Computational Bayesian techniques applied to cosmology

Sonke Hee

Churchill College

UNIVERSITY OF CAMBRIDGE

2017 April

A dissertation submitted for the degree of Doctor of Philosophy
This thesis presents work around 3 themes: dark energy, gravitational waves and Bayesian inference. Both dark energy and gravitational wave physics are not yet well constrained. They present interesting challenges for Bayesian inference, which attempts to quantify our knowledge of the universe given our astrophysical data.

A dark energy equation of state reconstruction analysis finds that the data favours the vacuum dark energy equation of state $w = -1$ model. Deviations from vacuum dark energy are shown to favour the super-negative ‘phantom’ dark energy regime of $w < -1$, but at low statistical significance. The constraining power of various datasets is quantified, finding that data constraints peak around redshift $z = 0.2$ due to baryonic acoustic oscillation and supernovae data constraints, whilst cosmic microwave background radiation and Lyman-$\alpha$ forest constraints are less significant. Specific models with a conformal time symmetry in the Friedmann equation and with an additional dark energy component are tested and shown to be competitive to the vacuum dark energy model by Bayesian model selection analysis: that they are not ruled out is believed to be largely due to poor data quality for deciding between existing models.

Recent detections of gravitational waves by the LIGO collaboration enable the first gravitational wave tests of general relativity. An existing test in the literature is used and sped up significantly by a novel method developed in this thesis. The test computes posterior odds ratios, and the new method is shown to compute these accurately and efficiently. Compared to computing evidences, the method presented provides an approximate 100 times reduction in the number of likelihood calculations required to compute evidences at a given accuracy. Further testing may identify a significant advance in Bayesian model selection using nested sampling, as the method is completely general and straightforward to implement. We note that efficiency gains are not guaranteed and may be problem specific: further research is needed.
Contents

Summary iii

Contents v

Declaration vii

Acknowledgements ix

1 Outline 1
  1.1 Chapter synopsis 2

2 Cosmological and gravitational framework 5
  2.1 The concordance model: ΛCDM 6
  2.2 Dark energy 24
  2.3 Gravitational waves 29

3 Statistical Framework 37
  3.1 Bayesian inference 38
  3.2 Nested sampling 46
  3.3 Implementation in cosmology 54

4 Development of the H3L method 59
  4.1 Introduction 60
  4.2 Method 61
  4.3 Application to toy-models 64
  4.4 Applications to the dark energy equation of state 74
  4.5 Conclusions 82

5 Time dependence in dark energy equation of state 87
  5.1 Introduction 88
  5.2 Datasets and Computation 89
  5.3 Results: dark energy equation of state reconstruction 94
  5.4 Results: Kullback-Leibler divergence and dataset analysis 98
  5.5 Conclusions 102

6 Double dark energy model investigation 105
  6.1 Introduction 106
  6.2 Conformal time development of a flat-Λ radiation-filled universe 107
6.3 Inclusion of matter .................................................. 108
6.4 Phenomenology ..................................................... 112
6.5 Analysis ............................................................... 116
6.6 Results ................................................................. 118
6.7 Discussion and Conclusions ....................................... 122

7 H3L method acceleration in a gravitational wave test of GR 123
   7.1 Introduction ......................................................... 124
   7.2 Statistical framework ............................................. 125
   7.3 Results: sine wave toy model .................................. 130
   7.4 Results: Kerr waveform model test of GR .................... 143
   7.5 Conclusions ........................................................ 149

8 Conclusions .............................................................. 153

Bibliography .............................................................. 157
DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit (60,000 words) for the relevant Degree Committee.
ACKNOWLEDGEMENTS

Without my supervisors, Mike Hobson and Anthony Lasenby (as well as Will Handley who was an additional de facto supervisor throughout the years), this thesis would be considerably emptier if not blank. Thank you for the direction and genius that guided the research and overcame all problems.

The work in this dissertation has substantially benefited from a steady stream of discussion with (mostly forced upon) a variety of brilliant researchers within the IoA and MRAO groups. Thanks to Farhan Feroz for initial help with using MultiNest, Jose Vázquez for sharing various codes for reference, Ed Higson for disseminating a masterful understanding of errors in nested sampling, and Alvin Chua for supervisorial guidance on gravitational wave astronomy. With immeasurable thanks to Karen Scrivener for looking after everyone in MRAO so well.

My motivation over three and a half years of research has been greatly aided by a fantastic office environment. I am grateful to Richard Wolstenhulme for a consistent and holistic work efficiency programme, to Will Handley for being the office oracle on all matters physics and Linux, Do Young Kim for being positively lovely all of the time, and Carina Negreanu for always being a great source of constructive criticism and fun discussions. The K34 office will be greatly missed.

There is much to be said for the ‘Cambridge experience’, and I am very grateful for the constant support offered and opportunities for personal growth provided. Churchill College has been central to both welfare and development, and I am especially grateful to Rebecca Sawalmeh and Shelley Surtees for making Churchill College such a welcoming place. I am also grateful for all the time, effort and wisdom that the Møller centre coaches have offered whilst I was fortunate enough to be involved in their training program.

Finally, the warmest gratitude goes to my family for getting me to where I am: immeasurable thanks to Anja Hee, Anissa Hee, Stephen Tucker and Cedrik Hee - you all know what you have done. Especially, thank you mum for the financial and moral support. I dedicate this thesis to Aira, Anne, and a plum-sized 3 month old baby undergoing a phase of accelerated expansion.
Although every generation of PhD student must think this, we are at a remarkable time to be studying cosmology. We are privileged to have a wealth of observations that constrain models of the universe, from large scale structure surveys that analyse the universe’s recent evolution to a clear image of the cosmic microwave background radiation that shows a snapshot of the universe’s distant past. Alongside rapid development in cosmological data precision, the Bayesian statistical framework to analyse this data has developed: as an example, the remarkable contribution nested sampling has made to the field of Bayesian model selection.

In order to fit observations, cosmological models posit the existence of dark energy. Dark energy refers to a hypothetical repulsive force that is required to explain observations of the recent accelerated expansion of the universe. Dark energy is poorly understood despite its dominant contribution to the current universe’s evolution. This thesis constrains dark energy behaviour phenomenologically using data-driven analysis techniques. It also presents a novel method of quantifying the various dataset contributions to such phenomenological constraints. Additionally a specific dark energy model extension is tested and shown to be competitive to the standard dark energy description, though the new model’s competitiveness stems from a lack of data with which to constrain the models. Several of the analysis techniques can and have been applied to other areas of cosmology. The methods provide useful tools for future research as new observational missions provide ever more powerful data.

A new type of observation has become available in 2016 with the first direct detection of gravitational waves by the advanced LIGO detectors. Future missions are being upgraded and developed, including further ground based observatories as well as space based gravitational
wave observatories. These missions will detect gravitational waves more clearly as well as at different frequencies. The first detection was heralded as the dawning of gravitational wave astronomy, and the future is rich with possibility regarding its application. A particularly interesting application is in testing general relativity in the strong field regime. This thesis expands on the use of gravitational waves to test general relativity by using a novel method to improve the accuracy of the Bayesian model selection calculations. The method was first developed for the aforementioned dark energy investigation and is generally applicable to any Bayesian model selection problem.

The thesis presents both cosmological results for understanding dark energy and the future potential of gravitational wave tests of general relativity. Alongside these are a heavy emphasis on the analysis techniques developed. Below is a more detailed overview of each chapter and their relation to the thesis generally.

1.1 Chapter synopsis

Chapter 2 presents an overview of the dark energy and gravitational wave physics used throughout the thesis. Specifically, section 2.1 describes the Einstein field equations of general relativity and how to obtain the standard cosmological model from these. A quick overview of the universe’s evolution history is presented to frame the uniqueness of the dark energy era we find ourselves in. Observations to constrain our universe are described, including the cosmic microwave power spectrum, baryonic acoustic oscillation measurements and supernova measurements. Section 2.2 reviews the vacuum energy interpretation of the cosmological constant which forms the ‘Λ’ part of the ΛCDM model. Alternate dark energy models are reviewed briefly to highlight how constraining the dark energy equation of state parameter can shape our understanding of permissible dark energy models. Section 2.3 reviews the gravitational wave equations, wave detections and tests of general relativity.

Chapter 3 reviews the Bayesian inference framework which underpins our ability to constrain cosmological models with data. The thesis is inherently Bayesian throughout, and time is dedicated to define parameter estimation and model selection. Section 3.2 describes the nested sampling algorithm which is used to compute parameter estimates and the model selection variables. Computational implementations (most notably PolyChord) are discussed, with an emphasis on their robustness and the errors computed on the quantities of interest for inference. Further discussions on applying this framework to cosmological investigations and the gravitational wave problem are presented.

Chapter 4 presents and verifies a new method developed in this thesis for calculating model selection quantities (posterior odds ratios and Bayes factors). Traditionally, these quantities are
difficult to obtain, with nested sampling providing the best method by computing ‘evidences’ for each model from which model selection is trivial. The new method avoids the expensive evidence calculation altogether and instead uses parameter estimation on a model selection hyper-parameter to obtain posterior odds ratios. The method is verified on a toy model and then applied to a phenomenological investigation of potentially time varying dark energy equation of state behaviour. The new method is shown to compute Bayes factors reliably when the traditional evidence calculations are not reliable.

Chapter 5 expands on the phenomenological dark energy investigation by investigating more fully the dependence of dark energy behaviour on the choice of datasets. A novel formalism of the Kullback-Leibler divergence is presented, where the Kullback-Leibler divergence is used to define the information addition when moving from prior to posterior parameter distributions (where the additional information is due to the datasets used). The novel formalism identifies the equation of state constraining power of the datasets as a function of time. The results show that $\Lambda$CDM is consistent with all potential time varying behaviour. Additionally, we observe that baryonic acoustic oscillations and supernovae data provide the strongest constraints in general, whilst Lyman-$\alpha$ baryonic acoustic oscillation measurements provide much needed additional constraining power at earlier times.

Chapter 6 continues to investigate the dark energy equation of state behaviour but from an alternative model perspective (rather than chapters 4 and 5 which present model-independent investigations). A model is presented where an additional matter density component is introduced with equation of state $w = -2/3$. The introduction is justified by a symmetry in the evolution history of the universe (in terms of conformal time). The model is not disfavoured compared to $\Lambda$CDM in a Bayesian model selection analysis, but the new model parameter estimates are consistent with $\Lambda$CDM. An additional model is tested as a natural extension in which the additional matter component’s equation of state is allowed to vary, again showing no preference for or against $\Lambda$CDM.

Chapter 7 utilises the novel method described in chapter 4 to improve the efficiency of a gravitational wave test of general relativity. The test is a model-independent phenomenological test of deviations from general relativity in the phase coefficients of the gravitational wave. Similar tests have been used by the LIGO collaboration to find that the observed gravitational waves show no sign of deviations from general relativity. The new method is able to compute Bayes factors significantly faster than when computing evidences first. For a toy sinusoidal model with a wave signal without deviations, the new method computes equivalently accurate posterior odds ratios with 24 times less computational resource (fewer likelihood calculations). This efficiency gain is also observed using a more physically relevant model: gravitational waves produced by Kerr model binary coalescence. For a mock data injection without deviation from
general relativity, the efficiency gain is on order $\approx 100$, though the analysis could not be completed as thoroughly as for the toy model due to general computational limitations.

Chapter 8 summarises the key results and discusses future applications. The four chapters of original work present primarily advancements in dark energy and gravitational wave data analysis. The new method for computing posterior odds ratios is applicable more widely to any model selection problem. The novel formalism of the Kullback-Leibler divergence is also applicable more generally to any function that is constrained by data, and has already been adopted by other authors. We conclude that generally the data analysis techniques are of interest widely in the astrophysical community as well as outside of it once the adoption of nested sampling prevails beyond the astrophysics community.
In this chapter we present the theoretical framework used throughout this thesis. The mathematical arena for this work is general relativity. We start with the Einstein field equations and move on quickly to cosmological and wave solutions. Specifically, we present the standard model of our universe, how to extend to alternate dark energy models, and the gravitational wave equations. The most detail is dedicated to testable observations within these topics as that is the focus of this thesis: data-driven model analysis to constrain dark energy physics and improving data-driven gravitational wave tests of general relativity.

In section 2.1 we present the textbook $\Lambda$CDM cosmological model of the universe, obtained by solving the Einstein field equations using the cosmological principle, and outline the dynamical behaviour of this model and how it can be tested by astrophysical observations. In section 2.2, extensions to the dark energy component of this model are analysed. Finally, section 2.3 presents the gravitational wave solutions to the Einstein field equations and describes the promising field of gravitational wave astronomy.

This short review uses a wide range of materials, with citations introduced as needed, whilst using as references the textbooks of Dodelson (2003) and Hobson et al. (2006) and also the very clear introductory chapters in the theses of Vázquez (2013), Handley (2016) and Chua (2017).
Chapter 2. Cosmological and gravitational framework

2.1 The concordance model: $\Lambda$CDM

Over the years many cosmological models have been proposed, from Einstein’s attempt at creating a static universe to ones containing only matter and radiation. At present the consensus is that the $\Lambda$CDM model fits all cosmological observations remarkably well. The building blocks of the universe that are familiar to us, namely ordinary (baryonic) matter and photons (radiation), only account for about 5% of this proposed model. It should seem odd to anyone that we believe a model where 95% of its contents are hitherto unknown to us, but cosmologists have been led to this conclusion by a host of interesting astrophysical observations.

In this section we outline what the $\Lambda$CDM model is and to discuss observational data relating to cosmology. We assume a basic knowledge of general relativity but do not rely on it heavily in this introduction as the thesis focusses on data analysis rather than theory; an understanding of spacetime, the related 4-vectors and tensors, and notational index raising and lowering would help but are not necessary to understand the principles.

2.1.1 Einstein field equations and cosmological solutions

John Wheeler famously remarked that “spacetime tells matter how to move; matter tells spacetime how to curve”. This direct relation between matter and spacetime curvature is captured even more succinctly by the Einstein field equations:

$$G_{\mu\nu} = \kappa T_{\mu\nu}$$  \hspace{1cm} (2.1)

where $G_{\mu\nu}$ is the Einstein tensor representing the curvature of spacetime and $T_{\mu\nu}$ is the energy-momentum tensor representing the matter and energy content. $\kappa=8\pi G/c^4$ is a constant which can easily be derived from approximation to the Newtonian gravitational field equations; it is composed of Newton’s gravitational constant $G$, speed of light $c$, as well as $\pi^b$. The subscripts $\mu$ and $\nu$ hide the complexity of the gravitational field equations. These subscripts (or indices) run from 0 to 3 to define ‘tensors’ in differential geometry, where $\mu=0$ are time coordinates and the rest are spatial. After symmetry considerations, equation (2.1) comprises 6 independent non-linear 2nd order partial differential equations, with solutions only possible by making assumptions about the matter and energy distribution or geometry of the spacetime.

To solve equation (2.1) for a model of our universe we will therefore make an assumption about its matter and energy distribution: the universe is homogeneous when viewed on sufficiently large scales. This particular assumption is known as the cosmological principle and, although impossible to confirm for the whole universe based on our limited singular vantage

\textsuperscript{a}Omitting the cosmological constant for now.

\textsuperscript{b}Note that Einstein’s birthday is on international $\pi$-day, another way in which they are connected.
point on Earth, it seems empirically justified: galaxy surveys suggest that our local universe matter distribution is homogeneous at a length scale of $63.3 \pm 0.7\, h^{-1}\text{Mpc}$ (Ntelis et al. 2017) (where $h \approx 0.7$ is suggested by the $\Lambda$CDM model).\(^d\)

We use the cosmological principle to create a maximally symmetric spatial component in the geometry. Choosing comoving coordinates so that an observer with constant spatial coordinates would observe this isotropic universe (amongst other conveniences; such observers are referred to as ‘fundamental observers’), one can obtain the Friedmann-Lemaître-Robertson-Walker (FLRW, or often FRW) metric:

$$ds^2 = a(t)^2 \left[ d\chi^2 + S^2(\chi) \, d\Omega^2 \right],$$

where

$$S^2(\chi) = \begin{cases} 
\sin^2(\chi) & ; k = 1, \\
\chi^2 & ; k = 0, \\
\sinh^2(\chi) & ; k = -1. 
\end{cases}$$

(2.3)

The metric above describes a spacetime with a simple time coordinate and a spatial component $a(t)^2 \left[ d\chi^2 + S^2(\chi) \, d\Omega^2 \right]$. The spatial component is constant in time aside from the aply named scale factor $a(t)$. The geometry of the spatial component is described by the function $S^2(\chi)$ as open, flat or closed, depending on the curvature parameter $k$. Additionally, $d\Omega^2 = d\theta^2 + \sin^2(\phi) d\phi^2$ is the solid angle element for a polar coordinate representation of an isotropic manifold, whilst $\chi$ is the comoving radial coordinate defining the comoving distance between fundamental observers.

The scale factor measures the distance separation (or scale) of the spatial component, specifically, a larger scale factor means that the physical distance is larger between two comoving spatial coordinates. Evolution of this scale factor defines the evolution of the universe, and will be the end product of our analysis.

Solving the Einstein field equations for the FLRW metric requires us to define the matter we wish to include in our model. For simplicity, it is normally assumed that the universe is permeated by a perfect fluid: a fluid which is fully characterised at each point by its density $\rho$ and isotropic pressure $p$. By the cosmological principle this perfect fluid will be homogeneous and stationary with respect to comoving coordinates (to maintain isotropy). A perfect fluid is therefore well characterised by the equation of state parameter $w$, such that $p = w \rho c^2$. Fluids in the $\Lambda$CDM model will have constant equation of state parameter for most of their evolution but $w(t)$ as a function of time could be analysed more generally. What is more, for multiple

\(^d\)Note that the nearest star to the solar system is slightly over 1pc distant, the Milky Way galaxy is just under 35kpc in diameter, the nearest galaxy is around 1Mpc from the Milky Way, the nearest galaxy cluster (Virgo) is over 15Mpc from us, the largest observed structure is about 3Gpc across (Her-CrB great wall), and the observable universe (particle horizon) is ~14Gpc.

\(^d\)A recent gamma-ray burst (GRB) survey analysis by Li & Lin (2015) suggests a homogeneity scale closer to 8000h$^{-1}$Mpc.
different perfect fluids we can combine them and define a total energy momentum tensor as

\[ T^\mu_\nu = \sum_i (T^\mu_\nu)_i = \sum_i \left( \rho_i + p_i/c^2 \right) u^\mu u_\nu - \sum_i (p_i g^\mu_\nu), \]  

(2.4)

where \( u^\mu \) is the 4-velocity of the fluid at a point. This multicomponent perfect fluid can itself be modelled then by a single perfect fluid with density as \( \rho = \sum_i \rho_i \) and \( p = \sum_i p_i. \) We are now able to solve the Einstein field equations using the metric of equation (2.2) and the energy-momentum tensor of equation (2.4).

Additionally we note that a constant term multiplying the metric, \( \Lambda g^\mu_\nu, \) can be added to the left hand side of equation (2.1) without loss of generality. Here \( \Lambda \) is known as the cosmological constant and \( ds^2 = g^\mu_\nu dx^\mu dx_\nu \) shows that \( g^\mu_\nu \) characterises the metric. Solving the Einstein field equations thus, one can obtain

\[ H^2 = \left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho + \frac{1}{3} \Lambda c^2 - c^2 k/a^2, \]

(2.5)

\[ \dot{H} + 3H^2 = \frac{\ddot{a}}{a} = -\frac{8\pi G}{3} \left( \rho + \frac{3p}{c^2} \right) + \frac{1}{3} \Lambda c^2, \]

(2.6)

where equation (2.5) is known as the Friedmann equation and equation (2.6) is the acceleration equation. We have also defined the Hubble parameter \( H(t) \equiv \dot{a}/a. \)

One final point before introducing the \( \Lambda \)CDM model is that generally the Friedmann equation can be written in a more convenient form using the definition of dimensionless density parameters, \( \Omega_i \equiv \frac{8\pi G}{3H^2(t)} \rho_i(t), \) which define the proportion of the total universe’s energy that is contained in a given perfect fluid at any given time. Rearranging equations (2.5) and (2.6), or from the equation for conservation of the energy momentum tensor \( \nabla_\mu T^\mu_\nu = 0 \) implicit in the Einstein field equations, one can obtain the continuity equation: \( \dot{\rho} + (\rho + p) \frac{3\dot{a}}{a} = 0. \) For the perfect fluids discussed we can integrate to obtain \( \rho \propto a^{-3(1+w)} \) such that pressure can be replaced by the equation of state parameter in the Friedmann equation to obtain a more intuitive form:

\[ \left( \frac{H}{H_0} \right)^2 = \sum_i \Omega_i 0 a^{-3(1+w_i)}, \]

(2.7)

where parameters with subscript ‘0’ denote evaluation at the current time. We note firstly that dark energy can be defined as a perfect fluid with an equation of state parameter (to be discussed below), and secondly that we may define a curvature density for notational simplicity (see below): we can therefore write \( \sum_i \Omega_i = 1 \) at all times. Equation (2.7) allows us to define the evolution of the size of the universe (the scale factor) based on current observations of what

\footnote{4-vectors are a consequence of the construction of spacetime as one manifold with 4 dimensions, and are defined as the rate of change of the 4 coordinates with respect to an appropriate parameter that defines the motion, usually proper time (or some other affine choice) along a spacetime curve.}
is in the universe, and for a single perfect fluid with constant equation of state we obtain

\[ a(t) \propto \begin{cases} 
  t^{2/3(1+w)} & : w \neq -1, \\
  e^{Ht} & : w = -1. 
\end{cases} \tag{2.8} \]

The equations can be solved numerically for more complex perfect fluid combinations. We are now as far as we can go in describing the universe using the FLRW metric without specifying the matter and energy content of our perfect fluid explicitly. At this point we stop being general and define a model as a sum of components in the universe. The typical components available to us are summarised in table 2.1 and discussed below. Interesting choices include the Einstein-de-Sitter model with only dust-like matter and no curvature, the Einstein-static model with cosmological constant such that \( \dot{a} = 0 \), and the de Sitter model with no big bang\(^6\).

Observations allow us to constrain what components are required and in what quantities, with the current standard agreed model being that of \( \Lambda \text{CDM} \) with the components of radiation (\( \Omega_r \)), matter (\( \Omega_m \); including baryonic and dark matter), negligible curvature (\( \Omega_k \)) and a cosmological constant density often referred to as dark energy (\( \Omega_{\Lambda} \)).

**Radiation** is characterised by an equation of state parameter \( w = \frac{1}{3} \). Its energy density evolves as \( \rho \propto a^{-4} \), a result of photon number density scaling inversely with volume \( a^3 \) and wavelength being redshifted along the direction of travel with universe expansion (hence energy is reduced). In a universe dominated with radiation energy density, such that \( \Omega \approx \Omega_r \), the scale factor evolves as \( a(t) \propto t^{1/2} \). The early universe was radiation dominated but \( \Omega_r \) is small now due to the \( a^{-4} \) scaling as the universe expands. We note that both photons and relativistic neutrinos contribute to the radiation energy density, but in this work the distinction is not needed.

**Matter** is characterised as a pressureless dust with \( w = 0 \). The energy density of matter therefore evolves with \( \rho \propto a^{-3} \), as the number density for matter particles will scale inversely with volume \( a^3 \). In a universe dominated by matter energy density the scale factor evolves as \( a(t) \propto t^{2/3} \). The energy density does not decline as rapidly as the radiation energy density (which included the additional redshifted factor in its scaling). Therefore, in a universe starting at small \( a(t) \), there is a point in the universe’s history at which the energy density transitions from a radiation dominated era to a matter dominated one. Physical implications of these eras are discussed further in section 2.1.2. In cosmological models the distinction is made between ordinary baryonic matter \( \Omega_b \) and dark matter \( \Omega_{dm} \) as these evolve differently given that dark matter does not interact with radiation other than through gravity. Therefore \( \Omega_m = \Omega_b + \Omega_{dm} \).

**Curvature** can be considered as a perfect fluid with equation of state parameter \( w = -\frac{1}{3} \), such that curvature density scales as \( \rho \propto a^{-2} \), as required for the curvature term in equation (2.5).

---

\(^6\)It also has no matter or radiation.
This mathematical convenience allows us to define \( \Omega_k = -\frac{c^2 k}{H^2 a^2} \), such that the density is negative for positive curvature \( (k > 0) \). Equation (2.5) shows that the scale factor can only decrease in a universe where \( \Omega_k \) is positive and greater than the sum of the remaining densities. In a matter and radiation only universe, a closed universe \( (\Omega_k = 1 - \Omega_r + \Omega_m > 1) \) causes a turning point in the scale factor and eventual return to \( a(t_{end}) = 0 \) in some finite time \( t_{end} \). This fate is typically referred to as the ‘big crunch’, for ominous reason. The physical implications of curvature being positive or negative change slightly with the introduction of \( \Omega_A \), as for some closed universes dark energy can avoid a big crunch.

**Vacuum** energy density is defined by a negative pressure with equation of state parameter \( w = -1 \). Density \( \rho_A \) is then constant, and the scale factor grows exponentially. The interpretation of this fluid as vacuum energy comes from the Einstein field equations, to be discussed further in section 2.2.1. More generally, a perfect fluid with negative equation of state parameter is termed dark energy. The vacuum energy density is constant, making it the only density that does not decrease in time, and it therefore dominates at later times in the universe’s evolution.

