seligman, M. H.

- Shaevitz, T. Sjöstrand, G. F. Smoot, S. Spanier, H. Spieler, N. J. C. Spooner, M. Srednicki, A. Stahl, T. Stanev, M. Suzuki, N. P. Tkachenko, G. H. Trilling, G. Valencia, K. van Bibber, M. G. Vincter, D. R. Ward, B. R. Webber, M. Whalley, L. Wolfenstein, J. Womersley, C. L. Woody, O. V. Zenin, and R.-Y. Zhu, “Review of Particle Physics,” *Physics Letters B*, vol. 592, pp. 1–5, July 2004, astro-ph/0406663.
- [120] S. M. Kasanda, “Initial conditions of the universe: Signatures in the cosmic microwave background and baryon acoustic oscillations,” *Ph.D thesis, University of KwaZulu-Natal*, 2012.
- [121] U. Seljak and M. Zaldarriaga, “A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies,” *Astrophys. J.*, vol. 469, p. 437, Oct. 1996, astro-ph/9603033.
- [122] S. Muya Kasanda, C. Zunckel, K. Moodley, B. A. Bassett, and P. Okouma, “The sensitivity of BAO dark energy constraints to general isocurvature perturbations,” *Journal of Cosmology and Astroparticle Physics.*, vol. 7, p. 21, July 2012, 1111.2572.
- [123] H. Kodama and M. Sasaki, “Cosmological Perturbation Theory,” *Progress of Theoretical Physics Supplement*, vol. 78, p. 1, 1984.
- [124] J. M. Bardeen, “Gauge-invariant cosmological perturbations,” *Physical Review D*, vol. 22, pp. 1882–1905, Oct. 1980.
- [125] D. Langlois, “Correlated adiabatic and isocurvature perturbations from double inflation,” *Physical Review D*, vol. 59, p. 123512, June 1999, astro-ph/9906080.
- [126] T. Moroi and T. Takahashi, “Effects of cosmological moduli fields on cosmic microwave background,” *Physics Letters B*, vol. 522, pp. 215–221, Dec. 2001, hep-ph/0110096.

- [127] D. H. Lyth and D. Wands, “Generating the curvature perturbation without an inflaton,” *Physics Letters B*, vol. 524, pp. 5–14, Jan. 2002, hep-ph/0110002.
- [128] J. García-Bellido and D. Wands, “Metric perturbations in two-field inflation,” *Physical Review D*, vol. 53, pp. 5437–5445, May 1996, astro-ph/9511029.
- [129] A. Linde and V. Mukhanov, “Non-Gaussian isocurvature perturbations from inflation,” *Physical Review D*, vol. 56, p. 535, July 1997, astro-ph/9610219.
- [130] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, “Adiabatic and entropy perturbations from inflation,” *Physical Review D*, vol. 63, p. 023506, Jan. 2001, astro-ph/0009131.
- [131] R. T. Cox, “Probability, Frequency and Reasonable Expectation,” *American Journal of Physics*, vol. 14, pp. 1–13, 1946.
- [132] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” vol. 57, no. 1, pp. 97–109, 1970.
- [133] A. Gelman and D. Rubin, “Inference from iterative simulation using multiple sequences,” *Statistical Science*, vol. 7, pp. 457–511, 1992.
- [134] G. O. Roberts and J. S. Rosenthal, “Optimal scaling for various metropolis-hastings algorithms,” *Statistical Science*, vol. 16, pp. 351–367, 11 2001.
- [135] E. Koblents and J. Míguez, “A comparison of nonlinear population Monte Carlo and particle Markov chain Monte Carlo algorithms for Bayesian inference in stochastic kinetic models,” *ArXiv e-prints*, Apr. 2014, 1404.5218.
- [136] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. pp. 79–86, 1951.

- [137] O. Cappé, R. Douc, A. Guillin, J. Marin, and C. P. Robert, “Adaptive Importance Sampling in General Mixture Classes,” *ArXiv e-prints*, Oct. 2007, 0710.4242.
- [138] E. Koblents and J. Míguez, “A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models,” *ArXiv e-prints*, Aug. 2012, 1208.5600.
- [139] D. Parkinson and A. R. Liddle, “Bayesian Model Averaging in Astrophysics: A Review,” *ArXiv e-prints*, Feb. 2013, 1302.1721.
- [140] B. Liseo and A. Parisi, “Bayesian inference for the multivariate skew-normal model: a Population Monte Carlo approach,” *ArXiv e-prints*, Feb. 2013, 1302.0977.
- [141] N. Chopin, “A sequential particle filter method for static models,” *Biometrika*, vol. 89, no. 3, pp. pp. 539–551, 2002.
- [142] L. Verde, H. V. Peiris, D. N. Spergel, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, E. Wollack, and E. L. Wright, “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Parameter Estimation Methodology,” *The Astrophysical Journal Supplement Series*, vol. 148, no. 1, p. 195, 2003.
- [143] M. R. Nolta, J. Dunkley, R. S. Hill, G. Hinshaw, E. Komatsu, D. Larson, L. Page, D. N. Spergel, C. L. Bennett, B. Gold, N. Jarosik, N. Odegard, J. L. Weiland, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright, “Five-Year Wilkinson Microwave Anisotropy Probe Observations: Angular Power Spectra,” *The Astrophysical Journal Supplement Series*, vol. 180, no. 2, p. 296, 2009.
- [144] D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard,

- L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters,” *The Astrophysical Journal Supplement Series*, vol. 192, no. 2, p. 16, 2011.
- [145] A. Venkatesan, “The Optical Depth to Reionization as a Probe of Cosmological and Astrophysical Parameters,” *Astrophysical Journal*, vol. 537, pp. 55–64, July 2000, astro-ph/9912401.
- [146] P. J. Steinhardt, “Cosmological Perturbations: Myths and Facts,” *Modern Physics Letters A*, vol. 19, no. 13, pp. 967–982, 2004.
- [147] L. A. Boyle, P. J. Steinhardt, and N. Turok, “Inflationary Predictions for Scalar and Tensor Fluctuations Reconsidered,” *Physical Review Letters*, vol. 96, p. 111301, Mar. 2006, astro-ph/0507455.
- [148] R. Bean, J. Dunkley, and E. Pierpaoli, “Constraining isocurvature initial conditions with WMAP 3-year data,” *Physical Review D.*, vol. 74, p. 063503, Sept. 2006, astro-ph/0606685.