Together, the energy densities \( \Omega_r, \Omega_b, \Omega_{dm}, \Omega_k \) and \( \Omega_A \) and the Hubble constant \( H_0 \) define the evolution of the scale factor. There are several measurements and constraints that can be placed on these: \( \Omega_k,0 \approx 0 \) and is set to zero for the \( \Lambda \)CDM model; \( \Omega_r,0 \) is very accurately measured and set to its measured value (Lahav & Liddle 2014); and \( \sum_i \Omega_i = 1 \). After the additional constraints there are 3 free parameters in the above description which can not be predicted \textit{a priori} by theory, but instead these are parameters to be fitted by observation. One significant observation is that the universe is not actually homogeneous\footnote{The universe has been observed to have an average density of 5 atoms per cubic meter, or \( \approx 10^{-26} \text{kg m}^{-3} \). On very small scales we note that people are dynamical overdensities by a factor of \( 10^{29} \), whilst on larger scales the Milky Way dark matter density of around \( 2 \times 10^{-20} \text{kg m}^{-3} \) (Read 2014) is an overdense region.}. To include the inhomogeneity of galactic structure formation in \( \Lambda \)CDM requires 2 additional parameters that define an initial small inhomogeneity distribution at early times, known as the primordial density fluctuations (to which we turn in section 2.1.3). These early universe fluctuations provide the seeds for all the observed structure in the universe. A final parameter is added to describe the ionization state of the universe, which defines how light from distant sources interacts with matter. Adding these 3 parameters to the previous 3 parameters completes the 6 parameter concordance \( \Lambda \)CDM model (Lahav & Liddle 2014).

We note that there are alternatives to the \( \Lambda \)CDM model. Models could be created using different procedures and initial assumptions when solving the Einstein field equations, but these can be problematic: an example would be the \( R_H = ct \) model (Melia & Shevchuk 2012) which initially received attention and then much criticism (Lewis 2013; Kim et al. 2016). As the \( \Lambda \)CDM model provides a very accurate fit, most alternative models look to expand the base 6 parameter \( \Lambda \)CDM model to include other phenomena; such as curvature, modified gravity or a
more complex dark energy component. So with full confidence in $\Lambda$CDM’s explanatory power, let us review some of the physical consequences of the $\Lambda$CDM model and observations one can make to probe the universe.

2.1.2 Evolution history of the Universe

Astrophysical observations inform us that the universe is currently expanding. The Friedmann and acceleration equations show us that, with the measured $\Lambda$CDM components, the scale factor will have been zero at some finite cosmic time in the past (at which point energy densities $\rho_i$ for radiation and matter tend to infinity). Starting in a radiation dominated era with $a \propto t^{1/2}$, it will have moved on to a matter dominated era with evolution $a \propto t^{2/3}$ and is currently in an exponential phase of expansion. Although conventional physics breaks down as we model the $a=0$ universe we have a surprisingly good understanding of the physics from significant fractions of a second through to the present. However, seconds are not always a particularly useful notion at various points of interest, so to describe the various events in the physical evolution of the universe we introduce several alternative parameterisations.

We have already seen the scale factor and its use to describe the geometric properties of the universe. Additionally, we have been using cosmic time which describes the proper time of a stationary fundamental observer, which is a useful reference for the time available for physical interactions within comoving fluids. A useful parameterisation for atomic and sub-atomic interactions is the temperature $T$ of the plasma in the universe. Due to the radiation wavelength scaling one finds that $T \propto a^{-1}$, hence the universe cools as it evolves and started off extremely hot. Temperature relates to particles energies $E$ through $E \propto k_B T$. The final parameterisation we wish to introduce is one which, unlike scale factor and temperature, does not depend on the underlying cosmological model: observed cosmological redshift $z$.

The redshift can be defined by considering a photon emitted by some source at time $t_E$ and arriving today at $t_0$. As the energy and frequency of a single photon are both inversely proportional to the scale factor, we immediately find that $\nu_E/\nu_0 = a(t_0)/a(t_E)$. Redshift is typically defined as the relative change in wavelength, $z = (\lambda_E - \lambda_0)/\lambda_0$, rearranging for
frequency we immediately obtain the usual cosmological redshift equation:

\[ a = \frac{1}{1 + z}. \]  

(2.9)

Using redshift is particularly useful as it is a quantity that we measure when we observe light from distant sources, which is looking back in time due to the photon travel time. Knowing that larger redshift photons were emitted at an early cosmic time means that stating a redshift unequivocally defines an observed event on the universe’s time scale (note that the proposed redshifts of events not directly observed may still be cosmology dependent). Cosmic time, scale factor or temperature can then, if needed, be calculated using knowledge of the model. Let us now describe the cosmological evolution story.

**The Big Bang** generally defines the spontaneous coming into existence of comoving coordinate spacetime, simultaneously at every coordinate, and the hot dense phase of the universe thereafter. No physical process is known to describe this earliest phase, and some hypotheses for the earliest events include a cyclic universe wherein an exponential expansion phase leads to another Big Bang (Penrose 2010) or a ‘cosmic egg’ which has an infinite static phase which broke into the inflationary period (Mithani & Vilenkin 2012), or string theory models such as the ‘ekpyrotic’ colliding brane model (Khoury et al. 2001) and other pre-big bang models (Gasperini & Veneziano 2007). Labelling the big bang as \( a = 0 \), it produces a typical age of the universe in the range \( 13-14 \) Gyr. The earliest time at which we typically begin modelling the universe is in the *Planck era* with temperature scale \( E_P \approx 10^{19} \text{GeV}/k_B \) at times \( t_P \approx 10^{-43} \text{s} \) with temperature \( T_P \approx 10^{32} \text{K} \) (Hobson et al. 2006).

**Inflation** occurs some time around \( 10^{-32} \) seconds after the big bang. This inflation is a period of exponential growth in the scale factor, and we will revisit it in more detail in section 2.1.3 when describing the inhomogeneity of the universe statistically. During inflation the scale factor rapidly grows by a factor of \( e^{N_s} \), where \( N_s \) is the number of *e-folds* and is typically in the range \( 50 - 70 \). Temperature cools equivalently. Loosely speaking quantum mechanical fluctuations are expanded to cosmological scales and provide the initial conditions for the evolution of the universe at large scales. Inflation is now a standard part of cosmological models as it solves several fine tuning issues and provides a mechanism describing the initial density perturbations that are the seeds for large scale structure in our universe to form.

**Reheating** of the universe is believed to occur shortly after inflation. The inflation field essentially deposits energy into the universe and the temperature very rapidly rises again. For thorough reviews see sections IX and X of Bassett et al. (2006) which describes the process with a focus on preheating and its relation to inflation theory, the review by Allahverdi et al. (2010) for a thorough treatment of the subject inclusive of relations to baryogenesis (creation

\[^b\text{Which owes the catchy name to Fred Hoyle, a fierce opponent of the theory (Hoyle et al. 2000).}\]
of matter, including the breaking of matter and anti-matter symmetry), and the review by Amin et al. (2015) which aims to highlight the general evolution (and observable signatures therein) of the universe between inflation and BBN (below). Reheating is an area of ongoing research, but the basic principle for the requirement of reheating is to fill the universe with baryonic matter and energy after inflation, producing suitable initial conditions for the next phase. Typically this process finishes with $E > 10^{15}$GeV.

**Big Bang Nucleosynthesis** (BBN) is the time when light elements are first formed in the universe. After reheating the universe continues to cool as the scale factor increases. Radiation before BBN has been energetic enough to destroy bound matter on short timescales. As the photon energies decline, subatomic particles are able to form protons and neutrons which bind to form elements once the photons have cooled below the nucleus binding energies. Between $E \approx 1\text{MeV}$ and $E \approx 0.1\text{MeV}$ (on order $10^3$ seconds into the age of the universe) the primordial light elements are formed, primarily Helium but also detectable amounts of deuterium and lithium. Cyburt et al. (2016) provides a good up to date review of BBN including Planck satellite data. This sequence of events largely concludes the rapidly evolving early phase of the universe.

**Recombination** of electrons and protons marks the next significant transition: photons are no longer in thermal equilibrium with baryonic matter. Since BBN, the universe continues to cool for some 400,000 years until electrons are able to bind to protons to form hydrogen atoms without photons breaking them apart. Photons no longer interact with the matter and the universe becomes transparent to light for the first time. The interaction-free propagation of light occurred simultaneously at every point in the universe and we find ourselves in a bath of these photons at all times arriving from all directions, having travelled from the surface of last scattering (the instance of free-streaming) to arrive in our detectors. This ‘first light’ can be detected and provides astronomers with a snapshot of the early universe. It is the largest photon redshift that can be measured, with $z \sim 1100$, and can be found in the microwave spectrum of light. This first light is called the cosmic microwave background (CMB) and due to its importance we will revisit it in section 2.1.3. We note too that some time before recombination the universe entered the matter dominated era.

**Reionisation** of the hydrogen gas (and at late times also of helium) occurs when the first stars form and release ionising radiation into the gas (though there could be other contributors to reionisation). This matter phase change means that photons once again interact with (ionised) matter and this scattering off matter has implications for our observations of astrophysical light sources. The stars that emit the ionising radiation form due to the gravitational collapse of the

---

1Electroweak symmetry breaking occurs between reheating and BBN at $E \approx 300\text{GeV}$, see Ghosh (2016) for a review.

3A historical misnomer. Nothing is ‘re’ combined but instead combines for the first time.
post-recombination density fluctuations. This star formation occurs after the ‘dark ages’ of the universe: the time period before stars formed (at redshifts around 20). Reionisation is largely over by redshift $z \approx 6$ (Fan et al. 2006). Assuming that reionisation occurs sharply, Planck Collaboration et al. (2016c) suggest it occurs at redshifts of 8–10. Loeb & Barkana (2001) provides a thorough review, and for a review with more modern data see Zaroubi (2013) as well as the discussion of parameters in the Planck 2015 data release (Planck Collaboration et al. 2016c).

Large scale structure formation (LSS) defines the remaining $\approx 10\text{Gyr}$ to the present, in which galaxies, galaxy clusters and superclusters form to create a ‘cosmic web’ of structure (Bond et al. 1996). The process follows that of gravitational collapse and can be modelled numerically, famously so by the millennium simulation (Springel et al. 2005). The initial small density fluctuations which caused matter to gravitationally collapse into stars also seeds LSS. With sufficient time the over densities become large and perturbation theory becomes inaccurate; non-linear dynamics are required to model LSS. A good review can be found in Springel et al. (2006). The observations section below outlines some of the interesting physics that one can obtain from galaxy surveys analysing LSS.

Dark energy began to dominate the evolution of the scale factor in the recent past (last few billion years) and the LSS period of the universe is believed to have an end in the future. Structure will disappear as the matter density declines and gravitationally bound systems slowly lose matter without being able to gain more (as the exponential expansion isolates systems). The exact future of the universe depends on the nature of the dark energy component: potential scenarios include the above mentioned slow heat death, a sudden change in all fields in the standard model of particle physics due to the decay of the vacuum (Burda et al. 2015) or an exponential expansion leading to a singularity in the scale factor at finite cosmic time, termed the big rip (Caldwell et al. 2003). Characterising dark energy is an active research field and much of the thesis is dedicated to the dark energy equation of state.

After this brief overview of the rich history of the universe a few themes are worth expanding on. Firstly, the inhomogeneity in the universe provides a wealth of information to constrain our cosmological models. Its usefulness is due to the deep connections between the quantum mechanical fluctuations during inflation and the inhomogeneities at all other times. As alluded to above, the CMB pattern shows the primordial density fluctuations in the very early universe whilst LSS surveys can map density fluctuations throughout the redshift range of operational scope. Section 2.1.3 will address the CMB and power spectra specifically due to their central importance for constraining cosmological models.

Another theme is that of observation. The CMB, reionisation redshift and structure formation, amongst others, provide direct probes of cosmology. There are many high quality
experiments already completed and planned for the future which define what is frequently referred to as the current era of ‘precision cosmology’. Section 2.1.4 aims to briefly define the datasets used in this thesis to probe cosmology, and also to frame the work in chapter 5 which presents a novel discussion of the relative powers of datasets to constrain our models.

The final theme is that of a dominant energy density that defines the evolution of the universe. Although not mentioned explicitly, the earliest universe was dominated by an energy density associated with inflation, which caused exponential scale factor growth with cosmic time. Thereafter the universe is radiation dominated and shortly before recombination the universe transitions to a matter dominated era. These two eras define fundamentally different energy scales and physical processes. This puts into perspective the significance of the recent transition into a dark energy dominated era: characterising dark energy will define the evolution of the universe for the future, and potentially for the remainder of all time as \( \Lambda \text{CDM} \) predicts no other energy density which can dominate at late times. This characterisation will be explored in section 2.2 of the introduction and also in chapters 5 and 6 of the thesis.

### 2.1.3 CMB and Power Spectra

In order to characterise inhomogeneities in the universe it is typical to use a power spectrum. In this section we will broadly discuss three power spectra. The first is the primordial power spectrum from inflation, which seeds all inhomogeneities in the universe and has its origins in quantum mechanics. The second is the CMB power spectrum from recombination, which is caused by the primordial fluctuations which presents the initial conditions for LSS. The third power spectrum is the matter power spectrum from LSS at various redshifts, which characterises the universe today. These are fundamental to modern cosmological observation, but before summarising each in more detail let us quickly define what is meant by a power spectrum.

The power spectrum describes the amplitude of fluctuations at different length scales in a signal. The signals we wish to study are the density contrasts \( \delta(x) \) of the density fields of fluids in the universe, where \( x \) is the 3 dimensional position vector in space. We note that these density contrasts are a random field with zero average fluctuation, which is to say that there are overdense and underdense regions in spacetime that sum to an average. We can define the two-point correlation function \( \xi(r) \) as the product of density contrasts for a given separation \( r \), averaged over every possible realisation of that separation (denoted by the symbols \( \langle \rangle \)):

\[
\xi(r) \equiv \left\langle \delta(x) \delta(x + r) \right\rangle. \tag{2.10}
\]

The correlation function describes the degree of clustering. For a homogeneous and isotropic density field it only depends on the separation distance and not direction, such that we can write \( \xi(r) = \xi(r) \). If the density field is Gaussian, then this correlation function contains all the
information of the field. Non-Gaussian fields would have higher order correlation functions (such as the three-point correlation function (Wolstenhulme et al. 2015)) which could contain a wealth of information relevant to the cosmological power spectra we discuss (Baumann 2009), but we will not consider non-Gaussianity in this work.

Taking the Fourier transform of equation (2.10) gives,

\[ \langle \hat{\delta}(k) \hat{\delta}(k') \rangle = (2\pi)^3 \delta^D(k + k') P(k), \]

(2.11)

where \( \hat{\delta}(k) \) is the Fourier transform of \( \delta(x) \), \( \delta^D \) is the 3 dimensional Dirac-delta function, and the power spectrum \( P(k) \) is the Fourier transform of the correlation function:

\[ P(k) = \int d^3 r \xi(r) e^{-i k \cdot r}. \]

(2.12)

Again the power spectrum only depends on \( k \) due to isotropy, letting us write \( P(k) = P(k) \). One further step is that, by convention, we use the dimensionless power spectrum \( \mathcal{P}(k) = \frac{k^3}{2\pi^2} P(k) \). We note that the above description is for matter power spectra, and the spherical power spectrum of the CMB is described below. This power spectrum is the preferred way in cosmology to represent density fluctuations for a number of reasons: physical interpretations are more intuitive in the Fourier domain, power spectrum plots are model independent and the power spectrum effectively compresses the information of Gaussian sky maps; see Tegmark (1997) for a good discussion of the merits in using CMB power spectra.

2.1.3.1 Primordial power spectrum

The exact shape of the primordial power spectrum is the subject of ongoing research, and data-driven free-form reconstructions have provided tight constraints over a significant range of wavelengths (Planck Collaboration et al. 2016f, section XX). The primordial power spectrum, \( \mathcal{P}_R(k) \), characterises the density fluctuations produced by inflation. It is typically parameterised as an almost scale invariant spectrum:

\[ \mathcal{P}_R(k) = A_s \left( \frac{k}{k_s} \right)^{n_s-1}, \]

(2.13)

where \( A_s \) is the amplitude, \( k_s \) is an arbitrary pivot scale, and \( n_s \) is referred to as the spectral index. Different inflationary models will produce different sets of parameters. Additionally, inflation proposes that there are different types of perturbation possible: scalar, vector and tensor. The above characterises the scalar, or curvature, perturbations of the matter. Vector perturbations are assumed negligible. Tensor perturbations are associated with gravitational waves and have to date not been observed.

Extensions on the parameterisation of equation (2.13) include adding a more complex power law behaviour such as a ‘running’ spectral index \( n_{run} \equiv \frac{dn_s}{d\ln k} \) or including tensor contributions.
via an equivalent power spectrum parameterisation which adds to the scalar perturbations. Baumann (2009) presents a very thorough introduction of the primordial power spectrum and inflation. Throughout this work the primordial scalar perturbation power spectrum is assumed. It is worth noting that the two parameters $A_s$ and $n_s$ are part of the 6 parameter $\Lambda$CDM model. $k_s$ is typically set constant at 0.05 Mpc$^{-1}$.

2.1.3.2 CMB power spectrum

The primordial power spectrum perturbations evolve after inflation as the universe goes through BBN and recombination. At the end of recombination the photons stream freely through space to produce the cosmic microwave background (CMB) as discussed in section 2.1.2. We then are able to construct a two dimensional all sky map $k$ of these photons. From this all sky map we wish to extract information about the primordial power spectrum, so we need to understand each of those three steps and how they transform the power spectrum. Firstly one can analyse the power spectrum on a sphere by expanding the temperature fluctuations of the observed CMB in spherical harmonics $Y_{lm}$:

$$\Theta(\hat{n}) \equiv \frac{\Delta T(\hat{n})}{T_0} = \sum_{l} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\hat{n}),$$

$$a_{lm} = \int d\Omega \ Y_{lm}^* (\hat{n}) \Theta(\hat{n}),$$

where $a_{lm}$ are the multipole moments defining the expansion. We can compute the integral in equation (2.15) by taking measurements of $\Theta(\hat{n})$ on the 2-sphere. For a given $l$ there are $2l + 1$ observations of $a_{lm}$ which define our expected variance on the multipole moments. Specifically, at low $l$, which corresponds to large angular scales, we have fewer measurements of $a_{lm}$ and the data becomes less statistically significant: this is known as cosmic variance. From the measurements of $a_{lm}$ we can compute the angular power spectrum as

$$C_l = \frac{1}{2\pi} \sum_{m} \langle a_{lm} a_{lm}^* \rangle .$$

If we rewrite this as $\langle a_{lm} a_{lm}^* \rangle = C_l \delta_{ll'} \delta_{m m'}$ the analogy to equation (2.11) becomes clearer. For more details see Baumann (2009) on which much of this section is based.

Relating the primordial power spectrum to the observed CMB is too complicated to discuss in detail here, a full treatment is given in Dodelson (2003). Here we quickly summarise some results from Baumann (2009) and Challinor & Peiris (2009) which present thorough discussions of the CMB anisotropies. Generally, the primordial power spectrum $P_R(k)$ is related to the

<sup>5</sup>Aside from the regions dominated by the Milky Way and other foreground noise.
CMB power spectrum \( C_l \) by

\[
C_l^{XY} = \frac{2}{\pi} \int k^2 dk \frac{d}{dk} \left( P_R(k) \right) \frac{\Delta X_l(k) \Delta Y_l(k)}{\text{inflation anisotropy}}
\]

(2.17)

where the labels \( X \) and \( Y \) refer to the temperature (T) and polarisation (E and B) modes that contribute to the total power spectrum. Figure 2.1 shows that the main contribution to the CMB power spectrum is from the \( C_l^{TT} \) temperature modes from scalar perturbations, so we will focus on these. The transfer functions \( \Delta X_l(k) \) can be written as a line-of-sight integral which contains contributions from the physical sources and geometric projection.

The effects have been well studied mathematically and codes such as CAMB are able to numerically compute theoretical CMB \( C_l \)s for \( \Lambda \)CDM parameter inputs, as well as a wide variety of other model extensions beyond \( \Lambda \)CDM (Lewis et al. 2000; Howlett et al. 2012). The general physical picture is described by acoustic physics. The initial perturbations from inflation cause peaks and troughs in the fluid density field. The matter collapses under gravity but is modulated by the dominant radiation pressure, such that matter oscillates on scales defined by the primordial fluctuations as well as the matter density. This process creates acoustic waves in the fluid with sound speed \( c_s \). Between the end of inflation and the end of recombination these imprint a fundamental scale at which matter and photons oscillated and therefore a scale at which they are correlated. Specifically, the longest wavelength oscillation is the first peak in \( l \), with subsequent peaks defined by its harmonics. Generally, the shape of the CMB power spectrum is very sensitive to densities and the primordial power spectrum and recent measurements of the CMB have provided the tightest cosmological constraints available.

### 2.1.3.3 Matter power spectrum

Measuring the density fluctuations in matter at any point in the universe produces a matter power spectrum. As LSS was seeded by the primordial anisotropies, the matter power spectrum provides another probe of the early universe and its evolution thereafter. As with the CMB, one can write the power spectrum for matter using transfer functions to be computed numerically:

\[
P_m(k) = \frac{4}{aH} \left( \frac{k}{aH} \right)^4 T_m^2(k) P_R(k),
\]

(2.19)

where the transfer function \( T_m(k) \) now models the growth of matter fluctuations through the evolution of the universe. It has been shown that at large scales (small \( k \)) \( P \propto k \) and that at small scales (large \( k \)) \( P \propto k^{-3} \). The precise form in between these limits is obtained by solving the full general relativistic Boltzmann equation (Dodelson 2003).
2.1. The concordance model: $\Lambda CD$M

2.1.4 Observing and measuring the universe

Feynman is quoted as saying that “if it doesn’t agree with experiment, it’s wrong.” Fortunately there are numerous high quality datasets available to cosmologists and these have shown very good agreement with the $\Lambda$CDM model. To provide a thorough review of all that is available is not within the scope of this short introduction. The aim is to define the datasets that will be used throughout the thesis.

The thesis focuses on datasets which provide likelihood codes as part of the operational output of a mission. Likelihoods are required for mathematical model analysis, to be defined more fully in section 3.1. These datasets can easily be split into several categories: measurements of the cosmic microwave background radiation (CMB), redshift distance measurements of Supernovae type Ia (SNIa), measurements of the baryon acoustic oscillation (BAO) scale in LSS surveys and BAO measurements in the Lyman-\(\alpha\) forest (Ly-\(\alpha\)). There are additional measurements that place constraints on the $\Lambda$CDM model, these include measurements of the BBN elemental abundances to define the helium fraction and measurements of the local Hubble constant.

Before describing the observations and relevant datasets it is worth noting that measuring distances to objects in astrophysics is a non trivial pursuit which depends on the cosmological model. Both BAO and SNIa data rely on redshift and distance relations to determine universe
Chapter 2. Cosmological and gravitational framework

evolution. Pointing a telescope at interesting features on the sky produces light spectra which can
be analysed to classify the object studied. Once the absorption line patterns in the spectrum
are identified a redshift can be stated which defines the wavelength shift in the signal compared
to the known spectral frequency features. Distance is less straightforward to calculate as it
requires knowledge about the metric, and therefore a cosmological model is assumed.

As the universe expands with scale factor $a(t)$, the distance between two comoving points in
spacetime increases. We can define the comoving distance $\chi$ from the metric of equation (2.2)
as the distance between comoving coordinate points in the space-only part of the metric. From
the definition of $\chi$, we can define proper distance as $d=a(t)\chi$ which measures physical distance
between the comoving coordinates at some cosmic time $t$. However, this proper distance is
difficult to measure directly as we are confined to the Earth\(^1\). Instead we can define operational
distance measurements which relate an observable quantity to distance via a known relation.

With sources of light we know that energy dissipates as $r^{-2}$ and on earth we can measure
the flux from a distant source. If we know the energy an object is outputting in a given time
frame, its luminosity $L$, we can calculate the distance via the relation $F=L/(4\pi d_L^2)$, where $d_L$
is the luminosity distance. In an expanding universe, however, the area of the sphere itself
depends on the scale factor. From the FLRW metric we can see that $A=4\pi S^2(\chi)$, where $S^2(\chi)$
is defined in equation (2.3), and we can obtain an equation relating flux and luminosity in an
FLRW universe. By comparing the FLRW equation for flux to the operational observations of
$d_L$ we can relate the luminosity distance to the metric:

$$d_L(z) = S(\chi)(1 + z).$$

(2.20)

This relationship describes the luminosity distance we would measure for a given object at $z$
depending on its comoving coordinate $\chi$. We cannot know the comoving coordinate of a source
we measure, but instead can calculate it at a given redshift, with the distance depending on the
evolution history of the universe:

$$\chi = c \int_0^z dz \: H^{-1}(z).$$

(2.21)

Recalling equation (2.5) and (2.7) we see that the luminosity distance therefore depends heavily
on the energy density content of the universe. This is useful, as measurements of luminosity
distances can therefore constrain the evolution history of the universe. A similar analysis can
be done for the operationally defined angular distance $d_\Lambda$: an object of proper diameter $l$
will subtend an angle on the sky of $\Delta \theta = l/d_\Lambda$, where $d_\Lambda$ is the distance to the source. Therefore,

\(^1\)Distance measurements on Earth were originally defined using a metre stick and currently by the speed of
light. Sadly, in space we cannot lay out rulers to objects and have not enough information about the photon travel
time to calculate distance.
for an object of known dimension we can measure the angle on the sky between the object’s boundaries and compute the angular distance. Again we can compare to the FLRW metric, where the angular part defines \( l = a(t)S(\chi)A\theta \) and we obtain

\[
d_A = \frac{S(\chi)}{1 + z}
\]  

(2.22)

Again we see that this relationship depends on the expansion history through \( \chi \) and therefore measurements of \( d_A \) constrain models of our universe. Note that the luminosity and angular distance are not the same, such that for an object where we know both luminosity and size \( a \) \( a \) priori we could compute both distances and be surprised to see that they are not equivalent. In a non expanding universe, however, they would be the same (as redshift would be 0), in line with what is left of our now questionable intuition of these matters. We note too that the comoving distance, and therefore proper distance, would be the same from both measurements and therefore a more intuitive picture of these measurements might be to consider them as reconstructing \( \chi(z) \).

One can compute a similar distance measure for number density counts of objects, where our knowledge that density is constant in space due to the cosmological principle allows us to relate measured volumes with FLRW metric volume. The various distance measurements discussed allow astronomers to map the universe. Typically different observations are able to measure different scales of distance, and each such measurement will have inherent uncertainties. Combining multiple observations allows us to cross-calibrate measurements and create accurate distance measurements across a range of redshifts and using a range of methods, this is known as the cosmological distance ladder. Let us now turn to the datasets which will be used throughout the thesis.

**Cosmic microwave background** (CMB) radiation measurements\(^{10}\) determine the CMB power spectrum. Experiments for constructing whole sky mappings were conducted by the COBE (Mather et al. 1990)\(^n\), WMAP (Bennett et al. 2003; Hinshaw et al. 2003; Spergel et al. 2003) and most recently Planck (Planck Collaboration et al. 2014b,c) satellites. Ground based telescopes focussing on smaller sky regions with a higher resolution include ACT (Fowler et al. 2010; Das et al. 2011) and SPT (Schaffer et al. 2011; Reichardt et al. 2012), whilst some early work was done on balloons. CMB experiments can measure both temperature and polarization of microwave photons across a range of microwave frequencies. The work in the thesis was started in 2013 when WMAP data release 9 (Hinshaw et al. 2013; Bennett et al. 2013) was the state of the art. WMAP was combined with the ACT and SPT datasets which measure smaller scale CMB features (higher \( l \)). The Planck satellite 2013 data (Planck Collaboration...

\(^1\)The initial discovery of the CMB was by Penzias and Wilson, whilst attempting to remove background noise in their detector, and received a Nobel prize.

\(^2\)COBE black body and anisotropy measurements won Smoot and Mather a Nobel prize.
et al. 2014b) was released shortly after and still used the WMAP9 polarization data. Planck satellite 2015 data was released nearer the end of the thesis work. Elements of this work have been conducted using each of these 5 datasets and will be stated as appropriate.

We note that the Planck data contains several different datasets (coded into likelihoods) which can be used together or in place of one another (Planck Collaboration et al. 2014a, 2016a) (similarly for WMAP). This is due to how the T and E/B mode measurements construct the CMB power spectrum, as discussed briefly in section 2.1.3, and also due to the instrumentation used to measure \( C_l \) at low-\( l \) and hi-\( l \). Additionally there are datasets using gravitational lensing (Planck Collaboration et al. 2016e): photon path distortion by large masses such as dark matter overdensities create distinct signatures that can be used to determine the power spectrum due to lensing. Constraints on the primordial power amplitude \( A_s \), and to a lesser extent the matter density \( \Omega_m \), can be obtained using only the lensing data.

The WMAP data constrained the CMB angular spectrum between \( 2 \leq l \leq 1200 \) by measuring temperature and polarisation in 5 frequency bands between 23–93 GHz (effective frequency) between 2001 and 2010. The precision measurements by WMAP reduced the volume of the \( \Lambda \)CDM 6 parameter space by a factor of 68,000 compared to pre-WMAP constraints\(^a\). The WMAP data can be combined with the ACT and SPT high multipole ground based experiments. The high-\( l \) data extends to \( l \approx 10000 \) but foregrounds are said to dominate above \( l \approx 3000 \): the useful data range is limited to below \( l = 3000 \) (Dunkley et al. 2011).

Planck data extended its satellite-only CMB analysis from \( 2 \leq l \leq 2500 \) by measuring, with higher resolution instruments than WMAP, the temperature and polarisation in 9 frequency bands covering the 25–1000 GHz spectrum between 2009 and 2013. The Planck 2015 data release is sufficiently precise to not significantly benefit from including other high-\( l \) CMB experiments. 2013 Planck data slightly benefits from high-\( l \) data but we have opted to only use Planck in our analysis. It is important to note that the Planck 2013 data analysis introduces 14 parameters which model instrumental noise, foreground signals and other non-CMB sources of power (this is 15 parameters for the 2015 data, whilst the WMAP+ACT+SPT combination uses only 3). These parameters are treated the same as the \( \Lambda \)CDM parameters when constraining parameter values and are termed *nuisance parameters*.

**Supernovae Type Ia** (SNIa) distance redshift measurements provide measurements of \( d_L \). A type Ia supernovae is generally believed to be a (carbon-oxygen) white dwarf star which accreted matter until it reached a critical mass (the Chandrasekhar mass) and a thermonuclear explosion occurred, at which point it became a supernovae. The exact system producing these is still open to debate (Maoz & Mannucci 2012; Wang & Han 2012) but the property that these systems become supernovae with approximately the same known mass makes them useful for

\(^a\)“[T]hereby converting cosmology from a field of wild speculation to a precision science” (NASA 2013)
luminosity distance measurements. As the luminosity of these sources is known, they are often referred to as “standard candles”. Their initial usage in cosmology to probe the luminosity distance is credited with identifying the accelerated expansion of the universe (Riess et al. 1998; Perlmutter et al. 1999)\(^p\) which supports the dark energy construction of our models. The measurements typically constrain late time universe evolution and are degenerate in matter and dark energy densities. Combining SNIa data with other probes breaks this degeneracy and provides the tightest constraints on late time evolution. For a short review on constraining models from observations see Astier (2012).

In this work we use two supernovae catalogues: the Union 2.1 catalogue by Suzuki et al. (2012) and the joint light-curve (JLA) catalogue by Betoule et al. (2014). The Union 2.1 catalogue consists of 580 supernovae combined from several different supernovae surveys, with a redshift range from \(z=0\) to \(z=1.5\) (Suzuki et al. 2012, figure 4). The JLA catalogue was available only later on in the thesis. It consists of 740 supernovae combined from multiple surveys with a range of \(0.01 < z < 1.2\) (Betoule et al. 2014, figure 8). Some of the supernovae overlap between the two catalogues, such that they cannot be used together, and the JLA catalogue typically provides tighter constraints (Betoule et al. 2014, figure 14).

Baryon acoustic oscillation (BAO) data measures the acoustic oscillations in the power spectrum of matter at a given redshift. The acoustic oscillations are those created by the sound waves at the time of recombination. A very good description can be found in Eisenstein et al. (2007) (especially figure 1 describing the power spectrum of various components evolving before and after recombination) and also in the thorough review of Bassett & Hlozek (2009) with informative figures showing various features of BAO on 2D grids. The acoustic oscillations define a preferred scale for matter clustering which can be measured at various redshifts to define an evolution history of the scales of the universe. If the scale is known, then the angular distance can be calculated. BAO are therefore often referred to as “standard rulers”\(^q\). BAO measurements have degeneracies between several parameters (such as between \(\Omega_m\) and \(H_0\), between the dark energy equation of state and \(H_0\), and within early dark energy models; see Aubourg et al. (2015) for a thorough discussion), and again the best results are obtained by combining datasets.

In this work we use several different BAO measurements. These can be classified into galaxy BAO measurements, where the galaxy matter power spectrum is used, and Lyman-\(\alpha\) BAO measurements which measure the intergalactic gas power spectrum using Ly-\(\alpha\) emission lines (along the line-sight from high redshift quasar spectra). Aubourg et al. (2015, Table II) summarise the BAO datasets that we have used, with their discussions directly applicable to

\(^p\)Another Nobel prize winning experiment in cosmology, for Perlmutter, Riess and Schmidt.

\(^q\)The scale is not strictly known \textit{a priori}, and \textit{statistical standard ruler} is a commonly used term to reflect this.
our work. To summarise, we use the galaxy BAO data from the SDSS III BOSS data release 11 (Anderson et al. 2014, DR11) which provides an up to 1% measurement of cosmic distance at redshifts 0.32 and 0.57, with nearly one million galaxies in the redshift range $0.2 < z < 0.7$. We supplement this with the less powerful BAO results of Beutler et al. (2011, 6dFGS), a 6% distance measurement at $z=0.106$, and Ross et al. (2015, MGS), a 4% distance measurement at $z=0.15$.

For the Lyman-$\alpha$ data ($\text{Ly}\alpha$) we use the Lyman-$\alpha$ forest spectra of the BOSS DR11 dataset, specifically we use two independent reductions of this data that can be combined. Note that the specific distance measurements of $\text{Ly}\alpha$ are not equivalent to the galaxy BAO distance measured, so the accuracy is not directly comparable, but both distance measures provide similar constraints on cosmological evolution. Firstly we use the auto-correlation BAO measurements of Delubac et al. (2015) (so called forest-forest correlation). These use 140,000 quasars in redshift range $2.1 < z < 3.5$ and produce constraints on $H_0$ with 2.6% accuracy (5% on the distance measured) at a redshift of $z=2.34$. Secondly we use the quasar-forest cross-correlation BAO measurements of Font-Ribera et al. (2014), which uses 160,000 quasars over a similar redshift range: a 3% precision on $H_0$, 4% on the distance measure and $z=2.36$. These complement the galaxy BAO measurements as they greatly extend the redshift range at which the BAO scale has been measured.

The above three types of dataset define the principal cosmological probes used in this thesis. Each individually has degeneracies within the parameter space, and combining the data provides the best constraints. A thorough review on these can be found in Weinberg et al. (2013). It summarises additionally the developments of each field, other means of constraining cosmological models and the use of gravitational waves as “standard sirens”. In section 2.3 we will return to gravitational wave data and frame the chapter 7 discussions on using gravitational waves to test GR, wherein we demonstrate a novel statistical technique for improving the efficiency of such tests. In the following section we review the topic of dark energy to frame the contributions this thesis makes in providing free-form constraints of dark energy and also in investigating a particular parameterisation.

### 2.2 Dark energy

The accelerated expansion of the universe can be well described by a hypothetical repulsive force that permeates the universe. This is typically referred to as dark energy. Dark energy dominates at later times in the universe’s evolution as the other gravitationally attractive energy sources become diluted due to volume expansion. Due to this late time domination it describes the future fate of the universe and, besides being a fundamentally intriguing mystery, understanding
dark energy is therefore important in completing a model of the universe.

We do not know what dark energy is and understanding the time evolution of the dark energy equation of state phenomenologically is a principal investigation in this thesis. Many models exist to explain observations, though current constraints on the universe’s late time evolution history are not typically strong enough to rule out many models (including a model proposed in this thesis in which a second dark energy component exists alongside the vacuum energy). Here we wish to briefly summarise the motivation for the vacuum energy description of dark energy that makes the $\Lambda$CDM model so successful as well as describing briefly the types of alternative models and their predicted equation of state behaviours.

2.2.1 $\Lambda$CDM vacuum energy and $wCDM$

Einstein’s insight that the curvature of spacetime is related to the matter within it leads to the tensor formalism in equation (2.1). It relates the Einstein tensor, $G_{\mu\nu}$ which represents curvature, to the energy-momentum tensor of matter, $T_{\mu\nu}$. In order to understand the dark energy formalism of the concordance $\Lambda$CDM model, let us quickly review the Einstein field equations in more detail.

To define the Einstein tensor we need to define the curvature of the metric $g_{\mu\nu}$. General relativity provides a possible solution in the form of the Riemann tensor $R_{\mu\nu\rho\sigma}$, also known as the curvature tensor. It is defined in terms of the metric and the first and second order derivatives of the metric. This tensor has the property that it is zero when the metric is flat. We can use this to define the Einstein tensor. As the gravitational potential in the Newtonian limit is given by $\nabla^2 \phi = 4\pi G \rho$, $G_{\mu\nu}$ should have terms no larger than linear in the second order derivatives of the metric tensor. Combining the curvature tensor formalism and the Newtonian conclusions allows us to write a generic expression for $G_{\mu\nu}$:

$$G_{\mu\nu} = aR_{\mu\nu} + bRg_{\mu\nu} + \Lambda g_{\mu\nu},$$

(2.23)

where the tensor $R_{\mu\nu}$ and the scalar $R$ are derived from the curvature tensor via a contraction of indices, and $a$ and $b$ are constants to be determined. A thorough introduction on this topic can be found in Hobson et al. (2006), but here it suffices to note that the quantities only depend on the metric and its derivatives. From the conservation equation of the energy-momentum tensor, $\nabla_\mu T^{\mu\nu} = 0$, we can determine the constants $a$ and $b$ by substituting the general form of $G^{\mu\nu}$ in for $\kappa T^{\mu\nu}$ and comparing to the Newtonian equation in the weak-field limit. Several steps and careful considerations lead to the more specific definition of the Einstein field equations of GR:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu},$$

(2.24)
\( \Lambda \) is called the cosmological constant because any arbitrary constant value multiplied by the metric \( g_{\mu\nu} \) will still satisfy the general form of the Einstein tensor in equation (2.23). This is the origin of \( \Lambda \). Now writing equation (2.24) in mixed components, such that \( R_{\nu}^\mu - \frac{1}{2} \delta_{\nu}^\mu R + \Lambda \delta_{\nu}^\mu = \kappa T_{\nu}^\mu \), and contracting the \( \mu \) and \( \nu \) indices we find that \( R = \kappa T + 4\Lambda \), allowing us to write an alternate form of the Einstein field equations as

\[
R_{\mu\nu} = -\kappa(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu}) + \Lambda g_{\mu\nu}.
\]

(2.25)

The left hand side depends only on the metric and its curvature, and so it is clear that the cosmological constant can be on either side of the equation relating curvature and matter. Which side it is on, however, has profound implications for the constant’s interpretation. Its original position on the left makes it a curvature term. In an attempt to provide a physical interpretation it is common to consider it on the side of matter, where it is an energy term.

If we define an exotic perfect fluid with equation of state \( p = w \rho c^2 = -\rho c^2 \), we can see that the energy-momentum tensor \( T^{\mu\nu} = (\rho + p/c^2)u^\mu u^\nu - pg^{\mu\nu} \) simplifies neatly to \( T_{\mu\nu} = \rho c^2 g_{\mu\nu} \). This appears as a constant term proportional to the metric tensor, providing an interpretation of \( \Lambda \).

We can rewrite equation (2.24) as

\[
R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \kappa(T_{\mu\nu} + T^\text{vac}_{\mu\nu})
\]

(2.26)

where \( T^\text{vac}_{\mu\nu} \) is the fluid with \( w = -1 \) above. The ‘vac’ refers to the term only depending on the spacetime vacuum defined by the metric. \( T^\text{vac}_{\mu\nu} = \rho^\text{vac} c^2 g_{\mu\nu} \) suggests that the density for such a fluid is the vacuum density, which can be related to \( \Lambda \) via \( \rho^\text{vac} c^2 = \frac{\Lambda c^4}{8\pi G} \). Now we can view \( \Lambda \) not only as a cosmological constant but a quantum mechanical constant defining the energy density of the vacuum. Summing the zero point energies of the known fields in quantum mechanics unfortunately produces an estimate 120 orders of magnitude higher than that which the cosmological model predicts. Reconciling this elegant formalism of the vacuum energy in GR with the impressively discrepant results in QM is an important step towards understanding the 70% of the universe that is labelled as dark energy. It is regarded as one of the big problems in cosmology and is being addressed in the mission statements of future experiments such as Euclid.

---

\(^4\)Einstein is purported (though questionably) to have called it his ‘biggest blunder’; in a letter to Lemaitre he called it ‘ugly’. 

---

We can compare the GR weak-field limit equations at low velocity to the Newtonian gravity for a spherical mass, \( M \). This comparison presents us with another insight about the nature of the above vacuum energy term. We assume \( g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \), where \( \eta_{\mu\nu} \) is flat spacetime and \( h_{\mu\nu} \) is a small perturbation. For a perfect fluid with \( p/c^2 << \rho \), considering the \( \mu = 0 \) and \( \nu = 0 \)
2.2. \textit{Dark energy} \hfill 27

component of equation (2.25) we have $R_{00} = -\kappa (T_{00} - \frac{1}{2} T g_{00}) + \Lambda g_{00}$. In the weak-field limit $g_{00} \approx 1$ and $R_{00} \approx -\frac{1}{2} \delta^{ij} \partial_{i} \partial_{j} h_{00}$ from definitions and to first order in $h_{\mu \nu}$. From $T_{\mu \nu} = \rho u_{\mu} u_{\nu}$ we find that $T_{00} = \rho c^2$ and $T = \rho c^2$ in the low velocity limit. Since $h_{00} = 2 \Phi / c^2$ and $\kappa = \frac{8 \pi G}{c^4}$, we have the Newtonian result with the cosmological constant term included:

$$\nabla^2 \phi = 4 \pi G \rho - \frac{\Lambda}{c^2}, \quad (2.27)$$

which for a spherical mass, $M$, produces the field strength

$$g = -\nabla \phi = -\frac{GM}{r^2} \hat{r} + \frac{\Lambda c^2 r}{3} \hat{r}. \quad (2.28)$$

Here, the repulsive effect of the cosmological constant is easily observed. The gravitational field strength consists of the normal Newtonian term and the term that has been introduced by $\Lambda \neq 0$ that creates a gravitational repulsion increasing linearly with distance $r$.

Choosing an equation of state $w = -1$ for a perfect fluid is needed for the link to $\Lambda$ and the vacuum. As described above, the $T_{\mu \nu}^{\text{vac}}$ term depends only on the vacuum, but this arises from choosing $w = -1$. To see this explicitly, choosing $w = -1 \pm \delta$, where $\delta \ll 1$, would result in the energy-momentum tensor having a small velocity term such that the result would no longer be a property of the vacuum only. Most simple extensions to the $\Lambda$CDM model, however, are those where the parameter $w$ is fitted for as a constant (also known as $w$CDM), and deviates from $-1$. These are not physically motivated and instead provide only an indication of non-$\Lambda$CDM behaviour (Suzuki et al. 2012). We now turn to models that aim to describe dark energy. Firstly, we introduce quintessence models, amongst other models, that predict the shape of $w(z)$ based on physical reasoning. We then review methods which aim to study this in reverse: providing a description of $w(z)$ based on the data.

2.2.2 Theoretical dark energy models

Quintessence models describe a time-varying equation of state $w(z)$ for a missing energy component, with the name deriving directly from fifth (quint) force (essence). A scalar field is proposed to be this missing component, specifically a homogeneous scalar field that is very weakly coupled to matter (Ratra & Peebles 1988; Armendariz-Picon et al. 2000; Tsujikawa 2013). As briefly mentioned earlier, the cosmological constant encounters a problem with observations, where observational and theoretical values for it differ by many orders of magnitude (Vikman 2005; Hobson et al. 2006). Typically this problem is rephrased as two related ones: firstly, “why does it appear that the cosmological constant only starts to dominate this late in the universe’s history?”, known as the ‘cosmic coincidence problem’ (Armendariz-Picon et al.
and secondly, that this would require fine-tuning to get exactly the value we observe today, specifically that it would require an unnatural energy scale for $\Lambda$ and a very small energy density early in the universe’s history (Armendariz-Picon et al. 2000; Vikman 2005; Wang et al. 2000; Zlatev et al. 1999). Scalar fields are proposed to overcome these problems.

For a homogeneous scalar field $Q$ rolling slowly down a potential $V(Q)$, the pressure is given by $p_Q = \frac{1}{2} \dot{Q}^2 - V$ and the density as $\rho_Q = \frac{1}{2} \dot{Q}^2 + V$. Hence the pressure can be negative if the kinetic energy is less than the potential energy (Steinhardt et al. 1999). For such fields, models predict $w \in [0, -1]$ (Vikman 2005; Steinhardt et al. 1999) but extensions to these exist. “Phantom” dark energy models have the super-negative equation of state $w < -1$ at some point in their evolution (Vikman 2005; Gupta et al. 2009) (possibly at all time, as crossing the “phantom divide line” PDL of $w = -1$, is not trivial (Zhang 2009; Vikman 2005)). Further extensions in this domain are the “quintom” models that combine a quintessence and phantom model to create a 2-field Lagrangian where the quintom field behaves as either quintessence or Phantom depending on which kinetic term is dominant (with them being equal equating to $w = -1$) (Vikman 2005; Gupta et al. 2009). Quintom models may be able to cross the PDL as the contributions of the fields change with time, such that their average crosses the PDL without an individual scalar field doing so. Additional scalar field models are tracker (Zlatev et al. 1999; Steinhardt et al. 1999) and k-essence (Armendariz-Picon et al. 2000) models. Tracker and k-essence models have the similarity that both exhibit attractor solutions in their equation of state, where a large number of initial conditions leads to the same behaviour at some point. Further insights into how the various models solve the coincidence problem are contained in the references, but by and large a dynamical equation of state creates the scenario where a wider range of initial conditions can lead to our observed late-time accelerated expansion being a natural phenomenon of such models, rather than some coincidence of nature.

Here is a good time to briefly mention several alternatives to the above mentioned quintessence models. Examples include braneworld models where our 3+1 dimensional space is embedded in a 3+1+d dimensional one (Maartens & Koyama 2010), curvature quintessence models (Capozziello 2002; Nojiri 2006) where sub-dominant terms in GR become increasingly significant at low curvatures, $F(R)$ modified gravity (Appleby & Battye 2007; Hu & Sawicki 2007; Starobinsky 2007), and “extended quintessence” where the quintessence scalar couples to the Ricci scalar (Matarrese et al. 2004; Pettorino et al. 2005). Such theories may exhibit cosmic speed-up. Furthermore, interacting dark matter (Amendola et al. 2007; Clemson et al. 2012; Lu et al. 2012) may predict why dark energy is dominating at the same time as when
large scale structure forms, and anisotropic universe models (Akarsu & Kilinç 2009; Marra et al. 2013; Valkenburg 2012) suggest that the universe’s local properties can create bias in our measurements of the expansion (though it is unlikely that this accounts for the entire observed acceleration of the expansion).

Given the variety of models and predicted behaviours of $w(z)$ that may explain our expansion history, it seems necessary to have precision $w(z)$ measurements to constrain which models are sensible and within these, to constrain parameters. A way in which to test these models would be to take the predictions of each and use a model selection criterion. To do this however, we would need to test every model. More sensibly, we may acquire an idea of what observational data suggests $w(z)$ to look like. In order to do the latter we need to have model-independent constraint on $w(z)$ throughout a decent redshift range. Several techniques exist for this, with Sahni & Starobinsky (2006) providing a review of some of these. In chapter 5 we carry out a detailed analysis of such a model-independent reconstruction across various combinations of cosmological datasets. Mortonson et al. (2013) provide a good review of dark energy constraints and the Planck 2015 dark energy analysis paper (Planck Collaboration et al. 2016d) provides an in depth review of several models and reconstruction techniques to analyse dark energy.

2.3 Gravitational waves

With the LIGO consortium’s recent detection of gravitational waves (Abbott et al. 2016b,d) we find ourselves witnessing the dawn of the field of gravitational wave astronomy. It marks a new type of observation that can gaze upon the universe beyond the electromagnetic spectrum; it opens up a wealth of opportunity for astrophysical observations that both complement current electromagnetic astronomy and also for observations that are new and unique. To the latter we wish to turn in this thesis by potentially using the detected gravitational wave signals as probes of the theory of general relativity itself.

We wish to briefly review in this section some of the background theory for gravitational wave astronomy. In section 2.3.1 we review linearised general relativity and the wave equation solutions that define gravitational waves. Section 2.3.2 describes the recent detections and the prospect of multi-band gravitational wave astronomy in the coming decades. Finally section 2.3.3 describes the tests of general relativity that one can conduct using gravitational wave data, including the tests of the recently discovered GW150914 and GW151226 which showed good consistency with standard general relativity.

This short section will complete the cosmological and astrophysical background overview.
Throughout we summarise using results from Hobson et al. (2006) and the thesis of Chua (2017), the latter having collaborated in the gravitational wave project described in chapter 7. The dark energy work and, especially, the gravitational waves work aims in part to improve computational analysis techniques. An introductory chapter summarising the statistical background will be presented hereafter, and parts of this introduction may at times borrow from it due to the strong overlap between detecting gravitational waves and Bayesian inference.

2.3.1 GR wave equation

Gravitational waves are described well by the linearised form of general relativity (GR). This formalism relies on a field approximation where the background spacetime is flat and perturbed only by a gravitational wave (GW). This flat spacetime is referred to as Minkowski spacetime and is defined by the Minkowski metric $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$, in Cartesian coordinates. We can then define the spacetime with the GW perturbation as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu},$$  

where $h_{\mu\nu}$ is a small perturbation due to the gravitational wave, with $|h_{\mu\nu}| \ll 1$. We wish to solve the Einstein field equations, of equation (2.1), using this metric approximation. We linearise the solution by assuming terms of order $O(|h_{\mu\nu}|^2)$ are negligible. In linearising we observe that the Einstein tensor (composed of the Ricci tensor and scalar) depends only on second order partial derivatives of the metric, and we note that any resulting solutions are only valid when the gravitational field is weak (far away from GW sources) or when the source can be approximated as Newtonian (spatial terms in $T_{\mu\nu}$ are small). Using the trace-reversed form of the metric perturbation, with $\tilde{h}_{\mu\nu}$ satisfying $h_{\mu\nu} = \tilde{h}_{\mu\nu} - \frac{1}{2}\text{tr}(\tilde{h})\eta_{\mu\nu}$, and imposing the Lorenz gauge condition ($\partial_{\nu}\tilde{h}^{\mu\nu} = 0$) we obtain a wave equation:

$$\Box \tilde{h}_{\mu\nu} = -16\pi T_{\mu\nu},$$  

where the flat space d’Alembert operator $\Box$ is the generalisation of the Laplace operator. The $-16\pi T_{\mu\nu}$ term on the right hand side is the stress-energy tensor source for the gravitational wave. One can obtain a similar strong-field solution without the simplifying assumptions discussed above, which uses the curved-space d’Alembert operator and additional source terms relating to higher order terms in $h_{\mu\nu}$. The general solution of the wave equation is a retarded integral over the source terms. Far from the source, as would be the expected case for any gravitational waves arriving on Earth, one obtains a monochromatic plane wave:

$$\tilde{h}_{\mu\nu} = \mathcal{R} \left[ A_{\mu\nu} e^{i(k_{\mu}x_{\nu} - \omega t)} \right].$$  

---

8A common alternative convention is to use $\eta_{\mu\nu} = \text{diag}(1, -1, -1, -1)$, which does not change the following results qualitatively.
The wavevector $k_4$ is a null four-vector (such that $k_4 k^4 = 0$) and the amplitude tensor $A_{\mu \nu}$ is orthogonal to the wavevector (such that $A_{\mu \nu} k^\mu = 0$). The Lorenz gauge condition and further gauge choices leave only two degrees of freedom in equation (2.31), whilst also enabling us to choose the transverse-traceless gauge such that $h_{\mu \nu} = \bar{h}_{\mu \nu}$. These remaining two degrees of freedom are usually expressed as the plus ($h_+$) and cross ($h_\times$) linear polarisation modes. Defining a plane wave as travelling along the $z$-axis, the polarisation modes in transverse-traceless gauge are defined as

\begin{align}
    h_+ &\equiv h_{11} = -h_{22}, \\
    h_\times &\equiv h_{12} = h_{21}.
\end{align}

This polarisation manifests itself as measurable changes in distance relations between objects in spacetime. Specifically, the spatial distance in relativity theory between points in spacetime changes depending on the gravitational wave properties. For two points separated by distance $l$ in the $x$-direction, the presence of a purely plus polarised gravitational wave would alter the spatial distance by $(g_{11})^{1/2} l \approx (1 + h_+ / 2) l$, whilst for two points separated in the $y$-direction by $l$ the separation goes as $(g_{22})^{1/2} l \approx (1 - h_+ / 2) l$. A similar effect is observed for purely cross polarised waves, as these are just a $45^\circ$ rotation in the plane of polarisation of the plus polarised waves. We note that the distortions in the time dimension are negligible when far from the source (when using the transverse-traceless gauge). The effect is therefore compression along one axis and stretching along the other axis as the signal $h_+$ oscillates with time.

The effective compression and stretching can be measured. This measurement is what the large gravitational wave observatories, such as LIGO, are concerned with; and it is a marvel of precision science\footnote{The usual quote is that LIGO’s accuracy is equivalent to measuring the distance to the nearest star to within the width of a human hair.}. As the wave passes through a region of space time, the $x$ and $y$ distances change ever so slightly. The magnitude of this can be interpreted via the dimensionless strain $h = h_+ + i h_\times$, which is given by the evolution of the gravitational wave source. An ability to detect this strain would provide a powerful probe on cosmological scales as the gravitational waves only interact very weakly with matter: the information carried by the waves can tell us much about the source. For this reason, and others, they are often referred to as “standard sirens”. This desirable property that allows them to travel through space without distorting the information they carry also makes them notoriously difficult to detect, as the interaction with the matter in our detectors is very slight.
2.3.2 Detection

When a sufficiently strong gravitational wave passes through a detector it produces a time varying strain signal $h(t)$. Specifically, two or more detectors can measure two independent detector-response functions $h_I$ and $h_{II}$ and reconstruct $h_+$ and $h_\times$ from these. In order to identify this signal amongst the background noise that is present in a detector, one can use template waveforms for modelled sources and identify using Bayesian inference whether a particular template matches well with the received signal. A good overview of this process is presented in Chua (2017) and citations therein.

Commonly modelled waveforms are those for coalescing binary systems, which produce intense gravitational waves in detectable frequencies and are expected to occur frequently in the local universe within reach of detector sensitivities. Chapter 7 uses a Kerr waveform model for a black hole and compact object (such as a neutron star) binary system and this will be discussed in more detail there. The waveform is essentially an oscillating signal during the inspiral phase whose frequency and amplitude increases during the merger phase, with amplitude declining in the post-merger-ringdown phase; the increasing frequency and amplitude have earned this style of signal the name ‘chirp’.

Other sources include continuous sources, such as asymmetric and rapidly spinning compact objects, and burst sources from short-lived high energy astrophysical processes. Generally the gravitational wave universe is expected to also be filled with background noise from various source types that are not sufficiently strong to be resolved by detectors. These superimpose to form a correlated stochastic noise background that can be searched for in a detector over longer periods of observation.

The gravitational wave events GW150914 (Abbott et al. 2016b) and GW151226 (Abbott et al. 2016d) detected by LIGO were both of the compact binary coalescence type. Specifically, the detected signals describe the merger of two stellar-origin black holes, of varying masses for the two events. A third candidate event LTV151012 was also observed to be consistent with a black hole merger (Abbott et al. 2016a). The LIGO collaboration’s papers summarising binary black hole mergers in the first observing run provides more detail on these events (Abbott et al. 2016a).

GW sources with a characteristic mass of less than $10^3$ solar masses, such as the observed signals in LIGO, radiate in the kilohertz range ($10^1 - 10^4$ Hz). Ground based detectors such as Advanced LIGO and Advanced Virgo (Collaboration 2015), as well as potential future experiments KAGRA (Somiya 2012) and the Einstein telescope (Punturo et al. 2010), are designed to be sensitive to signals in this frequency range. Millihertz frequency band signals ($10^{-4} - 10^0$) include super-massive black holes producing longer-lived signals detectable by
2.3. Gravitational waves

space based detectors (Ni 2016) such as LISA (Amaro-Seoane & et al. 2017), and in the more distant future DECIGO (Kawamura & et al. 2011) and TianQin (Luo & et al. 2016). Gravitational waves in the nanohertz frequency band ($10^{-9}$ to $10^{-6}$) can be detected using pulsar timing arrays (PTAs): PTAs attempt to use precise timing information from pulsars in the Milky Way to detect spatial distortions from gravitational waves over galactic distances. An international collaboration of three PTA teams (Hobbs & et al. 2010) is working on this, and detections might occur within the next 10 years (Taylor & et al. 2016).

In the coming decade, the prospect for multi-band gravitational wave astronomy looks very promising (Sesana 2016; Vitale 2016). Among the myriad of interesting potential developments is the ability of space based detectors to identify black hole binary mergers in advance, ready for detection by ground based gravitational wave observatories and concurrent electromagnetic surveys for a range of astrophysical data sources relating to the event. This pre-empting could additionally improve tests of GR by improving the ability to constrain the waveform model parameters (Vitale 2016).

2.3.3 GW astrophysics and tests of GR

Gravitational waves are produced by sources in the strong field regime of general relativity. Testing GR using GWs therefore provides insights into the more extreme scenarios in which we might expect GR to be violated. This can include simple tests of whether the matched GR waveform sufficiently explains the observed signal, or whether any general phenomenological deviations from GR are favoured, or tests of specific models themselves. Typically tests are carried out on individual events, but with more frequent future detections it will be possible to combine results to place tight constraints on any violations of GR. Here we briefly summarise the type of tests already carried out on the existing GW detections.

Tests of GR using the GW150914 detection were conducted by Abbott et al. (2016c) and broadly found no violation of GR. Initial tests included subtracting the most probable waveform from the detection and observing whether the residual signal is consistent with detector noise (which it was found to be). Thereafter a consistency check was made for the phases before and after inspiral in the binary coalescence signal. In analysing the signal split into 2 different phases along the time axis (inspiral and merger-ringdown), a consistency check can be made on the recovered waveforms in both phases. If there was a large discrepancy between the fit of the sections it might suggest that a transient glitch caused the signal, but for GW150914 this was again ruled out and the signal is consistent across the three phases with the GR model waveform.

Abbott et al. (2016c) also test parameterised deviations from the GR binary waveform phase coefficients. These tests are conducted on 3 phases of the coalescence process: early-
inspiral stage, intermediary stage (transition between inspiral and merger-ringdown) and the late merger-ringdown stage. For each stage, phase coefficients are introduced that characterise fractional changes to each of the GR phase parameters in the baseline GR model. Note that in the early stage these coefficients are the post-Newtonian (PN) theory phase coefficients, a commonly used analysis for deviations from GR. The results for the 3 stages again conclude that no deviation for GR is found throughout the binary coalescence. The analysis shows that parameter constraints on phenomenological deviations are consistent with GR, and also that Bayesian model selection using Bayes factors favours (slightly) the GR model over any extensions using phenomenological parameters.

Abbott et al. (2016a) expand on these same tests by extending the analysis to the GW151226 gravitational wave detection, and also combining the constraints from both to obtain even tighter phenomenological parameterisation constraints. Specifically, the early stage PN coefficients are more tightly constrained by GW151226 as more wave cycles fell within the detector frequencies, whilst the late time phenomenological parameters are poorly constrained due to the lower signal strength of this second GW detection. The residual signal tests are less insightful for GW151226 as the signal strength is too low.

The LIGO collaboration analysis is expanded on by Yunes et al. (2016), who discuss constraints on a wide range of generation and propagation mechanisms, alongside providing a good review of GR tests using GWs. As this thesis delves into the phenomenological parameterisation type of tests, we will not discuss the many other useful insights into GR that gravitational waves can provide. Another good review discussing model independent and model dependent tests of GR (specifically in relation to black holes) can be found in Yagi & Stein (2016), which includes a brief description of the PN parameterisation and discusses how gravitational and electromagnetic wave measurements can complement an analysis. For a comprehensive review on ground, space and PTA based GW detections and tests of GR, please see Yunes & Siemens (2013).

The work in chapter 7 aims to improve the efficiency of conducting a phenomenological test of deviation from GR waveforms. It builds on the tests proposed by Li et al. (2012), in which waveform phase deformation coefficients are systematically searched over to create a Bayesian model selection criterion on how favourable GR is compared to a generic deviation away from GR. This proposed test is very similar to the PN and other phenomenological parameterisation tests discussed in the LIGO GR test papers. These tests are model-independent in that they require no alternative to GR in order to check the consistency of GR with a detected signal. Chapter 7 will discuss in more detail the test of Li et al. (2012) and the related improvements obtained.

This concludes the GR introductory sections of the thesis, as well as the cosmological and
2.3. *Gravitational waves*

dark energy sections. We move on now to the statistical framework that is indispensable when attempting to make sense of any cosmological or gravitational wave data.
A typical thesis preface to Bayesian statistics might extol the virtues of the Bayesian approach over the frequentist approach. The dichotomy between frequentist and Bayesian is a valid one in the field of statistics: a frequentist approach defines probability as a distribution of outcomes from frequently repeated trials, whilst a Bayesian approach defines probability as a degree of belief in an outcome. Instead of describing why the Bayesian approach is more suited, we note instead that for cosmology there is only one game in town: Bayesian statistics is the preferred method of data analysis in cosmology. Trotta et al. (2008) provides a clear description of the dichotomy and highlights the growth in Bayesian applications up to 2007.

Here we review the Bayesian inference framework to constrain model parameters as well as to compare models themselves. Statistical theory is covered as well as computational implementation. Section 3.1 describes Bayes theorem and the functions required to use it. Section 3.2 describes the nested sampling procedure, which presents an effective method of conducting parameter estimation and model selection. Nested sampling is used throughout this work. Section 3.3 concludes with an overview of the codes used to turn the statistical theory into computation that constrains our cosmological and gravitational models.

Throughout this chapter we use the brilliant textbooks by MacKay (2003) and Sivia & Skilling (2006). The nested sampling discussion is based on the seminal papers by Skilling (2004) and Skilling (2006). The theses of Handley (2016) and Vázquez (2013) have been used for further inspiration, with the former providing a fluent overview of Bayesian inference and nested sampling, and the latter particularly clear on $\Lambda$CDM within this framework. Additional papers and reviews are cited as used, and these influence the wider points around their citation.
3.1 Bayesian inference

Bayesian inference provides a mathematical formalism to answer “what is the probability of my parameters given my observed data?” At first glance this may seem quite easy, the human mind is very good at inferring patterns from observations: from determining how happy or sad our friends are, to deducing the mng f sntncs without vowels. Formal logic deals less well with such problems. Specifically, the type of reasoning involved with mathematics is effective at starting from a set of assumptions and deducing the next steps (deductive reasoning) as a one-to-one or one-to-many function. The reverse is more difficult as there may be a host of possible starting points that can lead to the same outcome, often with no deductive definition to determine them.

Taking the above sentence “mng f sntncs” (intentionally without vowels) as an example, we see that there is no single deductive argument that leads back to the parameters that were originally present. How could we create an algorithm to decipher it? It seems mathematically sensible to take each vowel-less word and construct a set of all permutations of vowel-full words (with a word size limit, such as having no more than three consecutive vowels). For a word with \( N \) consonants there are \((1 + 5 + 5^2 + 5^3)^{N+1}\) combinations, equalling 24336 for the word ‘f’.

Using an acceptable word database we can search and pick out candidate words, on order 10 for ‘f’ depending on dictionary. At this point we still only have a list of possible parameters, and a combination of all candidate permutations for the 3-word vowel-full sentence. Now we might enter each candidate sentence into a search engine and allow the wealth of human knowledge (essentially another database, of sentences) to determine the probability of having received the vowel-less sentence for each set of parameters. We find, for example, that “meaning of sentences” was used in 109 million websites whilst “moaning oaf santonicas” was used in 80 websites, hence the former is a million times more likely, according to our algorithm.\(^a\) Note that we cannot be certain what was meant: scientific enquiry does not deal with truths and certainties but instead with models that fit data accurately.

The described algorithm provides an inductive argument for the probabilities of what the sentence is. It starts by defining our parameter space before we have knowledge of the sentence, specifically it says our sentence must be composed of words in the specified dictionary of acceptable words, and that no word has more than three consecutive vowels. In the jargon of Bayesian inference this is the prior on our parameter space. The next step is to find the likelihood that the vowel-full words could have created the received sentence. We constructed all acceptable sentences and used a large language usage database to define this likelihood based on frequency of usage in the dictionary. This is usually termed the likelihood function. If we

\(^a\)Spellcheckers work on a similar search, permute and recommend structure, with efficient searching and additional complexities in the recommendation.
correctly normalise over all combinations we finally obtain a probability distribution on the total word dictionary, which is termed the posterior in Bayesian inference. The posterior describes the probability of our vowel-full sentence given the vowel-less sentence data. Needless to say our human mind did not conduct the same searches, but it is hypothesized that brains generally use Bayesian inference principles, an interesting review of the “Bayesian brain” model and related topics is presented by Clark (2013).

The above example hints at the structure of Bayes theorem and its usage in inference. We will now define it more rigorously.

### 3.1. Parameter estimation

Bayes theorem is stated as,

\[
\Pr(X|Y, I) = \frac{\Pr(Y|X, I) \times \Pr(X|I)}{\Pr(Y|I)}.
\]

(3.1)

Here, \(X\) and \(Y\) are propositions and \(\Pr(X)\) specifies our belief that the proposition \(X\) is true. \(I\) is the background information and the constructs “,” and “|” can be read as “and” and “given that” respectively, where “|” implies that everything to the right is taken to be true. Therefore \(\Pr(X|Y, I)\) can be read as “the probability that \(X\) is true given that \(Y\) is true and that we have some background information \(I\) that is true”. Bayes theorem written as in equation (3.1) does not immediately seem useful for determining what, for example, we believe the value of \(\Omega_A\) to be in our cosmological model. Let us substitute the symbols to be more intuitive: we want to know what a parameter \(\Theta\) is, given some data that we have, \(\mathcal{D}\), whilst assuming that our model \(\mathcal{M}\) is true. It is common notation to drop the dependence on \(I\) as it can be assumed that each probability depends on the background information. Constructing the probability \(\Pr(\Theta|\mathcal{D}, \mathcal{M})\) allows us to define Bayes theorem for parameter estimation:

\[
\Pr(\Theta|\mathcal{D}, \mathcal{M}) = \frac{\Pr(\mathcal{D}|\Theta, \mathcal{M}) \times \Pr(\Theta|\mathcal{M})}{\Pr(\mathcal{D}|\mathcal{M})} = \frac{\mathcal{L} \times \pi}{\mathcal{Z}}.
\]

(3.2)

\(\Pr(\Theta|\mathcal{D}, \mathcal{M})\) is the posterior probability, it defines our parameter constraints after taking data into account. To calculate the posterior we therefore need to be able to compute 3 separate quantities. \(\Pr(\mathcal{D}|\Theta, \mathcal{M}) = \mathcal{L}\), the likelihood, is the probability of having obtained our data for a certain value of \(\Theta\). \(\Pr(\Theta|\mathcal{M}) = \pi\), the prior, is the probability of obtaining \(\Theta\) before we take the data into account, based only on our a priori knowledge. \(\Pr(\mathcal{D}|\mathcal{M}) = \mathcal{Z}\), known as the evidence or marginal likelihood, acts as a normalisation factor to ensure that our posteriors obey \(\int \Pr(X) dX = 1\).

The calculation of \(\mathcal{L}\) defines the probability of obtaining the data given some proposed parameters, for example, we can compute the probability that a coin lands on heads 9 times out of 10 (our data) given that we are told that the coin is unbiased (our parameter). In
cosmological applications the datasets are complex and likelihood calculations are non-trivial, and we will return to it shortly. The prior on the other hand is simpler. For priors, ignorance over a given wide range of possible parameter values is typically assumed and produces ‘flat’ priors. Other commonly used priors include Gaussian priors, which characterise that we have a favoured expectation value, and logarithmic priors that characterise our ignorance on the order of magnitude of a parameter. Prior choice is often criticised as being subjective, but flat and similarly descriptive priors are well justified to represent a given knowledge of the parameters. The maximum entropy technique can also be used to define priors (Jaynes 1968; Caticha & Preuss 2004) to remove ambiguity further.

As mentioned, in order to ensure that the posterior probabilities are true probabilities, the evidence acts as a normalisation constant. The evidence can therefore be calculated by,

\[ Z = \text{Pr}(\mathcal{D}|M) = \int_{\text{all } \theta} \text{Pr}(\mathcal{D}|\theta, M) \times \text{Pr}(\theta|M)d\theta = \int_{\text{all } \theta} L(\theta)\pi(\theta)d\theta. \]  (3.3)

As we know in principle how to calculate each of the quantities on the right hand side of equation (3.3), we can derive our posterior probabilities. This solves the problem of induction by using the deductive prior, likelihood and evidence functions. This process can be thought of as starting with a prior expectation on your parameter constraints and updating these expectations using data via the likelihood term.

The likelihood function is central to Bayesian inference. For flat priors the shape of the posterior is defined entirely by the likelihood function, with the prior and evidence adjusting the normalisation. For non flat priors the process of updating the parameter constraints is also entirely dependent on the likelihood function. Constructing the likelihood function for a complex dataset is non-trivial. In our cosmological model investigations this process will be done by the experiments that gather the data. After the data has been cleaned and compressed into a suitable format (such as a power spectrum, BAO measurements or distance-redshift relations) one can construct the probability that we would observe this data if the universe was as modelled by the parameters. For example, if our BAO measurements show that the scale factor is growing exponentially in late times then a model which suggests a big crunch will start to significantly diverge from the BAO data points. The likelihood function has to capture this numerically and can do so by, for example, computing differences between data points and the model prediction.

Posterior distributions are typically computed using either Markov chain Monte Carlo (MCMC) or nested sampling methods. MCMC samples from the posterior at random, with one or more chains exploring the parameter space to preferentially sample the higher probability regions. The MCMC proposal distribution, which defines how the chains explore the space, is chosen such that the density of samples will be proportional to the posterior distribution. From
3.1. Bayesian inference

this one can find the peak in a likelihood and therefore the significant region of the posterior probability. Nested sampling is an alternative approach to exploring the parameter space and will be discussed in detail in section 3.2

3.1.2 Model comparison

The above considerations provide the framework for estimating parameters, where \( \Theta \) can represent any number of parameters that we need in our model as the equations generalise easily for multiple dimensions. For model comparison we can use a similar approach. This time we may attempt to find the probability of a model \( M_1 \) given the observed data:

\[
\Pr(M_1|D) = \frac{\Pr(D|M_1) \times \Pr(M_1)}{\Pr(D)}.
\]  

(3.4)

If we wanted to compare the probability of a second model \( M_2 \), we can take the ratio of probabilities and use Bayes theorem:

\[
\frac{\Pr(M_1|D)}{\Pr(M_2|D)} = \frac{\Pr(D|M_1) \times \Pr(M_1)}{\Pr(D|M_2) \times \Pr(M_2)} = \frac{Z_1 \pi_1}{Z_2 \pi_2}.
\]  

(3.5)

where in the last step we have noted that the probability of the model (dependent only on the implicit background information) is its prior. If we take the logarithm of equation (3.5) we obtain the posterior odds ratio (POR; \( \mathcal{P}_{ij} \)) between two models \( i \) and \( j \):

\[
\mathcal{P}_{ij} = \ln \left( \frac{\Pr(M_i|D)}{\Pr(M_j|D)} \right) = \ln \left( \frac{Z_i \pi_i}{Z_j \pi_j} \right) = \ln(Z_i \pi_i) - \ln(Z_j \pi_j)
\]  

(3.6)

If we assumed equal belief in both models *a priori* then we obtain the more commonly used Bayes factor \( B_{ij} \) between two models, which quantifies the relative support in the data between the models:

\[
B_{ij} = \ln \left( \frac{\Pr(M_i|D)}{\Pr(M_j|D)} \right)_{\pi_i=\pi_j} = \ln(Z_i) - \ln(Z_j)
\]  

(3.7)

We see that to compare models we only need to calculate evidences and state our prior beliefs. This allows us to easily quantify our degree of belief in one model compared to another. A positive Bayes factor means that \( M_i \) is favoured and a negative Bayes factor means that \( M_2 \) is favoured, with qualitative assessment given by the Jeffreys guideline (Jeffreys 1961) described in section 3.1.3. The Bayes factors and PORs can be defined with or without the logarithm, and the convention to use the logarithm is assumed here. We note too that it is common to use equal priors, such that Bayes factors and PORs are equivalent. 

Equation (3.3) shows that the evidence calculations is a multi-dimensional integral over the whole parameter space, weighted by the prior. Naively one might wish to integrate this by quadrature: creating an interpolation grid across the parameter space and computing approximate volumes. However, even for modest dimensionality this brute force approach becomes
Chapter 3. Statistical Framework

intractable due to the many likelihood evaluations required as the algorithm scales exponentially with the number of parameters in the space. Additionally, the cosmological and gravitational wave models used in this thesis produce a computationally expensive likelihood calculation\(^\text{b}\), such that this approach does not work.

The MCMC approach, which works well for determining posteriors, is not reliable in estimating evidences for a number of reasons, including that MCMC does not sample the tails of distributions sufficiently well for the integral. Alternative methods are required. There are many good reviews describing these, such as the clear overview of methodologies presented in Friel & Wyse (2011), the discussion in Clyde et al. (2007) of relative merits of methods, and the wide methods discussed in Knuth et al. (2014) (as well as the textbooks used throughout this section). Applications directly to cosmology are also reviewed by Liddle et al. (2006); Trotta (2008). We quickly summarise a few techniques and discuss specifically why we have settled on nested sampling for our Bayesian inference problems.

**Laplace’s method** (Tierney & B. 1986) makes the assumption that the posterior distribution is approximately Gaussian. Expanding in parameter space produces an evidence estimate that depends only on the mean and covariance of the distribution, with the covariance given by the (Hessian) matrix of second order derivatives. This approximation will not be useful for our discussions as the Gaussian approximation is not suitable for complex likelihoods. Other approximations exist, such as the harmonic mean estimator (Newton & Raftery 1994), but again we avoid these.

**Thermodynamic integration** (also known as simulated annealing) (Kirkpatrick et al. 1983; Gelman & Meng 1998) uses MCMC runs with the likelihood raised to a power termed the inverse temperature \(\beta\) of the run. The inverse temperature starts at \(\beta = 0\), such that the likelihood is the prior, and increases via some schedule to the maximum \(\beta = 1\), producing the posterior. Averaging over the likelihood at a given inverse temperature, \(\langle \mathcal{L} \rangle_\beta\), allows the evidence to be computed as \(\log Z = \int_0^1 \langle \log \mathcal{L} \rangle_\beta \ d\beta\) (Sivia & Skilling 2006). The approach requires several runs of the MCMC exploration to produce an evidence such that it is generally more computationally intensive to obtain evidence calculations than parameter estimates and prohibitive for large parameter space models (Liddle et al. 2006). The method relies on defining a temperature schedule which achieves an accurate evidence calculation, and there are specific examples where the approach is not valid (see Sivia & Skilling (2006, section 9.6) for a discussion of how nested sampling overcomes these) alongside other extensions to the method to deal with shortcomings (Knuth et al. 2014). The parameter spaces explored in this thesis extend up to 30 dimensions, such that this method is too expensive (alongside likely shortcomings due to the complexity of the multi-modal parameter space).

\(^{\text{b}}\)On order of seconds for a computing core in the Cambridge Darwin HPC system
Savage-Dickey density ratio (Dickey 1971; Verdinelli & Wasserman 1995, SDDR) can be used to compute Bayes factors between two models that are nested: where one model is contained in the other by appropriate parameter choice $\Theta_s$. By running the ‘larger’ of the two models (the one that contains the nested model by setting parameters to $\Theta_s$) one can obtain an expression for the Bayes factor from the posterior and prior functions evaluated at $\Theta_s$. Extensions beyond the original formalism exist to broaden the scope of situations when the SDDR can be used (Cameron 2013). Trotta (2007) introduces this concept into cosmology, demonstrates its efficacy and discusses its expected range of applicability. As this approach is limited to nested models we do not adopt the method in this thesis.

Other MCMC methods such as reversible jump MCMC (Green 1995, RJMCMC) and product space MCMC (Carlin & Chib 1995, PSMCMC) attempt to compute Bayes factors directly from the MCMC chains, see Han & Carlin (2001) for a comparative review. PSMCMC samples over a discrete parameter that combines the parameter space of several models to create a joint space as the product of individual model spaces. This discrete parameter requires alteration of typical MCMC techniques and the method requires careful selection of priors to ensure accurate sampling of the increasingly complex space (Han & Carlin 2001). RJMCMC avoids sampling over the entire product space by defining a proposal rate for jumping to a new model, typically defining a more complex algorithm. Such methods have problems dealing with large dimensionality and expensive likelihoods (Knuth et al. 2014), and as a result are deemed unsuitable for our cosmological applications. Importance sampling can be used to estimate evidences too, whereby the posterior is averaged over a function that is simpler to sample from (Neal 1993), but it requires reliably constructing this simpler function (Knuth et al. 2014) and is not used in this thesis (beyond its implementation in MultiNest (Feroz & Skilling 2013)).

Nested sampling (Sivia & Skilling 2006; Skilling 2004; Skilling 2006, NS), to be described in section 3.2, provides a method of obtaining an evidence calculation from a single run. The accuracy of the evidence calculation depends on the thoroughness of exploration of the parameter space, such that dealing with highly multi-model parameter spaces, high dimensionality or curved degeneracies is an implementation problem and not a methodological problem. Specifically, nested sampling has no requirements on the shape of the likelihood space it can explore, it does not require the use of local gradient information and no a priori setting up of functions that requires information of the likelihood. The main problems with nested sampling are reported to be its computational cost (Friel & Wyse 2011) and a difficulty in sampling accurately (Knuth et al. 2014) but, in recent years, significant progress has been made on efficiency (Handley et al. 2015a,b) and an increasing body of literature successfully using these approaches shows promise in the fields of cosmic shear (Joudaki et al. 2017), pulsar timing arrays (Lentati et al. 2016), galaxy cluster parameterisation (Rumsey et al. 2016),...
| $|B_{ij}|$ | Odds   | Probability | Qualitative conclusion |
|--------|--------|-------------|------------------------|
| < 1.0  | $\lesssim 3 : 1$ | < 0.750 | Indistinguishable       |
| 1.0    | $\approx 3 : 1$  | 0.750   | Slight                 |
| 2.5    | $\approx 12 : 1$ | 0.923   | Significant            |
| 5.0    | $\approx 150 : 1$| 0.993   | Decisive               |

Table 3.1: Table of Jeffreys guideline for assessing a Bayes factor. Jeffreys guideline concerns the far left and far right columns that relate significant Bayes factor values to their qualitative assessment, as a general approach to assessing the Bayes factors that is widely agreed on. The middle two columns, reproduced from Trotta (2007, table 1) but easily calculated, relate the Bayes factors to more intuitive probabilities: the odds of one model being favoured over another and the probability of the favoured model being correct, respectively.

and cosmological inflation (Planck Collaboration et al. 2016f), to name a few (alongside the publications associated with this thesis)\textsuperscript{c}.

### 3.1.3 Jeffreys guideline and the Occam factor

To qualify the results of model comparison using posterior odds ratios, or Bayes factors, it is common practice to use Jeffreys guideline (Jeffreys 1961). For two models $i$ and $j$, Jeffreys guideline allows us to conclude whether the data prefers either of the models. Table 3.1 summarises Jeffreys guideline. From equation (3.6) we see that $P_{ij} = P_{ji}$, and similar for $B_{ij}$ from equation (3.7). The Jeffreys guidelines therefore tell us that a log evidence difference between two models $\pm 1$ suggests the two are indistinguishable in statistical probability, and provides a qualitative scale up to $\pm 5$, at which we decide one or the other models is decisively favoured.

To preface some of the discussions on Bayes factors in the following chapters, it helps to understand the Occam factor and our expectation for evidence calculations between models. The Occam’s razor principle is that a simpler theory should be favoured over a more complex theory if both fit the data equally well. An Occam factor is the amount by which a more complex model is penalised. Bayes theorem incorporates this effect automatically in the construction of the evidence due to the posterior and prior construction. Following MacKay (2003, page 349), let us review a simple hypothetical example of a 1 dimensional model, where $Z = \int_{\Omega} L(\theta) \pi(\theta) d\theta$ has a well peaked Gaussian likelihood and a flat prior. The evidence can be estimated using

\textsuperscript{c} Citations are for papers utilising the PolyChord algorithm, and we note that the slightly older MultiNest algorithm has brought nested sampling to many more fields.
Laplace’s method as the height of the integrand times its width:

\[ Z = L(\theta_{\text{max}}) \times \pi(\theta_{\text{max}}) \sigma_w |D \]

\[ = \frac{L_{\text{max}}}{\text{Fit}} \times \frac{\sigma_w |D}{\sigma_w} \],

\[ \quad \text{(3.8)} \]

\[ \quad \text{(3.9)} \]

where \( \theta_{\text{max}} \) is the parameter for the peak of the likelihood, \( \sigma_w |D \) is the width of the Gaussian likelihood, and \( \pi(\theta_{\text{max}})=1/\sigma_w \) for a flat prior (such that \( \sigma_w \) is the prior range or width). We can then identify this as a maximum likelihood term multiplied by a term which is the ratio of the posterior width to the prior width. This latter term is always less than or equal to one, as the posterior cannot be larger than the prior by definition. We identify the latter term as the Occam factor associated with the parameter when fitting the data\(^\text{d}\). A parameter whose posterior width is very small is one which requires much fine tuning to fit the data, and the Occam factor penalises it more than a parameter which describes the model equally well but accepts a wide range of values.

We can use the Occam factor to understand Bayes factor ranges. We will see that typically our model comparisons involve a base model, such as \( \Lambda \)CDM or general relativity as a theory, and add to it an extension, such as dark energy or non-GR parameters. In such cases, if we add a parameter that does nothing to improve the constraining power of the model, the term \( L(\theta_{\text{max}}) \) is unchanged. The addition of the parameter, however, incurs an Occam factor defined by the width of the parameter compared to the prior width. Occam factors due to compressing the prior space down by a fraction of \( \frac{1}{2} \), \( \frac{1}{5} \) and \( \frac{1}{10} \) are \(-0.7\), \(-1.6\) and \(-2.3\) respectively. A model with a poorly constrained parameter may therefore be practically indistinguishable from a model without that parameter. If the data are completely insensitive to an additional parameter \( \theta' \) then \( \Pr(\theta' |D)=\pi(\theta') \) and the evidence is unchanged. We note that Occam factors add in log space, such that including additional parameters that do not improve the fit to data significantly creates a gradual reduction in Bayes factors (Sivia & Skilling 2006, page 93).

The general interpretation of the Occam factor approximation extends to multiple dimensions and to non-Gaussian likelihoods. Adding parameters to a base model will increase the evidence if it improves the likelihood fit to the data, and it necessarily incurs an Occam factor. In this thesis we will use the Occam factor approximation to determine whether an additional parameter has added to the maximum likelihood: if the change in evidence is equal to the approximate Occam factor based on the contraction of the parameter’s posterior, then we conclude it has not significantly improved the likelihood fit to the data. This provides a stricter analysis than Jeffreys guideline alone which can overlook that an ‘indistinguishable’ model has merely added a poorly constrained parameter without improving the descriptive power of the model.

\(^\text{d}\)We take the logarithm of this for comparison on the same scale as Bayes factors.
3.2 Nested sampling

As discussed in section 3.1.2 there are several methods to compute Bayes factors which rely on Markov chain Monte Carlo techniques (MCMC) for parameter exploration. Here a Markov chain is a sequence of points in the parameter space where the probability of choosing a subsequent point depends only on the properties of the current point. Transition probabilities define the probability distribution of the possible next points from the current one such that the density of steps traces out a chosen distribution, given sufficient time (Neal 1993). A Monte Carlo method is one which uses repeated random sampling to obtain a result. Visually we can consider this as a random walk on a hill, the hill is the posterior function, where the walker chooses a random direction each step with a slight preference to walk uphill. The density of footprints allows us to calculate the height of the hill, and conveniently we can recruit multiple walking friends to deal with complicated terrains. Tuning parameters can define how steep the hill is or how big the steps are (as a function of direction).

Nested sampling provides an alternative to the MCMC method. Nested sampling is a Monte Carlo method which samples randomly from an increasingly restricted prior volume to compute an evidence calculation, and parameter estimation as a by-product of the exploration. Visually, nested sampling is more akin to teleporting randomly around the metaphorical hill, but such that the teleporter rejects locations that would put you lower down the landscape. The evidence is calculated using your altimeter and the assumption that you move towards the centre of the hill by the same average step each time. There are many scenarios where we can envisage the teleporting method being more effective, such as for multiple hills (if there are more hills than walkers in MCMC, some may be missed as an individual walker is unlikely to traverse valleys) or having to find a hill in a very large area (the MCMC walkers will spend much time on flat ground whilst the teleporter exponentially reduces the search area). In MCMC many challenges are overcome by carefully tuning the walking behaviour. In nested sampling the challenge is designing an effective teleporter.

The papers by Skilling (2004); Skilling (2006) present the method more fully, here we present an overview with a focus on the robustness of evidence calculations and the estimates on error. The MultiNest papers (Feroz & Hobson 2008; Feroz et al. 2009, 2013) and PolyChord papers (Handley et al. 2015a,b) present good introductions with a computational implementation focus. Detailed analysis on the errors on evidence calculations is presented in Keeton (2011) whilst an analysis on the errors of parameter estimation is presented in Higson et al. (2017).
3.2. Nested sampling

3.2.1 Nested sampling evidence calculation

Nested sampling turns the multi-dimensional evidence integral in equation (3.3) into a one dimensional integral:

\[ Z = \int_{0}^{\infty} X(L) dL, \quad (3.10) \]

where \( X(L) \) is the survival function of the likelihood. It is known as either the “prior mass” or the “prior volume”, and can be computed as

\[ X(L_1) = \int_{\{\theta: L(\theta) > L_1\}} \pi(\theta) d\theta. \quad (3.11) \]

The prior mass integral is evaluated over the region of parameter space contained within the iso-likelihood contour. The iso-likelihood contour is defined as the set of points in the parameter space which have the same likelihood value: \( \{\theta: L(\theta) = L_1\} \). The prior mass function \( X(L) \) therefore describes the total prior volume contained in the iso-likelihood contours at each likelihood value. We observe that \( X(L) \in [0, 1] \), where \( X(L) = 1 \) when \( L = 0 \) (we are integrating over the entire prior space) and \( X(L) = 0 \) when \( L = L_{\text{max}} \) (we have found the peak). Using \( L(X) \), the inverse of \( X(L) \), the evidence integral can be rewritten as

\[ Z = \int_{0}^{1} L(X) dX. \quad (3.12) \]

If we can numerically calculate the likelihoods \( L_i = L(X_i) \) for a set of decreasing values of \( X_i \), then the evidence is trivially given via standard quadrature methods as the weighted sum

\[ Z \approx \hat{Z} = \sum_{i=1}^{N} L_i w_i, \quad (3.13) \]

where the weights are defined by the integration rule and are functions of the prior mass\(^*\).

Nested sampling therefore requires a different type of exploration to the MCMC random walks: the nested sampling algorithm works by starting the exploration at low values and gradually exploring ever higher likelihood regions. The prior mass can be shown to shrink approximately exponentially with iteration \( i \): \( \langle X_i \rangle \approx e^{-i/N_{\text{live}}} \). The evidence can therefore be calculated from the computed likelihood values and the estimated prior masses. Additionally, the posteriors can be computed from the nested sampling exploration by creating posterior samples at each point sampled (Skilling 2006). Errors and parameter estimation are discussed in more detail in section 3.2.3.

Figure 3.1 describes the conversion from a 2D likelihood map to the 1D prior mass function. Plot (a) shows a coloured contour map of the likelihood space. Example points 1 to 4 of the exploration are highlighted by their positions and likelihood values \( L_1 \) to \( L_4 \). For each curve,

\(^*\)For example, the trapezium rule is used in POLYCHORD such that \( w_i = \frac{1}{2}(X_{i-1} - X_{i+1}) \).
the prior mass is the area enclosed by the contours. The points must be ordered such that $X$ is decreasing, which is equivalent to choosing points of increasing likelihood. Plot (b) follows the function $L(X)$ as the algorithm progresses through the 4 points. The evidence is then a numerical integration of plot (b).

To facilitate faster and more accurate exploration, multiple points are evolved through the likelihood simultaneously. These points are called “live points”. $N_{\text{live}}$ live points are initially scattered randomly across the parameter space. The algorithm evolves in iterations which replace the lowest likelihood live point with a new randomly drawn point which has higher likelihood (thus ensuring that prior mass is reduced by each iteration). The newly sampled point must be drawn randomly from within the entire iso-likelihood contour with uniform probability across the space. Exploration that does not uniformly sample within the entire iso-likelihood will bias the evidence calculation, such that the final value of the evidence is wrong or such that the estimate of the error on the evidence is wrong.

Finally, we note that searching within an iso-likelihood contour becomes increasingly difficult with higher dimensional spaces, multi-modal spaces, or highly degenerate spaces. Accomplishing this task is left to the sampling algorithm and is therefore implementation specific. Solutions for the constrained sampling include ellipsoidal rejection sampling (Feroz et al. 2009), MCMC slice sampling (Handley et al. 2015b), Metropolis-Hastings MCMC sampling (Elsheikh et al. 2014; Brewer & Foreman-Mackey 2016) and genetic algorithms (Qian & Zheng 2017).

Figure 3.1: A pictorial representation of how the nested sampling method may turn a 2-D parameter space in (a) into the 1D likelihood curve given by equation (3.12) shown in (b). Reproduced from Feroz et al. (2013)
3.2. Nested sampling

3.2.2 Nested sampling evidence errors

We are interested in parameter estimation and model selection applications of the nested sampling algorithm. The errors on the evidence calculation have been well studied since inception (Skilling 2006), and recently an in-depth analysis of parameter estimation analysis has been presented (Higson et al. 2017). Assuming perfect nested sampling, such that there is no implementation bias, the source of evidence calculation errors is given by the prior mass shrinkage approximation. The prior masses are unknown during sampling as the iso-likelihood contour is generally not known. Prior masses are modelled statistically as a succession of shrinkages \( t_i \) at each step \( i \), such that \( X_i = t_i X_{i-1} \).

For a single live point run, the initial live point samples uniformly from the whole prior. On average it will cut the prior mass into half: \( \langle X_1 \rangle = \frac{1}{2} \). In generating the next point, we sample uniformly from the remaining prior volume, with \( t \sim \mathcal{U}[0, 1] \), and on average the result will be \( \langle X_2 \rangle = \frac{1}{4} \). Generally we can write the prior mass as a product of their shrinkage factors: \( X_i = \prod_{j=0}^{i} t_j \), where the shrinkage factors are each uniformly distributed. As iteration step \( i \) gets large, the central limit theorem can be used to show that the prior mass becomes log-normal distributed (Handley 2016):

\[
\log X_i \approx -i \pm \sqrt{i}.
\]  

(3.14)

The error on the prior mass propagates through to the evidence via the nested sampling weights in equation (3.13). Therefore nested sampling computes a distribution on the evidence (typically the log-evidence) with error \( \approx \sqrt{i} \).

When we have \( N_{\text{live}} \) live points exploring the space simultaneously, the shrinkage of an iteration is from the lowest likelihood point to the next lowest likelihood point. Therefore the next lowest likelihood point has the largest prior mass out of the set of \( N_{\text{live}} \) prior masses defined by the \( N_{\text{live}} \) points. As the prior volume of each individual point is sampled uniformly from within an iso-likelihood contour, the shrinkage of a given step is that of the largest value of \( N_{\text{live}} \) uniformly distributed probabilities. Therefore we can define the probability of the \( i^{th} \) shrinkage as well as the expectation and uncertainty:

\[
\Pr(t_i) = N_{\text{live}} t_i^{N_{\text{live}}} , \quad \langle \log t_i \rangle = -\frac{1}{N_{\text{live}}}, \quad \sigma(\log t_i) = \frac{1}{n}.
\]  

(3.15)

Again these errors propagate through to the prior masses as

\[
\log X_i \approx -\frac{i}{N_{\text{live}}} \pm \sqrt{\frac{i}{N_{\text{live}}}}.
\]  

(3.16)

The number of iterations a nested sampling run is expected to take can be defined by the log-compression from prior to posterior, where most of the final evidence contribution is from
Chapter 3. Statistical Framework

a region $\mathcal{X} \sim e^{-H}$ (Feroz & Hobson 2008):

$$H = \int \log \left( \frac{d \Pr(\theta|D)}{d\mathcal{X}} \right) d\mathcal{X} = \sum_i \frac{L_i w_i}{\mathcal{Z}} \log \left( \frac{L_i}{\mathcal{Z}} \right), \quad (3.17)$$

where $H$ is the relative entropy. As prior mass scales exponentially into this log-region it will take $i/N_{\text{live}} \approx H$ iterations to explore the parameter space. Substituting the number of iterations into equation (3.17) and noting that the evidence is just a linear combination of the prior mass functions, we obtain the result that the error on the evidence scales as $\sqrt{H N_{\text{live}}}$. We immediately notice that a greater number of live points takes a greater number of iterations (more computational cost) and produces a better estimate of the evidence. Another important point is that the error on the evidence calculation is due to the product of each individual shrinkage factor estimation error. Alternative estimates for the errors on nested sampling evidences can be found in Keeton (2011); Feroz et al. (2013), but throughout this thesis the above discussion informs our understanding of the nested sampling evidence calculation.

3.2.3 Nested sampling parameter estimation errors

Posterior estimation is different in two ways. Firstly, the combination of weights and likelihoods takes a different form when constructing posterior samples:

$$p_i(t) = \frac{L_i w_i(t)}{\sum_j L_j w_j(t)} = \frac{L_i w_i(t)}{\mathcal{Z}(t)}, \quad (3.18)$$

where the posterior samples $p_i(t)$ are the importance weights for the parameter points defined by the set of samples in the run. The importance weights are used to compute the expectation of a function of the parameters via the expression

$$\langle f(\theta) \rangle_t = \int f(\theta) \Pr(\theta) d\theta \approx \sum_i p_i(t) f(\theta_i). \quad (3.19)$$

The propagation of the shrinkage factor errors is explicitly shown as a dependence on the vector of all shrinkage factors $t$. Depending on the function to sample, the errors can be larger or smaller than the errors on the evidence. Higson et al. (2017) found that they are typically smaller as this source of error only depends on the relative weights of the samples, rather than in the evidence calculation where the error of each shrinkage factor was implicitly included.

However, the second difference between the evidence and parameter estimation error calculations is a source of error only present in parameter estimation. From figure 3.1 we observe that each contour in likelihood parameter space is approximated with a single point on the $\mathcal{L}(\mathcal{X})$ space. Higson et al. (2017) show that this approximates the mean of the function across the whole iso-likelihood contour to the value $f(\Theta)$ at the point sampled. Specifically, the function $f(\Theta)$ will vary as $\Theta$ varies on the iso-likelihood contour defined by a prior mass. The
iso-likelihood contour is typically not known, nor is it averaged over in the algorithm, and this introduces errors that depend on the intricate shape of each iso-likelihood contour.

No approximation is available for the error due to compressing the iso-likelihood contour, but Higson et al. (2017) provide a general method of computing errors for any given function of the parameters from the set of explored points. As nested sampling runs can be combined to form new valid nested sampling runs of greater total live point (Skilling 2006), it is possible to revert a given nested sampling run with $N_{\text{live}}$ live points back into a set of $N_{\text{live}}$ single live point nested sampling runs. Each of the single live point runs has a set of posterior samples, weights, likelihoods and an evidence. This single live point run is termed a thread.

With a set of single live point runs, the threads, it is possible to randomly recombine a selection of them to generate valid new nested sampling runs that are distinct from each other and the original run of all $N_{\text{live}}$ live points. Each new run will contain a subset of the total information of the original run. Analysing a set of such runs can identify the robustness of parameter estimation calculations via standard mean and standard deviation calculations on the set of results produced. The method is shown to be robust, as expected given the derivation from standard nested sampling results.

To summarise, nested sampling presents a method of calculating evidences and parameters from the likelihood space. A single nested sampling run provides errors on evidences as well as errors on functions of the posterior samples. The errors due to the shrinkage factor approximation in prior mass generally differs between the evidence and parameter estimation results. An additional source of error exists in parameter estimation due to the iso-likelihood compression into a single point whereby information is lost on the iso-likelihood contour which would be required for the estimate of a parameter. To understand nested sampling further requires understanding its computational implementation.

### 3.2.4 PolyChord and MultiNest

The nested sampling algorithm PolyChord (Handley et al. 2015a,b) is primarily used in this thesis. There are several implementation specific factors that will be mentioned in later chapters, such as additional parameters beyond $N_{\text{live}}$ which influence the sampling (and not the physical model), analysis of robustness and perceived convergence, and potential issues with exploration when things are not working well. A brief understanding of the sampling algorithm used will therefore be useful. Additionally we briefly discuss MultiNest (Feroz & Hobson 2008; Feroz et al. 2009; Feroz & Skilling 2013), a nested sampling algorithm with good low-dimensionality performance, to highlight implementation choices and also as it will be mentioned on occasion in later chapters. The most important difference between MultiNest and PolyChord is the sampling procedure.
**MultiNest** creates an ellipsoidal approximation to the iso-likelihood contour of the lowest likelihood point at a given iteration and samples uniformly from within the ellipsoid until it finds a replacement point of higher likelihood. This is termed rejection sampling as candidate samples are rejected if they do not meet the minimum likelihood criterion. For an expensive likelihood calculation, the dominant computational cost at a given iteration is in the $N_{\text{reject}}$ number of points that are rejected before finding a new point, with the ‘acceptance rate’ defined as $1/N_{\text{reject}}$. Ellipsoids are construct over a set of live points to enclose them minimally (with a clustering algorithm defining what sets of live points to construct an ellipsoid for). As the true iso-likelihood contour will almost certainly include regions outside of this approximation, the ellipsoid is expanded by a factor of $f$ to ensure no region is cut off by the ellipsoid, which is determined by the MultiNest-specific runtime parameter $\text{eff} = 1/f$. As $\text{eff}$ is reduced the iso-likelihood is more likely to be contained by the ellipsoid and the evidence calculation becomes more robust, the algorithm also takes longer as the acceptance rate decreases.

**PolyChord** draws a new live point by consecutively uniformly sampling on lines of random orientation in the parameter space. This is termed slice sampling (Neal 2000) and is an MCMC approach. At the lowest likelihood live point, a line (or chord) is chosen with random orientation in the multi-dimensional space. Points are sampled along this chord to find boundaries on either side where the likelihood is below the iso-likelihood contour. Once boundaries are known, a new point is chosen using rejection sampling on the 1D chord. From this new point, another chord is drawn and the procedure repeated, with each new slice-sampled point walking the MCMC chain through the iso-likelihood volume. The samples become decorrelated and the number of repeated steps to take is given by the user defined variable $N_{\text{rep}}$. The dominant computational cost for an expensive likelihood is the $N_{\text{rep}}$ repeated chords to find a single new live point, with each chord needing to find boundary points (likelihood calculations are required to test points on the chord) and then selecting the new MCMC point with rejection sampling on the boundary. The larger the number of chords, the more uniformly and thoroughly the parameter space is explored: the evidence calculation becomes more robust and the algorithm takes longer.

Intuitively, MultiNest and rejection sampling seems less computationally intensive as we can picture sampling from a 2D contour, such as described in figure 3.1, without discarding and wasting many points in the rejection step. Sampling the lightest green contour in figure 3.1 from within an ellipsoid expanded out as far as the yellow curve might only take 2 or 3 attempts before we find a valid new live point. With PolyChord one needs to first determine the boundaries along the chord, which might take 3 or 4 trials along each direction, then sample the new point, and also repeat this a few more times to decorrelate the MCMC chain. This expected difference in efficiencies between PolyChord and MultiNest is certainly realised
for low dimensionalities, and the rejection sampling method of \textsc{MultiNest} is best for low dimensionalities (Handley et al. 2015b). 

In higher dimensionalities the volume on the shell of the contour becomes an increasingly large ratio of the total volume. The expansion $f$ in the \textsc{MultiNest} ellipsoids must essentially be applied to each dimension to avoid the iso-likelihood contour in any dimension falling outside the ellipsoidal approximation. Hence the volume to sample within expands as $f^{N_{\text{dim}}}$, where $N_{\text{dim}}$ is the dimensionality of the parameter space. For higher dimensions it roughly follows that rejection sampling scales exponentially with dimension. MCMC techniques on the other hand can scale better with dimension. \textsc{PolyChord} draws 1D chords in the space, such that adding a dimension requires you to draw more samples to decorrelate the extra correlation. The decorrelation time, or number of steps, is proportional to the number of degrees of freedom, which is the dimensionality. Therefore $N_{\text{rep}} \propto N_{\text{dim}}$, and the scaling for decorrelating points with higher dimensions is not exponential. Handley et al. (2015b) describe scaling for \textsc{MultiNest} and \textsc{PolyChord} in greater detail and describe how \textsc{PolyChord} scaling is found to be $\sim O(N_{\text{dim}}^{3})$. They note that complex parameter spaces may have different scaling and that generally a rejection sampler is the most efficient choice for simple or low-dimensional problems. In this thesis \textsc{PolyChord} is the preferred nested sampler due to the predicted benefits for the complex and higher dimensional spaces that this thesis studies\footnote{\textsc{PolyChord} also has ongoing development, and therefore very useful tech-support. \textsc{MultiNest} is now well tested but the developer has moved on from it.}.

It cannot be known \textit{a priori} what values of search parameters (such as $N_{\text{live}}$, $N_{\text{rep}}$ or $\text{eff}$) are needed to explore the parameter space sufficiently well. One can test the robustness of evidence calculations and parameter estimations by repeating the algorithm. This is highly recommended, even if undesirable due to computational costs. Throughout this thesis there will be repetitions of the nested sampling algorithm to ensure accuracy. When likelihoods are not prohibitively expensive, the preferred method adopted is to conduct repetitions at the same runtime settings and observe whether the quoted errors are equivalent to the statistical standard deviation on the distribution of repetitions (as well as checking means). When likelihoods are expensive, the preferred method adopted is to start with runtime parameters below the expected optimum values and increase these to their optimum over a limited handful of runs. The latter allows one to form conclusions on the convergence of the runs with respect to sampling parameters. Intuitively, this convergence measure seems to avoid having to repeat runs of equivalent sampling parameters a statistically significant number of times, but of course this is not the case: any results based on a low number of samples is subject to risk from random variation. As computational time is limited on high performance computing servers, we unfortunately cannot escape this reality and instead ensure to transparently reflect our tests
of robustness throughout.

### 3.3 Implementation in cosmology

The discussions on the Bayesian framework and the nested sampling algorithm provide the background for the computational work in this thesis. We now wish to comment briefly on how these will be implemented in the specific cosmological and gravitational wave studies of the coming chapters. The thesis uses well established cosmological codes, likelihood routines written by collaborators and also specific toy models. Here we will describe the cosmological codes. Codes written by collaborators will be commented on as they arise and toy models will be described in detail as they arise.

For the cosmological codes we primarily rely on CAMB (Lewis et al. 2000; Howlett et al. 2012) to produce cosmological power spectra, likelihood codes provided by experimental collaborations, CosmoMC (Lewis & Bridle 2002) to wrap up likelihoods and CAMB with an MCMC sampler for parameter estimation, and the CosmoChord (Handley et al. 2015a,b) code to replace the MCMC sampler in CosmoMC. CosmoChord is the PolyChord implementation designed specifically for use in CosmoMC. Additionally we utilise the PPF extension of CAMB (Fang et al. 2008).

#### 3.3.1 Cosmological parameterisation

The description of the ΛCDM model at the end of section 2.1.1 defined 6 parameters: 3 for the evolution of the scale factor, 2 for the primordial anisotropies and 1 for the reionisation history. These define the model intuitively but a better set of parameters exists for computational work. Better in this context specifically means that the likelihood space produced by the parameters is more Gaussian, primarily the computational parameterisation aims to minimise degeneracies. This is important for MCMC codes which are widely used (both when these issues were first resolved and now that one might be able to move on). Although nested sampling deals well with such degeneracies, we keep the same parameterisation for ease of comparison to the community.

The CosmoMC parameterisation in the code uses the following six parameters. $\Omega_b h^2$ is the physical baryonic matter density and $\Omega_c h^2$ is the physical dark matter density. $100\theta$ is $100 \times$ the ratio of the sound horizon to the angular distance, at the surface of last scattering and can be used to derive $H_0$. $\tau$ is the optical depth at reionisation from which the reionisation epoch median redshift can be derived. $n_s$ and $\ln(10^{10}A_s)$ define the primordial power spectrum. These are the parameters used in the CosmoMC convention and other parameters can be derived in turn from another significant reason being so as to not have to rewrite established codes.
them. Generally the convention of base parameters to use does not affect the final results, but we note that a flat prior choice on all parameters in one convention does not necessarily imply a flat prior choice in a different convention. In cases where the likelihood strongly constrains the data, this prior discrepancy between conventions will not matter.

A likelihood code associated with a dataset may introduce nuisance parameters which are added to the cosmological base parameters in order to model data or instrumentation specific issues. As noted near the end of section 2.1.4, the Planck data typically has on order 15 of these whilst the JLA data introduces 2. A >20 dimensional parameter space is not trivial, fortunately the MCMC implementation can utilise a fast-slow parameter separation technique whereby the nuisance parameters can be changed and a new likelihood calculated without an entire recalculation of the \textsc{camb} power spectra. This allows the MCMC chains to take many steps in the computationally relatively fast nuisance parameter directions to easily decorrelate samples. The effect of fast-slow sampling is to greatly speed up the likelihood calculations for the nuisance parameters whilst simultaneously sampling over them very thoroughly so that they do not pose a difficult challenge for the MCMC program. A similar approach has been implemented in the \textsc{CosmoMC/OND} nested sampling code (Handley et al. 2015b).

Alongside the base $\Lambda$CDM and nuisance parameters are any parameters which are included as part of a model extension. Typically the dark energy work in this thesis will introduce up to 10 additional parameters. These parameters are slow in the sense that they require a recalculation of the full likelihood code to obtain a new likelihood.

### 3.3.2 CosmoMC and \textsc{camb}

\textsc{camb} produces power spectra for the CMB $C_l$s and for the matter power spectrum $P(k)$ described in section 2.1.3.3. A good summary of workings are found in the notes of Lewis (2014). The power spectra are computed by approximately solving the transfer functions to high precision (Lewis et al. 2000; Howlett et al. 2012), and it is therefore referred to as a Boltzmann code (after the Boltzmann equations, see Dodelson (2003) for details). The details of the approximations and computational implementation of \textsc{camb} are discussed further in Lewis et al. (2000), and the citations therein. Howlett et al. (2012) describe how the CMB power spectra are reconstructed with an accuracy sufficient for high precision cosmology using Planck and similar datasets. For sensible precision settings, the \textsc{camb} code can approximate the spectra for given cosmological parameters within a matter of seconds.

A dark energy extension module for \textsc{camb} is used to facilitate general quintessence models. The module uses a so-called parameterized post-Friedmann (PPF) approach adopted from modified gravity theories to solve the problem that single dark energy scalar fields cannot easily be made to cross the $w = -1$ line in their time evolution (Fang et al. 2008). The PPF module
is incorporated easily into \texttt{CAMB} by design and defining the equation of state is trivial. In chapters 4 and 5 we modify the PPF module to parameterise a dark energy equation of state function for testing time varying dark energy behaviour. In chapter 6 we again modify the PPF module to introduce a second dark energy component, this essentially amounts to duplicating the \texttt{CAMB} $\Omega_A$ perturbation growth code for a second similar dark energy component, and again is not difficult to do given the modular coding used in \texttt{CAMB} for the dark energy effects on power spectra.

Bayesian inference is carried out by the \texttt{CosmoMC} code package (Lewis & Bridle 2002). It facilitates MCMC exploration of cosmological model posteriors by combining likelihood codes, the \texttt{CAMB} code and an MCMC implementation tuned for the $\Lambda$CDM model. The various datasets described near the end of section 2.1.4 provide likelihood codes already designed for inclusion in \texttt{CosmoMC}, with many already present in the out-of-the-box version of \texttt{CosmoMC}. The priors need to be defined for the model parameters, and sensible defaults are already presented in the default configurations. Throughout this thesis we state our priors for model selection and parameter estimation problems as they arise. We note too that \texttt{CosmoMC} places implicit prior constraints on the user defined priors such that no parts of the parameter space are permitted which ‘break’ the calculation of the \texttt{CAMB} code. There are further top hat priors on $H_0$ which may restrict some extreme prior choices too. Generally these prior considerations affect neither parameter estimation, as those regions not permitted are far from posterior peaks, nor model selection, as they are applied across all models and may only have limiting Occam factor effects given that the regions not permitted are far from the posterior peaks.

The MCMC module in \texttt{CosmoMC} can easily be replaced by the \texttt{CosmoChord} nested sampler, which is the \texttt{CosmoMC}-ready implementation of the \texttt{PolyChord} algorithm discussed in section 3.2.4. This enables evidence calculations using nested sampling and also removes the otherwise necessary MCMC tuning for specific models. \texttt{CosmoChord} works for our proposed dark energy extensions without the requirement to tune any search specific algorithm parameters, aside from ensuring that $N_{\text{live}}$ and $N_{\text{rep}}$ are sufficient for the dimensionality. Additionally the \texttt{CosmoChord} algorithm implements the same fast-slow mechanism which allowed \texttt{CosmoMC}’s MCMC implementation to explore cheaply and thoroughly the nuisance parameters of certain data likelihoods.

The combination of \texttt{CAMB} and likelihood codes, via \texttt{CosmoMC}, with the \texttt{PolyChord} sampler, via \texttt{CosmoChord}, provides a powerful Bayesian inference package. The rest of the thesis will make extensive use of these codes in chapters 4, 5 and 6, with modifications to them described as arising. Chapter 7 will make extensive use primarily of the \texttt{PolyChord}
package and a likelihood provided by a collaborator which models Kerr waveform production as realistically measurable by ground based gravitational wave detectors. We have now concluded the statistical overview and are ready to apply the techniques throughout the remaining work presented in this thesis.
This chapter introduces a new method to compute posterior odds ratios (and the more commonly used Bayes factors). Posterior odds ratios are fundamental to Bayesian model selection as they characterise the posterior probability ratio between two models: a direct quantification of model preference. The method that this chapter develops obtains posterior odds ratios without calculating evidences, but instead uses parameter estimation. The evidence calculations is typically the most difficult part of Bayesian inference, whereas parameter estimation for models is comparatively simple.

The method’s validity is demonstrated on a toy model and its utility is demonstrated by application to a phenomenological dark energy equation of state investigation. Evidences could not be calculated for the dark energy problem using the MultiNest nested sampling algorithm, as the dimensionality and parameter space complexity was too large. As parameter estimation was robust, however, the new method excelled. This method requires no simplifying assumptions or restrictions on the types of models that can be compared, making it widely applicable.

This method is used again in chapter 7 where applications to a toy model and a gravitational wave physical model highlight efficiency gains. The dark energy investigation presented here is expanded on in chapter 5 which also introduces a fuller treatment of the dataset information content. This chapter is concerned with validation of the method and an investigation of the dark energy equation of state using Planck 2013 data; it is adapted from the publication by Hee et al. (2015).
4.1 Introduction

Comparing two or more models given some data is central to the scientific method. The field of model selection within statistical inference attempts to address this problem, and numerous techniques for choosing between models exist, including: Akaike’s Information Criterion (Akaike 1974), Schwarz’s Bayesian Information Criterion (Schwarz 1978) and the Bayesian evidence (Jeffreys 1961; MacKay 2003). Throughout the thesis the focus is on Bayesian model selection using the evidence \( Z \) (also known as the prior predictive or marginal likelihood) and posterior odds ratios \( P_{ij} \) (a generalisation of the more commonly used Bayes factors \( B_{ij} \)), as this technique is inherent to Bayes theorem and both are widely used throughout cosmology and astrophysics (Liddle et al. 2006).

Posterior odds ratios provide a quantitative means for selecting between models and are usually calculated directly from the evidence of each model. A thorough discussion of techniques to calculate evidences and posterior odds ratios is presented in section 3.1.2, where the method of nested sampling is conclusively favoured for complex and higher dimensional likelihoods. The executive summary is that calculating evidences is a difficult task: many MCMC methods struggle to do this well for general parameter spaces and nested sampling solves this but inevitably still requires significant computational expense.

In this chapter, a method is proposed to calculate posterior odds ratios without the problems associated with evidence calculations or simplifying assumptions. In this new method, posterior odds ratios are calculated directly from a set of models explored simultaneously without constraints on the forms these models might take. The new method circumvents the challenges associated with accurate evidence calculations by computing posterior odds ratios using Bayesian parameter estimation, which is typically a more reliable and computationally less expensive task. Additionally, parameter estimation algorithms are more commonly used and therefore the method provides an easy means for extending existing knowledge to the domain of model selection. This evidence-free Bayes factor calculation is achieved by introducing a parameter that selects between models, and allows the calculation of posterior odds ratios from the posterior probability of this new model selection parameter. Let us note that similar approaches have been proposed previously (Hobson & McLachlan 2003; Goyder & Lasenby 2004; Brewer & Donovan 2015), but these typically rely on the use of sampling techniques capable of jumping between parameter spaces of different sizes, such as reversible jump MCMC (Green 1995), which require special sampling methods that are often very computationally demanding. The new approach is much simpler, requiring no special sampling methods, provided the number of models under consideration is specified \textit{a priori}. The new method is related to the class of product-space MCMC methods originally proposed by Carlin & Chib (1995) (see also Sisson.
This chapter sees the application of the new method applied to toy models and the cosmological problem of constraining the dark energy equation of state. The equation of state of dark energy is discussed near the end of section 2.1.1, and throughout section 2.2.2, and in this chapter there will be a particular emphasis on determining the complexity of redshift-evolution supported by data and determining potential deviations from \( \Lambda \)CDM. Both the toy and cosmological applications are solving the problem of how many nodes are required in a piecewise linear model to reconstruct a one-dimensional function. With the number of nodes defining the models, one can show explicitly that this new method agrees with the evidences-based approach for calculating posterior odds ratios on models.

The rest of the chapter is organised as follows. Section 4.2 defines the new method and discusses the statistical framework for calculating posterior odds ratios using parameter estimation techniques. Thereafter, results are presented in section 4.3 for a toy model data fitting problem and in section 4.4 for the cosmological problem of characterising the dark energy equation of state parameter as a function of redshift. The toy model application serves to validate the method on a fast likelihood, facilitating extensive testing, whilst the cosmological application highlights how to practically and robustly apply the new method to an expensive likelihood. The cosmological model analysis was completed between 2014 and 2015, and utilised suitably modern cosmological datasets at that time. Additionally, the analysis was carried out before Higson et al. (2017) presented a method for obtaining errors on parameters within nested sampling and before Handley et al. (2015a) had presented PolyChord; best practical procedure for applying the new method will therefore have changed slightly, essentially making application of the method easier and more robust. These changes will be noted when applicable. Findings and conclusions are summarised in section 4.5.

4.2 Method

The new method described in this chapter can calculate posterior odds ratios using parameter estimation techniques, such that the method avoids calculating evidences directly. The method places no constraints on the models that can be considered and has the advantage of being simple to implement and undisruptive for members of the community familiar with Bayesian parameter estimation techniques, especially those using nested sampling already. The method can be implemented in any nested sampling algorithm that can deal with rapidly varying parameter spaces. For the sampling methods discussed in section 3.2.1 and section 3.2.4 we observe that ellipsoidal sampling does not struggle with such spaces (as an ellipsoid can span over the rapidly varying range and sample uniformly on either side of any such cliff feature).
and nor does MCMC slice sampling (as the boundary points allow a uniform sampling across a range with a rapidly varying region).

Let us consider a number of different models \( M_n \) \((n = 1, 2, \ldots, N)\), and combine these into a single hyper-model \( M \). The parameters of \( M \) are the integer variable \( n \) that ‘switches’ between the models \( M_n \), and the union \( \theta \) of the parameter vectors \( \theta_n \) of each individual model. Note that, if there is some overlap between the parameter vectors \( \theta_n \) and \( \theta_{n'} \) of two different models, then the coincident parameters are notionally included only once in the union \( \theta \). In practice, the parameter \( n \) can be implemented as a continuous parameter and a suitable binning used to convert it to an effective integer parameter, thereby simplifying the implementation (provided the technique used to explore the parameter space does not rely on gradient information).

Indeed, the implementation of the new method is, in general, straightforward, since one needs only to write a simple ‘wrapper’ hyper-likelihood function for \( M \), which calls the existing likelihood function for the appropriate individual model \( M_n \) depending on the (integer) value of \( n \).

In general, the parameter vectors \( \theta_n \) and \( \theta_{n'} \) for different models will be of different dimensionalities. In the case of nested models, where \( \theta_n \subset \theta_{n+1} \), such problems are usually accommodated using reversible-jump Markov chain Monte Carlo (RJMCMC) methods, which are capable of making transitions between spaces of different dimensionality. In principle, such methods might also be used in the case of non-nested models, even in the extreme case where \( \theta_n \) and \( \theta_{n'} \) have no parameters in common, although such applications appear not to have been widely explored.

Here let us adopt a different approach that accommodates nested and non-nested models equally well, including the extreme case mentioned above, and avoids the algorithmic complication and computational expense of RJMCMC methods. The only assumption required is that \( N \) (the number of models under consideration) is known \emph{a priori}. Although this seems an innocuous requirement, it does constitute a mild limitation. Consider, for example, the classic nested problem of fitting a polynomial of unknown degree to a set of \((x, y)\) data points. In the new approach, one is required to fix the maximum allowed degree \( N \) of the polynomial in advance, whereas this is not necessary in the traditional RJMCMC approach. Nonetheless, in realistic applications such a limitation is not too severe.

By fixing \( N \), the full parameter space \((\theta, n)\) is determined \emph{a priori}, and is of fixed dimensionality, so it may be explored using standard sampling methods, such as MCMC or nested sampling (MacKay 2003; Skilling 2006; Brewer et al. 2011). Explicitly, suppose at some MCMC step or nested sampling iteration one considers the point \((\theta, n)\) (possibly after suitable binning of the continuous parameter \( n \) to obtain an integer value). For any given value of \( n \) so obtained, the union parameter space may be partitioned into those parameters \( \theta_n \) on which the
model $M_n$ depends and the remaining parameters $\theta_n$ that are not used by $M_n$. The ‘wrapper’ hyper-likelihood function thus may pass only the parameters $\theta_n$ to the likelihood function for the appropriate model $M_n$. The remaining parameters $\phi_n$ are thus ‘ignored’, which is equivalent to assigning a constant likelihood value over this subspace. By considering the full space $(\theta, n)$, however, the sampling method will typically need to accommodate moderate to large dimensionality, most likely possessing multiple modes and/or pronounced degeneracies. In practice, nested sampling is well suited to such problems, and therefore we can adopt it here.

Once one has obtained a set of posterior samples from the space $(\theta, n)$, one may calculate $Pr(n|\mathcal{D}, M)$ by simply marginalising out all other parameters to produce a marginalised posterior probability:

$$ Pr(n|\mathcal{D}, M) = \int Pr(\theta, n|\mathcal{D}, M) d\theta $$

$$ = \frac{1}{Z_M} \int L(\theta, n) \pi(\theta, n) d\theta, $$

where $Z_M$ is the evidence for this hyper-model $M$. Since for any given value of $n$ the union parameter space may be partitioned into those parameters $\theta_n$ on which the model $M_n$ depends and the remaining parameters $\phi_n$ that are not used by $M_n$, one may write the likelihood in (4.2) as $L(\theta_n)$ and the priors as $\pi(\theta|n)=\pi(\theta_n|n)\pi(\phi_n)\pi(n)$, where $\pi(n) \equiv Pr(n|M)$. Hence (4.2) becomes

$$ Pr(n|\mathcal{D}, M) = \frac{\pi(n)}{Z_M} \int L(\theta_n) \pi(\theta_n|n) d\theta_n, $$

where we have used the fact that the integral over the priors for unused parameters is unity, namely $\int d\phi_n \pi(\phi_n) = 1$. We can recognise the integral in (4.3) as the evidence $Z_n$ of the model $M_n$, so that we have

$$ \pi(n)Z_n = Z_M Pr(n|\mathcal{D}, M). $$

Of interest for model selection are the posterior odds ratios between two models, $M_i$ and $M_j$:

$$ \mathcal{P}_{ij} = \ln \left[ \frac{Pr(n=j|\mathcal{D}, M)}{Pr(n=i|\mathcal{D}, M)} \right], $$

where the $Z_M$ cancels. Thus, the posterior odds ratio is given simply by the ratio of values of the posterior $Pr(n|\mathcal{D}, M)$ for the two models, which is obtained using the parameter estimation formulation of Bayes theorem and the process of marginalisation, without the need to calculate evidences directly. The key feature is that the unused parameters $\phi_n$ marginalise out to unity. Moreover, the posteriors on $\phi_n$ should simply equal the priors on $\phi_n$. Visual inspection of these posteriors thus provides a useful check that the method is performing correctly.

A potential downside to this method is the requirement that the prior probabilities of the models are specified in advance. For signal detection problems with an unknown number
Figure 4.1: Illustration of the nodal reconstruction, which flexibly allows the parameter estimation process to define the preferred shape of \( y(x) \) from the data by linearly interpolating nodes whose amplitudes, positions (for internal nodes) and number can vary as required. The figure shows the interpolation process, and highlights how nodes can be positioned inside the unshaded prior space (with sorting of node positions such that \( x_i < x_{i+1} \)).

of sources, for example Hobson & McLachlan (2003); Feroz & Skilling (2013), this is in principle undesirable but in practice a suitable prior choice can always be found. Additionally, if calculating posterior odds ratios for another model \( M_{N+1} \) was desired, after having completed the analysis for the first \( N \) models, then a repetition of the method with only this new model and the most favourable model is possible, at a computational cost of exploring the most favourable model\(^a\) a second time.

It is also important to note, however, that at the time of these investigations the new method could not produce an estimate of the error on the posterior odds ratios in a single computation, whereas this is possible when calculating evidences directly using nested sampling. As a result, multiple repeat runs are used throughout this chapter to obtain errors on posterior odds ratios (and Bayes factors).

### 4.3 Application to toy-models

This section demonstrates the approach by applying it to some toy-models. The next section applies the method to constraining the dark energy equation-of-state as a function of redshift using recent cosmological datasets.

In both applications we will model a one-dimensional function \( y(x) \) using a piecewise linear interpolation scheme between a set of nodes and ask the model selection question “how many nodes are needed to fit the data?” Thus a set of nodes \( y_i(x_i) \) is placed in the plane, where the amplitude \( y_i \) and the position \( x_i \) are model parameters to be varied. At \( x_{\text{min}} \) and \( x_{\text{max}} \) fixed-position nodes are placed with varying amplitude only, such that for the model defined by

\(^a\)The most favourable is best used, in light of discussions on the size of error bars in section 4.3.3.
4.3. Application to toy-models

\[ y(x) = \sin(2\pi x) \]

\[ y(x) = \text{line}(2\pi x) \]

Figure 4.2: Data points plotted in the \((x, y)\) plane for each dataset (a) and (b). The unshaded region represents the prior space for the \(y_i\) amplitudes and \(x_i\) positions of the nodes, over which a uniform prior is assumed (with sorting of the node position parameters such that \(x_i < x_{i+1}\)).

For internal nodes there are \(2 + 2n\) parameters. As shown in Figure 4.1, linear interpolation is used to construct \(y\) at all points (with \(y(x)\) set constant outside the range \([x_{\text{min}}, x_{\text{max}}]\)). Of course, other interpolation schemes between nodes may be used, such as splines, although these are not considered here. The application of spline interpolation to constraining \(w(z)\) is described by Vázquez et al. (2012c).

A specific model is defined by how many nodes are used in reconstructing \(y(x)\). Comparing multiple models with increasing numbers of nodes identifies how many nodes are needed to fit the data, in other words the preferred complexity inherent in the data. As the final result, one can plot either \(\Pr(y|x, n)\), where \(n\) denoted the number of nodes in the most favoured model, or \(\Pr(y|x)\) averaged over all models weighted by their posterior odds ratios (PORs) (Parkinson & Liddle 2013; Planck Collaboration et al. 2016f). Either approach identifies clearly the nature of the data constraints on \(y(x)\).

The key strength of the reconstruction is its free-form nature, which can capture any shape of function in the \((x, y)\) plane by adding arbitrarily large numbers of nodes. Providing the model selection criterion penalises over-complex models appropriately by weighing ‘goodness-of-fit’ against the numbers of parameters in the model (Occam’s razor), identifying how much complexity the data support is performed in a clear and unambiguous manner by the favoured number of nodes. Model selection techniques can thus be used to solve questions on the constraining power of the data, as successfully shown in various cosmological applications (Vázquez et al. 2012b; Vázquez et al. 2012c; Planck Collaboration et al. 2016f).

The nodal reconstructions are clearly nested models. Since the general approach of the new method does not require this, for completeness we will also review a non-nested model selection problem by comparing a 2-internal node reconstruction with a sinusoidal model. The
Chapter 4. Development of the H3L method

Figure 4.3: Posterior odds ratios (or Bayes factors) for datasets (a) and (b) defined by Figure 4.2. $B_{n,n}$ denotes the Bayes factor for the models with $n$ and $n^*$ internal nodes. Histograms represent posterior odds ratios with respect to the most probable model. White, light grey and dark grey bars are for the vanilla, Post$(n)_{25}$ and Post$(n)_{50}$ results respectively. Error bars shown are sample standard deviations obtained from 10 repeat trials. The posterior odds ratios agree well between methods.

Figure 4.4: Average timing data for datasets defined in Figure 4.2 and the vanilla, Post$(n)_{25}$ and Post$(n)_{50}$ results defined in the text. The shaded regions show the approximate number of likelihood calculations made for each model $n$ and the solid lines show the cumulative numbers. More detail and an analysis of the timing benefits of using the new method are given in section 4.3.4. Considering error bars on the posterior odds ratios for the different methods, it is clear that the Post$(n)_{50}$ method (darkest plots) can produce comparable accuracy in less likelihood calls than the vanilla method (lightest plot).

rest of this section presents the results obtained and highlights further strengths and weaknesses of the new method.
4.3. Application to toy-models

4.3.1 Fitting a function to data

Consider a set of $j_{\text{max}}$ data points $\{(x_j, y_j), j=1, \cdots, j_{\text{max}}\}$ with experimental errors $\{(\sigma_{x_j}, \sigma_{y_j})\}$ on each of the points. Assuming there is a functional relationship between the independent variable $x$ and dependent variable $y$, captured by $y=f(x)$, then the likelihood of observing these data is given by:

$$
\Pr(x_j, y_j|\sigma_{x_j}, \sigma_{y_j}, f, X_-, X_+) = \prod_{j=1}^{j_{\text{max}}} \int_{X_-}^{X_+} dx_j \exp \left[ -\frac{(x_j-X_j)^2}{2\sigma_{x_j}^2} - \frac{(y_j-f(X_j))^2}{2\sigma_{y_j}^2} \right],
$$

(4.6)

where $X_-, X_+$ are the end points of the uniform region in which the data points may be found a priori. A Bayesian derivation of this likelihood can be found in section 4.3.2; for more detail see Sivia & Skilling (2006). The integral is calculated numerically using standard quadrature techniques.

Given the data, the Bayesian approach is to use this likelihood to infer the probability distribution of the parameters in some parametric form of the function $f$. We can do this for the family of functions described above, and use posterior odds ratios to determine how many nodes optimally reconstruct the function.

In this toy model section, 2 different datasets are tested, shown in Figure 4.2. The traditional evidence-based approach and the new method for calculating posterior odds ratios are compared for each dataset. The constraints on $y(x)$ given the data are also discussed.

Dataset (a) has 47 datapoints drawn uniformly in $x$ from the function $y=\sin(2\pi x)$ in the range $x \in [0, 1]$, with each point adjusted in $x$ and $y$ by random Gaussian noise with mean=0 and $\sigma=0.05$ (error bars on datapoints are $\sigma$)\textsuperscript{b}. Dataset (b) has 49 datapoints drawn as in (a) but from a piecewise-linear function coinciding with the function $y=\sin(2\pi x)$ at $x=0$, 0.25, 0.75, 1, so that it is very difficult by eye to distinguish the two datasets as being drawn from different functions. Let us call the function used in (b) line$(2\pi x)$ for brevity. Clearly, a linearly interpolated nodal model with $n=2$ internal nodes can represent this function exactly.

For each of the datasets models with 1 internal node up to 7 internal nodes are tested (i.e. 3 total nodes up to 9 total nodes or 2 line segments up to 8 line segments), using PolyChord (Handley et al. 2015a,b) to calculate evidences (the vanilla method henceforth) and again using PolyChord to implement the new method (Post($n$) method henceforth)\textsuperscript{c}. PolyChord is a relatively new nested sampler and was found to be very suitable for this problem. Uniform

---

\textsuperscript{b}50 points were drawn initially for each dataset, but some fell outside the prior range, due to the Gaussian noise, and were not included.

\textsuperscript{c}Note the marginalised posterior probability on $n$ is calculated from the chain\_unnormalised.txt file using the standard nested sampling technique (Skilling 2006). It is important to use this file over the usual chain.txt file and set up PolyChord to output all inter-chain points of the algorithm. This ensures good reconstruction of $\Pr(n|D, M)$ over the lower probability regions in light of the computing ‘log-sum-exp’ problem.
Chapter 4. Development of the H3L method

priors are used on the $y$ amplitudes of nodes, and sorted uniform priors on the $x$ position parameters of nodes, where the $x$ priors are uniform but forced to adhere to $x_i < x_{i+1}$ to avoid the scenario where the $n$ internal nodes are interchangeable with each other. Equal prior probabilities are assigned for each model, so that PORs are equal to Bayes factors.

Each dataset is analysed 10 times for each method to determine the statistical uncertainty on the derived PORs. In each case the PORs are normalised to the model with the highest evidence in the vanilla method. Errors on the posterior odds ratios are given as the sample standard deviation from the 10 repeats.

\texttt{P/o.sc/l.sc/y.scC/h.sc/o.sc/r.sc/d.sc} was run with $N_{\text{live}}=25N_{\text{dim}}$ live points initially to obtain the results labelled Post\(_n^{\text{1n}_n^25}\), where $N_{\text{dim}}=2n+2$ is the number of parameters to be explored (the dimension of the space) and the number of live points, $N_{\text{live}}$, is the only tuning parameter associated with the \texttt{P/o.sc/l.sc/y.scC/h.sc/o.sc/r.sc/d.sc} sampling algorithm. To highlight accuracy and timing considerations when using the method, the analysis is repeated with $N_{\text{live}}=50N_{\text{dim}}$ to obtain the results labelled Post\(_n^{\text{1n}_n^250}\).

4.3.2 Line fitting Likelihood

We aim to fit a parametric function $y = f(x)$ to a set of $j_{\text{max}}$ data points \(\{x_j, y_j\}\), where we have some knowledge of the errors on these measurements \(\{\sigma_{x_j}, \sigma_{y_j}\} (\{j = 1, \cdots, j_{\text{max}}\})\). In order to fit the function, one needs to calculate the likelihood of observing the data \(\{x_j, y_j\}\), given the function $f$, the observed errors and any additional assumptions made $I$:

$$\Pr(\{x_j, y_j\}|{\sigma_{x_j}, \sigma_{y_j}, f, I}).$$

(4.7)

To model the “error bars”, we can assume that each of the data points \((x_j, y_j)\) is drawn from a separable Gaussian distribution with covariance \(\text{diag}(\sigma_{x_j}^2, \sigma_{y_j}^2)\). The distribution will be centered about some true value \((X_j, Y_j)\), where these values are unknown and will need to be marginalised over as nuisance parameters in the final calculation. If each of these distributions are independent from each other, we arrive at the likelihood:

$$\Pr(\{x_j, y_j\}|{X_j, Y_j, \sigma_{x_j}, \sigma_{y_j}}) = \prod_{j=1}^{j_{\text{max}}} \frac{1}{2\pi\sigma_{x_j}\sigma_{y_j}} \exp \left[ -\frac{(x_j - X_j)^2}{2\sigma_{x_j}^2} - \frac{(y_j - Y_j)^2}{2\sigma_{y_j}^2} \right].$$

(4.8)

To marginalise out the nuisance parameters, we can place our prior assumptions on them. Let us assume that the true $X_j$ values are drawn uniformly in some range $X_- < X_j < X_+$, and that the true $Y_j$ obey the functional relationship: $Y_j = f(X_j)$. Given this, the probability distribution is:

$$\Pr(\{X_j, Y_j\}|f, X_-, X_+) = \begin{cases} \frac{1}{X_+-X_-} \prod_{j=1}^{j_{\text{max}}} \delta \left[ Y_j - f(X_j) \right] & : X_- < X_j < X_+ \\ 0 & : \text{otherwise} \end{cases}$$

(4.9)
where $\delta$ is the Dirac $\delta$-function. Multiplying (4.8) and (4.9) together and marginalising out $\{X_j, Y_j\}$ by integrating yields the likelihood:

$$
Pr(\{x_j, y_j\}|(\sigma_{x_j}, \sigma_{y_j}), f, X_-, X_+) = \prod_{j=1}^{j_{\text{max}}} \int_{X_-}^{X_+} dx_j \frac{\exp \left[ -\frac{(x_j - x_i)^2}{2\sigma_{x_j}^2} - \frac{(y_j - f(x_i))^2}{2\sigma_{y_j}^2} \right]}{2\pi \sigma_{x_j} \sigma_{y_j} (X_+ - X_-)}
$$

(4.10)

This procedure may be straightforwardly extended to consider correlated error bars where the covariance matrix of (4.8) is no longer diagonal. One may also adjust (4.9) if some additional knowledge is known about the independent variables $X_j$. For further details the reader is referred to Sivia & Skilling (2006).

### 4.3.3 Results for nested nodal models

The posterior odds ratios (or Bayes factors) for the vanilla method with $N_{\text{live}}=25N_{\text{dim}}$ and Post($n$) method with $N_{\text{live}}=25N_{\text{dim}}$ and $50N_{\text{dim}}$, per dataset, are shown in Figure 4.3 and show good agreement between the two methods regardless of $N_{\text{live}}$. From this we can conclude that the methods produce consistent posterior odds ratios. As one might expect, for the line($2\pi x$) dataset, the preferred model has $n=2$ internal nodes, whereas a larger number of nodes is preferred for the sin($2\pi x$) dataset. The reconstructions of the favoured models for each of the methods are shown in Figures 4.5 and 4.6. The reconstructions are identical in all key features between methods. The Post($n$)$_{50}$ graph is not plotted as it was very similar. Finally, the timing data in Figure 4.4 suggests that Post($n$)$_{25}$ results were faster to obtain by about a factor of 2.5 when using the same $N_{\text{live}}$ per parameter, however this comes at a cost in accuracy as the errors on the vanilla posterior odds ratios are clearly tighter than the Post($n$)$_{25}$ results. Post($n$)$_{50}$, however, takes less time to produce similar accuracy for the significant posterior odds ratios. In general we observe that the new method can produce Bayes factors faster than the vanilla method in a systematic manner. Timing considerations and discussed further in section 4.3.4.

The important discrepancies between the vanilla and Post($n$) methods are in the errors on the posterior odds ratios, where one can identified 2 issues: firstly for large negative posterior odds ratios the errors from the Post($n$) method are quite large and, secondly, the errors on the vanilla method are tighter for equivalent $N_{\text{live}}$. The first discrepancy might be expected given that POLYCHORD, and nested samplers in general, rapidly converge to the central peak(s) in a distribution, thus spending less time in lower likelihood regions and sampling those regions proportionately less thoroughly. Given that each model investigated is a separate mode in the computation, a model with low likelihood will be less thoroughly explored than the models with larger likelihoods – making the calculation of $Pr(n|\mathcal{D}, \mathcal{M})$ less reliable for these models. This is, however, desirable behaviour. Spending compute time only on probable models reduces the
Figure 4.5: Reconstructions of $y(x)$ using the vanilla method of explicitly calculating evidences to obtain posterior odds ratios. Plots are from one of the 10 trials, arbitrarily chosen, and are of the model with the largest posterior odds ratio, i.e. (a) 6 internal node model, (b) 2 internal node model. Each figure shows the posterior probability $\Pr(y|x, D, M)$, in normalised slices of constant $x$ to show the deviation from the peak $y$ at each $x$, binned in 100 bins in both $x$ and $y$. The colour bars to the right show the credible region intervals that the probabilities represent at a given slice in $x$ as calculated from the inverse of the cumulative distribution function on $\Pr(y|x, D, M)$, see Planck Collaboration et al. (2016f) section 8.2 equation 68 for details. The $1\sigma$ and $2\sigma$ intervals are plotted as black lines for clarity and the cube-helix colour scheme by Green (2011) is used for linearity in grey scale. In white is plotted the underlying function from which the data was sampled, and even with less than 50 datapoints a good reconstruction is obtained.

Figure 4.6: Reconstructions of $y(x)$ for the Post($n$)$_{25}$ results to obtain posterior odds ratios. Plots are for comparison to the vanilla results of Figure 4.5, and are plotted in the same way. The Post($n$)$_{25}$ results agree well with the vanilla method results in all key features.
overall time taken to find the most probable model(s), whilst the less probable models are still sampled sufficiently well to identify them as less probable.

The second discrepancy is more significant but equally predictable. The number of live points in POLYCHORD defines how fully the space is explored. For the vanilla method, the $N_{\text{live}} = 25 N_{\text{dim}}$ calculation provides adequate sampling per model, whilst for the Post$(n)$ method a similar number of live points needs to explore several models simultaneously, effectively reducing the live points available to explore each model and producing larger errors. This suggests that users need to ensure that algorithm tuning parameters such as $N_{\text{live}}$ are chosen appropriately and check that the results on repetitions of the algorithm are consistent. The Post$(n)_{50}$ results demonstrate clearly that results are confidently extracted in comparable compute-times when best practice is adhered to. Being aware of the increased modality of the space that is inherent to the method and ensuring that the sampling algorithm adequately handles such complex parameter spaces helps ensure accurate results.

Finally, it is worth making some brief comments on the ‘physical’ results of the model selection process for each of the datasets. In dataset (a) a more complex underlying shape in $y(x)$ is identified needing more nodes than dataset (b), consistent with the distinction between $\sin(2\pi x)$ and line($2\pi x$). It should be noted too that over-fitting (adding more parameters than needed) is not heavily penalised for dataset (b), as discussed in section 3.1.3, as observed in the slow decrease in Bayes factors after the favoured model is found – this is standard behaviour (Sivia & Skilling 2006, p. 93) and can be understood by considering the Occam factor associated with a parameter which is constrained without increasing the fit of the model (MacKay 2003, p. 349). In general the model selection and nodal reconstruction technique produces strong conclusions on the shapes of the $y(x)$ plane, given the data in each case, and clearly identifies the inherent complexity of the various datasets, as is desired.

### 4.3.4 Efficient computing of Bayes factors

Using the datapoints in figure 4.7 to test the vanilla and Post$(n)$ methods demonstrates that the new method may outperform the evidences approach in a systematic fashion that makes the approach desirable for common astrophysical and cosmological problems.

Running the nodal reconstruction technique with models of 1 internal node up to 13 internal nodes (3 to 15 total nodes) produces the Bayes factors and timing results shown in figure 4.8. The timing data shows the number of posterior points, and thus likelihood calculations up to a factor of the PolyChord efficiency, that each method makes for each of the nodal reconstruction models (shaded plots), alongside the cumulative number of likelihood calculations of these models (line plots).
Using the vanilla method, completing the evidence calculation for each model means that adding increasingly complex models is increasingly computationally expensive. In the Post(n) method, however, the model space is rapidly traversed from lower likelihood regions to higher likelihood regions, so that computationally expensive models with low likelihoods (or more correctly, with lower Bayes factors compared to other models in the space) are explored rapidly by the nested sampling algorithm. This is clearly identified by the fact that the Bayes factors and the number of likelihood calculations peak at the same model (4 internal nodes) and tail off similarly for models on either side of this.

It is worth noting, however, that the Post(n) method may perform more likelihood calculations for the most probable models, because the additional overhead of setting up the other parameters and populating their dimensions with live points (because $N_{\text{live}} \propto N_{\text{dim}}$ was used) means that the algorithm progresses more slowly.

Astrophysical and cosmological problems where a number of models of increasing complexity are explored may therefore benefit from using this method. The gravitational wave analysis in chapter 7 showcases this boost in efficiency very clearly. It is not generally guaranteed, however, as with the vanilla case one may have identified a drop off in the Bayes factors beyond $n = 8$ and stopped testing the more complex models thereafter. Nonetheless, the Post(n) method could provide an efficient means of verifying the drop off. For example one might run the above model analysis with $\pi(n)=[4, 9, 10, 11, 12, 13]$ as a fast means of verifying the shape, knowing that the Bayes factors for more complex models will be low and therefore the computational cost of obtaining the Bayes factor will be low. Any gains in performance must be considered against the need for repetition of the algorithm to obtain an estimate of the error on the Bayes factors. The nested sampling parameter estimation error analysis by Higson et al. (2017) is strongly recommended to deal with this issue, as it provides a robust nested sampling based technique for obtaining errors on any posterior function from a single nested sampling run. We note that this analysis could not include it as it was carried out before the results of Higson et al. (2017) were known, but chapter 7 presents a thorough analysis using this parameter estimation error analysis, using firstly a new toy model and then a gravitational wave model.

### 4.3.5 Results for non-nested models

The new method does not require that the models be nested. A model is nested inside another ‘larger’ model if setting some parameters to specific values in the larger model allows one to obtain the smaller nested model. The nodal reconstructions are clearly nested in this sense.
4.3. Application to toy-models

Figure 4.7: A set of 11 datapoints defining a spike in the $x$-$y$ plane. This dataset is tested with models of 1 internal node up to 13 internal nodes (3 to 15 total nodes).

Figure 4.8: Bayes factors with respect to the most probable model (top) and timing data (bottom) for the vanilla method and the Post($n$) method using $25N_{\text{dim}}$ and $50N_{\text{dim}}$ number of live points. Note that the large error bars on the dataset in figure 4.7 allow models that underfit with less than 3 internal nodes (1 at each vertex of the spike signal) to be probable. The timing data is measured by the number of likelihood calculations the algorithm makes. The shaded regions show the time taken on each nodal-reconstruction model for the vanilla (lightest colour plotted), Post($n$)$_{25}$, and Post($n$)$_{50}$ (darkest colour plotted) methods. Observe that the shapes of the Post($n$) method timing data coincides with those of the Bayes factors, as explained in the text, and thus outperforms the vanilla method in obtaining Bayes factors accurately.

A quick demonstration that the new method also works for non-nested models is sensible for completeness.

Datasets (a) and (b) are tested against two models. The first model is the sinusoid function $y(x) = A \sin(2\pi Bx + C) + D$ and the second model is the 2 internal node reconstruction, so that one expects dataset (a) to favour the sinusoidal model and (b) to favour the linear model. Parameters $A$ and $B$ are scale parameters for the amplitude and frequency respectively; these are assigned logarithmic priors in the range $[0, 5]$. Parameters $C$ and $D$ are shift parameters and uniform priors are assigned in the ranges $[-\pi, \pi]$ and $[-1.5, 1.5]$ respectively. These priors
reflect sufficient coverage of the prior space defined in Figure 4.2 and are adequate for comparing the vanilla and new methods. It is important to note that in this test, both the vanilla method and Post\( (n) \) method used \( N_{\text{live}} = 25 N_{\text{dim}} \). For the vanilla method this resulted in \( N_{\text{live}} = 100 \) for the sinusoidal model and \( N_{\text{live}} = 150 \) for the 4 node model, whilst for the Post\( (n) \) method the parameters were searched simultaneously (along with \( n \)) to give 11 parameters and \( N_{\text{live}} = 275 \).

The posterior odds ratios for dataset (a) favour the sinusoid by \( 1.94 \pm 0.93 \) and \( 2.01 \pm 1.08 \) units, for vanilla and Post\( (n) \) methods respectively. The posterior odds ratios for dataset (b) favour the linear model by \( 13.82 \pm 1.02 \) and \( 14.87 \pm 2.58 \) units, respectively for vanilla and Post\( (n) \) methods. Taking into account the previous discussion, it is clear that the new method produces posterior odds ratios consistent with the vanilla method. The Post\( (n) \) method here was about 5 per cent slower for dataset (a) and 30 per cent slower for dataset (b). However, with the significantly larger number of live points that the Post\( (n) \) method used, the fact that the methods are of comparable time is a desirable result and suggests that the unconstrained parameters for a given \( n \) are not significantly increasing the compute time of those isolated nodes in the parameter space.

In general we can conclude that the discussions in section 4.2 regarding unconstrained parameters is correct. When parameters were reviewed for the chains files produced in a given model, the parameters that were not used by that model were distributed according to their priors. This is one of the core strengths and novelties of the method and allows posterior odds ratios to be calculated without constraints on the models to be compared. This verifies that the method works for non-nested models, and we can now proceed to apply it to a cosmological application using the nodal reconstruction.

### 4.4 Applications to the dark energy equation of state

Having validated the approach on a toy problem, let us now apply the new method to a cosmological application, for which the vanilla method is not computationally suited. The aim is to demonstrate the method in a typical model selection application to obtain posterior odds ratios efficiently and with estimates of the error that do not require excessive repetition of long computations. This investigation probes the dark energy (DE) equation of state parameter \( w(z) \) as a function of redshift to update the work of Vázquez et al. (2012c), using more modern datasets. It further showcases the usefulness of the nodal reconstruction approach, briefly described in section 4.3 and more fully in Vázquez et al. (2012c), in defining the complexity supported by the data and identifying features in \( w(z) \), adding to the body of work using the reconstruction (Vázquez et al. 2012b; Vázquez et al. 2012c; Aslanyan et al. 2014; Planck Collaboration et al. 2016f).
4.4.1 Method

CMB data from the Planck 2013 data release (Planck Collaboration et al. 2014b,c,d) is combined with the WMAP 9-year polarisation data (Hinshaw et al. 2013), baryonic acoustic oscillation (BAO) from the BOSS data release 11 (Anderson et al. 2014) and supernovae type Ia (SNIa) data from the Union 2.1 catalogue (Suzuki et al. 2012) to provide constraints on DE behaviour. The investigation focusses on the redshift range \( z \in [0, 2] \) in the reconstruction, with constant values \( w(z) = w(2) \) set when \( z > 2 \). The CosmoMC code package (Lewis & Bridle 2002) is used throughout, which contains the camb code (Lewis et al. 2000; Howlett et al. 2012), and the MCMC sampler is substituted for the MultiNest nested sampling plugin running in constant efficiency mode (Feroz & Hobson 2008; Feroz et al. 2009, 2013), which is a well established nested sampling implementation for evidence calculations and parameter estimation, and was the sampler used by Vázquez et al. (2012b); Vázquez et al. (2012c) thereby enabling a direct comparison. To facilitate deviations away from the standard \( \Lambda \)CDM equation of state parameter \( w = -1 \) one can implement the ‘Parameterized Post-Friedmann’ framework (PPF) modification to camb (Fang et al. 2008). For further details on the method and datasets see section 3.3 as well as the papers by Vázquez et al. (2012c) for the method and paper Planck Collaboration et al. (2014c) for a discussion of the datasets.

Using posterior odds ratios to identify the optimal number of nodes tells us the complexity of \( w(z) \) features supported by the data. Further, the nodal reconstruction, as shown in the toy model, is highly adept at identifying constraints in the \((w, z)\)-plane. Of particular interest is whether deviations in \( w(z) \) away from the successful \( \Lambda \)CDM cosmological model are supported by modern data and to identify which DE extensions are favoured. Theories incorporating deviations from \( w = -1 \) include quintessence scalar fields for \( w > -1 \) (Ratra & Peebles 1988; Caldwell et al. 1998; Tsujikawa 2013) and phantom DE models with super-negative \( w < -1 \) (Caldwell 2002; Sahni 2005). The possibility of crossing of the phantom divide line at \( w = -1 \) in dynamical models has also been considered (Zhang 2009). Modified gravity or brane-world models also make predictions about \( w(z) \) (Sahni 2005). Thus, paramount to understanding DE is determining \( w(z) \).

To do this we can compare 6 models, in order of increasing complexity: \( \Lambda \)CDM with \( w = -1 \), \( wCDM \) with \( w \) constant in \( z \) but allowed to vary in amplitude, \( tiltCDM \) with \( w(z=0) \) and \( w(z=2) \) allowed to vary and linear interpolation for \( w(z) \) between them (0 internal node model), and then nodal models with 1, 2 and 3 internal nodes respectively. Models are abbreviated to \( \Lambda \), \( w \), \( t \), 1, 2 and 3 respectively, where appropriate. Priors on each \( w \) parameter are uniform on the range \([-2, 0]\) and were chosen to be conservative. Checking the robustness of results with respect to prior choice is useful, see Vázquez et al. (2012c) for such an analysis, but was not
done here. Priors on each $z$ parameter are uniform on $[0, 2]$ such that for more than one internal node $z_i < z_{i+1}$ (i.e. sorted uniform priors as in the toy model). The previous work by Vázquez et al. (2012c) found that $\Lambda$CDM was favoured, whilst the 2 internal node model had the second largest evidence, pointing to structure in $w(z)$ that could not be captured by a constant equation of state parameter $w_{\text{CDM}}$, or even the 1 internal node model. The results in this section show clearly that Planck 2013 era datasets do not have this feature and only $\Lambda$CDM can be considered favoured.

An important point is that the Planck data require the addition of 14 so called nuisance parameters. These must be sampled and, together with the 6 parameters of CDM models, produce an at least 20 dimensional parameter space. As MultiNest is a rejection nested sampling algorithm, it is expected that computation times increase significantly in higher dimensions as the volume on the shell increases\(^d\). MultiNest has the algorithm search parameters $N_{\text{live}}$ and $\text{eff}$, where decreasing $\text{eff}$ (in constant efficiency mode) typically achieves more accurate results more effectively than increasing $N_{\text{live}}$.

With the new method, a technique for obtaining errors on posterior odds ratios from a single run is presented in Higson et al. (2017). This method is highly recommended and requires minimal alteration of the nested sampling algorithm (to output the order of live point deletion). At the time of the current chapter’s investigation there was no way to estimate the errors on the posterior odds ratios from a single run, and attaining these from the chapter’s results is best done via repeat simulation (and the calculation of sample standard deviations from these). Therefore 3 repetitions were performed, each using $N_{\text{live}}=500$ with $\text{eff}=0.01$ (the repeat runs). The default July 2014 CosmoMC priors for the 20 CDM and nuisance parameters are used throughout with the priors mentioned above for additional model parameters; an overview is shown in table 4.1. The MultiNest constant efficiency mode had to be used to attain feasible computing times, similarly the search parameters could not just be increased arbitrarily. With these MultiNest search parameters and constant efficiency mode, it was found that the edges of the priors were not sampled effectively. The error is reproducible with a 20-dimensional Gaussian test likelihood with a covariance matrix given by Planck chains. To ensure this problem had no impact on the results, firstly a prior for an unconstrained parameter was added, the $\theta_{\text{uniform}}$ parameter in table 4.1, which should produce a flat posterior as it is not implemented in the likelihood. Observing the edge effects problem on this parameter gives a clear indication of the severity of the problem for other parameters, and allows us to reconsider parameter estimation conclusions if needed. It is a type of sanity check essentially. Secondly convergence

\(^d\)Specifically, it constructs multi-dimensional ellipsoids to estimate sampling within an iso-likelihood region, as required by nested sampling. The ellipsoids expand by a fraction to ensure no viable regions of the true iso-likelihood contour are outside this estimate. Points are sampled inside these ellipsoids and rejected until meeting the nested sampling criterion.
Table 4.1: The 30 priors that define the parameter space. The top set of parameters are the CDM parameters, the middle ones show the nuisance parameters associated with the Planck 2013 data release, and the bottom set are the parameters introduced by dark energy model extensions, including \( n \) for selecting between models and \( \theta_{\text{uniform}} \) for testing a MultiNest edge-effect problem. Planck Collaboration et al. (2014c) has more details about the CDM and nuisance parameters, whilst the dark energy extension parameters are defined in the text.

was tested for the marginalised posterior on \( n \) with respect to search parameter changes to ensure that the parameter estimation results were robust. A single further run was performed to check for convergence, using MultiNest with the search parameters \( N_{\text{live}}=1000, \text{eff}=0.005 \) (full run) for which the edges of the prior were sampled effectively. Given the concerns about the accuracy of the MultiNest evidence calculation for Planck data (due to nuisance parameters, high dimensionality, and the need for constant efficiency mode), the new method combined
Chapter 4. Development of the H3L method

Figure 4.9: The posterior odds ratios obtained from the new method comparing the 5 DE extension models to \( \Lambda \)CDM. The error bars on each histogram are the sample standard deviations of the 3 repeat runs. It is clear that the 2 sets of results agree very well, with discrepancies between them small compared both to the error bars and the absolute values used to draw conclusions based on Jeffreys guideline. This shows that the results are robust with respect to changes in \textsc{MultiNest} search parameters, as required. Numerical results are given in table 4.2.

<table>
<thead>
<tr>
<th>Bayes factor</th>
<th>Full run</th>
<th>Repeat averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B_{Aw} )</td>
<td>(-2.41 \pm 0.03)</td>
<td>(-2.55 \pm 0.03)</td>
</tr>
<tr>
<td>( B_{At} )</td>
<td>(-3.26 \pm 0.11)</td>
<td>(-3.43 \pm 0.11)</td>
</tr>
<tr>
<td>( B_{A1} )</td>
<td>(-3.54 \pm 0.32)</td>
<td>(-3.97 \pm 0.32)</td>
</tr>
<tr>
<td>( B_{A2} )</td>
<td>(-3.89 \pm 0.40)</td>
<td>(-4.50 \pm 0.40)</td>
</tr>
<tr>
<td>( B_{A3} )</td>
<td>(-4.31 \pm 0.63)</td>
<td>(-4.94 \pm 0.63)</td>
</tr>
</tbody>
</table>

Table 4.2: Summary of the Bayes factors from the 4 computations. The full run and repeat averages columns show results using the \textsc{MultiNest} search parameters discussed in the text. For both columns, the errors are sample standard deviations of the 3 repeat trials. The results agree well within 1\( \sigma \) credible region intervals for all but the \( B_{Aw} \), where a larger discrepancy occurs due to small error bars despite a small difference in log-units. The results show clearly that the new method implementation is robust to changes in \textsc{MultiNest} parameters.

with the 2 robustness checks thus provides a valuable alternative way to obtain posterior odds ratios.

### 4.4.2 Results

The posterior odds ratio results for the full run and the 3 repeat runs are shown in Figure 4.9 and table 4.2. The key points are firstly that the posterior odds ratios are consistent with each other, demonstrating convergence of \( \Pr(n|\mathcal{D}, \mathcal{M}) \) with respect to \textsc{MultiNest} search parameters, and secondly that the \( w(z) \) investigation clearly favours \( \Lambda \)CDM.

The toy model showed that error bars on posterior odds ratios will depend on how thoroughly
Figure 4.10: The $w(z)$ priors, $w(z)$ reconstructions and parameter constraints for each of the 5 model extensions beyond $\Lambda$CDM. The leftmost plot is the prior space on the function $w(z)$ as a result of the flat priors on amplitude and position parameters and the central plots show the posterior on $w(z)$ defining the data and model constraints on the $w(z)$-plane. These plots show the posterior probability $\Pr(w|z)$ similar to Figure 4.5. Here it is the probability of $w$ as normalised in each slice of constant $z$, with colour scale in credible region interval values shown. The 1σ and 2σ credible region intervals are plotted as black lines. Note that the prior on $w(z)$ in $w$CDM does not appear flat in this plotting style despite being so. Comparing the priors of the other 4 reconstructions to the flat $w$CDM prior it is noticed that the priors on $w(z)$ are slightly favouring the central values closer to $w=-1$ as expected when calculating priors analytically. The posterior shows that the data constrains $w(z)$ strongly compared to the priors. Rightmost are the 1D and 2D marginalised posteriors of the additional model parameters, with end-node amplitude posteriors shown as 1D plots and internal nodes shown as 2D plots. Although labels are too small to see, the plots span across their prior range defined in table 4.1, with a dashed line across $w=-1$. Plots were produced using GetDist and with the cubehelix colour scheme by Green (2011) for linearity in grey scale.
Figure 4.11: Summarising the DE model extension results for the constraints on the $w(z)$ plane. The 5 extension models, excluding $\Lambda$CDM, are weighted by their evidences to give a model averaged plane reconstruction (Parkinson & Liddle 2013; Planck Collaboration et al. 2016f), and plotted as in Figure 4.10. When including $\Lambda$CDM, approximately 85 per cent of the central credible region is contained in the line $w=-1$ due to the strength with which $\Lambda$CDM is favoured by the posterior odds ratios, almost 2$\sigma$. The two plots show the prior space (left) contracting down to the posterior odds ratio averaged $w(z)$ plane reconstruction (right), as discussed in the text. It is clear that $\Lambda$CDM is well within the favourable region, with the 1$\sigma$ contours easily containing $w=-1$.

the sampling explores the space. Note that the error bars used are the sample standard deviations from the posterior odds ratios of the 3 repeat runs. The repeat run posterior odds ratios are consistent with the full run and sufficiently tight to resolve differences to make conclusions based on Jeffreys guideline, suggesting that the space is well explored. This convergence on reruns, together with the convergence between different MultiNest search parameters, suggests that the posterior odds ratio results are robust. Additionally, the edge effect problem previously mentioned was thoroughly checked for using an unconstrained parameter $\theta_{\text{uniform}}$. The posterior of $\theta_{\text{uniform}}$ was close to flat for all runs. The edge effect problem presumably affects all parameters a small amount, as the strength of this effect is different between the different MultiNest search parameter settings whilst the posterior odds ratios are consistent, it suggests that the posterior odds ratios are not significantly biased. The 4 runs show that the posterior odds ratios are accurate, and we can quote the results as those of the full run combined with the errors from the 3 repeat runs as upper estimates for those of the full run (as repeats of a more well sampled run will produce tighter estimates, shown in the toy model analysis when doubling $N_{\text{live}}$).

From these posterior odds ratios it is clear that $\Lambda$CDM is the only favourable model. The decrease in posterior odds ratios with an increase in the number of parameters to model DE suggests that further additions of parameters to model deviations from $\Lambda$CDM are penalised
more strongly by the Occam’s razor principle than the gain in constraining power that they provide. One can estimate the Occam factor associated with adding an additional nodal amplitude parameter, using the analysis in (MacKay 2003, page 349), as $\sigma_{w|D}/\sigma_w$, where $\sigma_{w|D}$ is the width around the peak of a Laplace approximation inside the evidence integral and $\sigma_w$ is the prior width. As discussed in section 3.1.3, we can estimating $\sigma_{w|D}/\sigma_w$ for non-Gaussian parameters with a full width half max (FWHM) calculation of the 1D marginalised $w$-amplitude posterior. Doing this for the $w$CDM model’s additional parameter yields a drop in the Bayes factor due to the approximated Occam factor of $-2.63$. The observed $-2.41 \pm 0.03$ therefore suggests that the parameter is not improving the likelihood fit to the data significantly. Doing something similar for the 3 internal node model gives an Occam factor of $-0.45$ (using the average of the 5 amplitudes; assuming that an additional $z$-position parameter is unconstrained as there are no additional $w(z)$ features it would constrain). This is the anticipated decay in the posterior odds ratio when adding unnecessary nodes, and the Bayes factor drop from 2CDM to 3CDM at $-0.42$ suggests that 3 nodes already saturate the $w(z)$ space.

A clear and strong conclusion from this analysis is that there is considerably less evidence for deviations from $\Lambda$CDM in the Planck era datasets used here than in the WMAP era datasets used by Vázquez et al. (2012c), which is consistent with other results (Planck Collaboration et al. 2014c; Shafer & Huterer 2014). The next most favoured model is the next simplest one, $w$CDM, and at a posterior odds ratios of $-2.41 \pm 0.03$ it is almost significantly disfavoured according to the Jeffreys guideline. All other models are significantly disfavoured at between 3.3 to 4.3 log units.

The constraints in the ($w$, $z$)-plane for each of the model extensions beyond $\Lambda$CDM, shown in Figure 4.10, do however indicate some deviations from $w=-1$. Typically the data seem to favour the phantom region, potentially more so at the ends of the considered redshift range and less so at redshift 0.4–0.7, where the data gives the tightest constraints. However, the 1 $\sigma$ and 2 $\sigma$ contours clearly indicate that these effects are not significant. At all $z$ and for all models, $w=-1$ is comfortably within the peak of the $Pr(w|z)$ distribution and more so in the regions where we observe strong data constraints, suggesting that any deviations or apparent systematic patterns are dominated by a lack of data. The plane reconstructions also support the model selection conclusions that $\Lambda$CDM is significantly favoured over other models, as the constraints in the data do not deviate from $w=-1$ beyond even $1\sigma$.

The correct Bayesian way to view the $w(z)$ plane reconstructions for all models considered is to sum over all the models whilst weighting by the Bayesian evidence, or equivalently posterior odds ratios. This is exceptionally easy to implement with the new method, as a program like GetDist (included with CosmoMC) can use the chains file produced by the new method to correctly weight all the models automatically whilst marginalising out the parameter
Figure 4.11 shows this for the 5 DE extension models beyond $\Lambda$CDM. When plotting with $\Lambda$CDM the plot is centered on $w=-1$, with 85 per cent of the peak credible region contained in the $w=-1$ line, and thus a plot showing only the model extensions is more insightful. The plane reconstruction shows clearly the constraining power of the data at different redshifts as our knowledge of $w(z)$ moves from the prior on the left to the posterior on the right. The result is a tightly constrained function of $w(z)$ slightly below $-1$ for all redshifts, suggesting a small favouring of the phantom region at an insignificant level. Most importantly, $\Lambda$CDM is fully compatible, well within $1\sigma$ of the model extension results, as is expected given the Bayesian model selection analysis. This insignificant deviation away from $w=-1$ explains clearly why $\Lambda$CDM is so heavily favoured.

Of practical importance is the strength with which the nodal reconstruction identifies features, and especially that the reconstruction is data driven. Most of the datasets that can constrain $w(z)$ are in the redshift range $z \in [0.5, 0.8]$ and this is shown by where the reconstructions most tightly constrain the plane. This reconstruction technique is clearly of merit and in the future, with more powerful datasets, can hopefully act as a tool to identify features (if any) in $w(z)$. At present, the work here can only suggest that dark energy models with $w(z)$ close to $-1$ are needed. Finally, the posteriors of the CDM parameters are plotted in Figure 4.12 for each of the 6 models tested. The posteriors of the DE extensions agree well with the $\Lambda$CDM values, as can be expected given that there is no significant deviation from $w=-1$.

4.5 Conclusions

A novel method for calculating posterior odds ratios was validated using a toy model application and then applied to a cosmological model selection problem.

The new method uses Bayesian parameter estimation on a parameter that switches between models, via a hyper-likelihood that wraps around the individual model likelihoods, to infer posterior odds ratios (or Bayes factors if desired) without calculating evidences. It uses novel partitioning of the parameter space via the parameter $n$, and marginalisation of posterior probabilities, to allow sampling of a variable length parameter space when moving between models, thus facilitating any models to be tested without restriction and without reversible jump Monte Carlo techniques. To use the method one needs to have a parameter estimation algorithm capable of sampling from multi-modal spaces and to decide which models one wants to test a priori.

The toy model demonstrated clearly that the method is valid and consistent with the existing method of calculating posterior odds ratios by evaluating evidences. We can conclude that the new method is not necessarily faster, despite avoiding evidence integrals, for 2 reasons.
Firstly, to get errors on the posterior odds ratios it requires rerunning several times, whereas nested sampling algorithms such as MultiNest and PolyChord can attain error estimates of evidences from a single run. Secondly, the parameter space needs to be explored comparably thoroughly in both methods, as shown by the increase in error bars on the posterior odds ratios in the toy model when spending less computational time on the new method.

A peculiar feature of the new method in combination with nested sampling (which likely applies to other samplers too) is that computation time dedicated to a model is dependent on how strongly the model is favoured over others. Less favoured models become depopulated with live points as the nested sampling algorithm removes lowest likelihood points. As a result, we observed that less favoured models typically had less accurate posterior odds ratio calculations, which helps to reduce computing time, but still in such a way that they were always identifiable as less favoured. The reduction in computing time can be substantial, especially in applications where there are a number of computationally expensive models with low posterior odds ratios.

The toy models illuminated precautionary measures that best be adhered to by users. As with all Bayesian parameter estimation, robustness of posterior probabilities to changes in algorithm-specific tuning parameters needs to be tested for and in the case of the new method, where a posterior is used to infer evidence ratios, it is especially important to check this. It is best to test that the posterior odds ratios obtained from the posterior on \( n \) are consistent on repetitions of the algorithm and also that the error bars attained from repetitions are sufficiently small if needing to make judgments based on Jeffreys guideline. The toy model also highlighted the strength of the nodal reconstruction in identifying features in \( y(x) \) plane reconstruction problems. We can conclude that it is a useful tool for analysing the complexity supported by the data and add to the volume of literature using it (Vázquez et al. 2012b; Vázquez et al. 2012c; Aslanyan et al. 2014; Planck Collaboration et al. 2016f).

Thereafter, taking the above considerations into account, the new method was used to attain posterior odds ratios in a cosmological context where direct evaluation of evidences can be computationally demanding and problematic. We investigated an application of the nodal reconstruction technique to reconstruct the dark energy redshift-dependent equation of state parameter \( w(z) \), analysing the dynamic behaviour supported by modern datasets in a search for deviations from the \( \Lambda \)CDM model \( (w=-1) \). This work was principally an update on a paper using WMAP era data by Vázquez et al. (2012c). The conclusion was reached that \( \Lambda \)CDM is significantly favoured above any nodal reconstruction applied. Additionally, the model allowing \( w \) to vary as a constant is almost significantly disfavoured at \( -2.41 \pm 0.03 \) log-units of the posterior odds ratio with respect to \( \Lambda \)CDM. We can also conclude that additional parameters are systematically disfavoured: increasing the complexity of the \( w(z) \) reconstruction decreases posterior odds ratios with respect to \( \Lambda \)CDM. The Occam’s razor effect penalises additional
parameters when using posterior odds ratios to do model selection and, as $\Lambda$CDM is an excellent fit to current cosmological data, the addition of parameters to extended beyond $\Lambda$CDM adds less to the constraining power of the models than the Occam factor penalises.

The robustness of the results and methods were confirmed in several ways. Figure 4.12 shows that the CDM parameters of each of the dark energy extension models agree well with the $\Lambda$CDM values, as is expected given that all models agree well with $w = -1$. Further, a potential problem in sampling the edges of priors in high-dimensions was identified with MultiNest when using constant efficiency mode, but through tracking an unconstrained parameter $\theta_{\text{uniform}}$, it was shown to be insignificant given the final search parameters used. General robustness of the new method was confirmed too by repeating the calculation of $\Pr(n|D, M)$ with different search parameters and showing that the value of $\Pr(n|D, M)$ had converged with respect to algorithm tuning parameter.

Finally, the cosmological application demonstrated the strength of the new method, attaining posterior odds ratios without needing evidence calculations and effectively dealing with parameter spaces of varying length. Errors on the posterior odds ratios were attained through repeat runs with a faster sampling parameter setup which doubled to confirm that the posterior odds ratios were converged and accurate. As such a robustness check is important for any parameter estimation or model selection problem, where an algorithm uses tuning parameters for the sampling, this approach should come at little extra cost in practice. Even for the standard evidence calculation approach, testing of the runtime parameters requires repetition to ascertain robustness, and this is also the case for when calculating errors from a single run using the method of Higson et al. (2017).
As MultiNest converges to the peak likelihood regions, the datapoints output to the chains file are more sparse for some of the models. Typically $\Lambda$CDM had 8 times more points than $w$CDM with which to accurately reconstruct these posteriors. The lower posterior odds ratio models had less still and this leads to a lower quality reconstruction for the less favoured models. Nevertheless, it is clear that the models agree well and there are no significant deviations from the $\Lambda$CDM values of the CDM parameters, as can be expected given the only slight deviation from $w=-1$ in each model.
This chapter analyses in detail the phenomenological equation of state behaviour of dark energy in relation to the datasets used. We put on hold temporarily the new method to obtain posterior odds ratios developed in chapter 4. In that previous chapter we also saw the power of the nodal reconstruction in providing data-driven constraints on the dark energy equation of state across a redshift range. In this chapter, we firstly update the dark energy analysis with more recent data and secondly introduce a novel analysis of the constraining power of datasets.

We introduce the Kullback-Leibler divergence to quantify the information content in the transformation from a parameter’s prior to posterior. We extend the existing single parameter formalism to analyse a functional relationship. Specifically, we compute the Kullback-Leibler divergence on the prior and posterior of the equation of state parameter at a given redshift. Plotting this information content as a function of redshift identifies the constraining power of the data across the redshift range.

We use this novel tool to analyse the constraining power of a variety of different dataset combinations to conclude comprehensively on the effects of datasets on dark energy equation of state constraints. Such analysis could well be used widely in other areas of cosmology to identify gaps in data quality or plan future missions. The work in this chapter is adapted from the publication of Hee et al. (2017).
Chapter 5. Time dependence in dark energy equation of state

5.1 Introduction

The nature of dark energy (DE) remains a significant outstanding problem in cosmology. The $\Lambda$CDM model considers a constant equation of state (EoS) parameter $w = -1$ motivated by vacuum energy. The most frequent generalisation of the $\Lambda$CDM dark energy EoS is to allow an alteration of the time-independent EoS parameter so that $w \neq -1$ (hereafter referred to as $w_{\text{CDM}}$). Allowing $w$ to vary in time $w = w(z)$ probes quintessence DE models. Many quintessence models (Ratra & Peebles 1988; Caldwell et al. 1998; Tsujikawa 2013), including phantom DE (Caldwell 2002; Sahni 2005), as well as modified GR theories (Sahni 2005) make predictions for the behaviour of $w(z)$ which may be tested against cosmological datasets (Planck Collaboration et al. 2016d). A short review of these features and parameterisations is presented in section 2.2 of the introduction. Time-dependent behaviour can also be investigated by choosing equations that are simple or mathematically appealing, to test as a DE model. These phenomenological models include the CPL (Chevallier & Polarski 2001; Linder 2003), JPB (Jassal et al. 2004) and FNT (Felice et al. 2012) models. Lastly, free-form approaches attempt to avoid any commitment to particular equations and instead aim to allow the observational data to define any time-dependent features in $w(z)$ (Huterer & Starkman 2003; Zunckel & Trotta 2007; Zhao et al. 2008; Serra et al. 2009; Lazkoz et al. 2012; Vázquez et al. 2012c). Other free-form reconstruction methods include Gaussian processes (Holsclaw et al. 2010a,b; Seikel et al. 2012). We refer the reader to an older review by Sahni & Starobinsky (2006) which describes the general reconstruction process and new results by Planck Collaboration et al. (2016d) for further reading on dark energy constraints.

In this chapter we will see the use of Bayes factors combined with a ‘nodal’ free-form reconstruction method to investigate the constraints on $w(z)$. This nodal method reconstructs a function using a spline between nodes whose amplitudes and positions can vary, as first proposed by Vázquez et al. (2012b) and also used in chapter 4. This approach has also been used by Vázquez et al. (2012c); Aslanyan et al. (2014); Planck Collaboration et al. (2016f) and has the benefit of remaining general and allowing the cosmological datasets to define the posteriors without being restricted by a specific model.

The first aim of this chapter is to investigate potential deviations from the $\Lambda$CDM constant dark energy equation of state using Bayesian model selection and the most up to date datasets available at the time of writing this thesis (Planck 2015 data release and similar). This analysis will supersede the preliminary Planck 2013 results obtained in section 4.4, as the analysis will use more constraining datasets and a deeper analysis (as the previous chapter investigation was written to demonstrate the new method). The second aim is to analyse the constraining power of the datasets on $w(z)$ by using the Kullback-Leibler divergence (KL divergence; $D_{KL}$). In
doing so, a novel function of the $\mathcal{D}_{KL}$ is introduced which characterises the KL divergence of a posterior function $f(x)$ at a given $x$. This novel use of $\mathcal{D}_{KL}$ has subsequently been utilised by other authors due to its intuitive explanatory power (CORE Collaboration et al. 2016). Observational data are improving in quality with many upcoming missions promising to increase our ability to understand DE models. Assessing the datasets in the manner this chapter proposes provides a robust, quantitative measure of DE information that may easily be compared with past or future missions.

The chapter is structured as follows: We first review the datasets and computational techniques to be used in section 5.2. An analysis of $w(z)$ constraints from Planck satellite era cosmological datasets is presented in section 5.3 and the analysis of these additional datasets using the $\mathcal{D}_{KL}$ approach is presented in section 5.4. We will review conclusions in section 5.5, considering the findings in relation to $\Lambda$CDM and constraints on $w(z)$ and comment on the efficacy of the techniques used for quantifying dataset constraining power and information content.

### 5.2 Datasets and Computation

This chapter updates the work of Vázquez et al. (2012c) as well as the dark energy section of chapter 4, where time dependent behaviour in $w(z)$ within a CDM universe is identified using a sequence of nodal reconstructions weighted by their Bayes factors. In addition, this chapter introduces the Kullback-Leibler divergence to analyse information content, following similar work by Trotta et al. (2008); Bridges et al. (2009) as well as expanding on the formalism of the KL divergence to improve the detail of information it provides. Here we briefly summarise the methods and implementation.

#### 5.2.1 Datasets

In order to investigate possible time-dependent behaviour in the dark energy equation of state the investigation uses likelihood codes from Planck CMB measurements, baryonic acoustic oscillations (BAO), type-Ia supernovae (SNIa) and Lyman-α BAO data (Lyα). Each of these is discussed in more detail near the end of section 2.1.4 and references therein. For the CMB data, low-$l$ TEB and high-$l$ TT likelihoods from the Planck satellite 2015 data release (Planck Collaboration et al. 2016a,b,c) are used, which will be referred to as Planck. For the BAO data, the BOSS data release 11 likelihoods (Anderson et al. 2014) are used; BAO for short. For the SNIa data, the JLA supernovae catalogue likelihoods (Betoule et al. 2014) are used; JLA for short. For the Lyα data two datasets are used: firstly the likelihood code described by Font-Ribera et al. (2014) ($A_{Ly\alpha}$; BOSS cross-correlation) and secondly the likelihood codes
Table 5.1: The 31 priors that define the parameter space. The top set of parameters are the CDM parameters, the middle ones show the nuisance parameters associated with the datasets, and the bottom set are the parameters introduced by the free-form dark energy model extensions. Planck Collaboration et al. (2016b) has more details about the CDM and nuisance parameters, whilst the dark energy extension parameters are defined in the text.
### Table 5.2: Jeffreys guideline for interpreting posterior odds ratios.

As \( P_{ij} = -P_{ji} \), negative values imply model favouring is reversed.

<table>
<thead>
<tr>
<th>Posterior odds ratio</th>
<th>Favouring of ( M_j ) over ( M_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0.0 \leq P_{ij} \leq 1.0 )</td>
<td>None</td>
</tr>
<tr>
<td>( 1.0 \leq P_{ij} \leq 2.5 )</td>
<td>Slight</td>
</tr>
<tr>
<td>( 2.5 \leq P_{ij} \leq 5.0 )</td>
<td>Significant</td>
</tr>
<tr>
<td>( 5.0 \leq P_{ij} )</td>
<td>Decisive</td>
</tr>
</tbody>
</table>

### Figure 5.1: Piecewise linear interpolation function.

\( n \) internal nodes \((x_i, y_i)\) are placed in the rectangle bounded by \((x_{\text{min}}, y_{\text{min}})\) and \((x_{\text{max}}, y_{\text{max}})\), where the positions \( x_i \) and amplitudes \( y_i \) are model parameters to be varied. At \( x_{\text{min}} \) and \( x_{\text{max}} \) fixed-position nodes are placed with varying amplitude only, such that for the model defined by \( n \) internal nodes there are \( 2 + 2n \) parameters. Linear interpolation between the nodes \((x_i, y_i)\) is used to construct \( y(x) \) at all points, with \( y(x) \) set constant outside the range \([x_{\text{min}}, x_{\text{max}}]\).

from Delubac et al. (2015) (\( \text{B}_{\text{Ly}} \); BOSS auto-correlation). For a good summary of the BAO data see Aubourg et al. (2015). Using the above notation, the whole dataset combination can be referred to as \( \text{Planck} + \text{BAO} + J\Lambda + A_{\text{Ly}} + B_{\text{Ly}} \).

### 5.2.2 Computational tools

To carry out Bayesian inference, the CosmoMC (Lewis & Bridle 2002) code package is used, containing the Boltzmann CAMB code (Lewis et al. 2000; Howlett et al. 2012). These are both discussed in greater detail in section 3.3.2. Here we note that the default Metropolis-Hastings sampler of CosmoMC is substituted for the \textsc{PolyChord} nested sampling plug-in (Handley et al. 2015a,b), an effective nested sampling implementation (Sivia & Skilling 2006; Skilling 2004; Skilling 2006) for evidence calculations and parameter estimation with proven efficacy using Planck era data (Planck Collaboration et al. 2016f). This nested sampler is used throughout the thesis with good results. Aside from the \( \text{Ly}_\alpha \) datasets, all datasets used are default CosmoMC options. To facilitate deviations from the standard \( \Lambda \text{CDM} \) equation of state parameter \( w = -1 \) the
“Parameterized Post-Friedmann” framework (PPF) modification to CAMB (Fang et al. 2008) was implemented, which has sound speed equal to $c$ and no scalar anisotropic stress. The free-form reconstruction is the nodal reconstruction as proposed by Vázquez et al. (2012c) and successfully used in several cosmological applications to date (Vázquez et al. 2012b; Vázquez et al. 2012c; Aslanyan et al. 2014; Planck Collaboration et al. 2016f). Chapter 4 uses this reconstruction technique too, where its efficacy was demonstrated for both a toy model and a $w(z)$ reconstruction task similar to the one in this chapter.

5.2.3 Nodal reconstruction

Let us review the nodal reconstruction in slightly more detail than was done in chapter 4, as it forms the basis of the analysis of this chapter. We can model a one-dimensional function $y(x)$ using a piecewise linear interpolation between a set of $n$ nodes (Figure 5.1), where the positions of the nodes are model parameters to be varied. Alternative interpolation schemes may be used, for example, the cubic spline studied by Vázquez et al. (2012b), although these here not considered here since the continuity requirements of the interpolation functions and its derivatives limit its ability to model sharply changing functions $y(x)$.

A model is defined by how many nodes are used in reconstructing $y(x)$. We can use Bayes factors to compare models with increasing numbers of nodes, which quantify how many nodes are needed to fit the data. Further, as each posterior sample defines a function in $y(x)$, we can calculate the posterior probability of $y$ in normalised slices of constant $x$, Pr($y|x, D, M$), to obtain the plane reconstruction of a model. We can plot these as a function of $\sigma$ credible region intervals to show the statistical significance of deviations from the maximal $y$ at each $x$. One can plot Pr($y|x, n_*$), where $n_*$ denotes the number of nodes in the most favoured model. In order to identify the nature of constraints from various models, one should also plot Pr($y|x$) averaged over all models weighted by their posterior odds ratios (Parkinson & Liddle 2013; Planck Collaboration et al. 2016f; Hee et al. 2015).

A key strength of this reconstruction procedure is its free-form nature, which can capture any shape of function in the $y(x)$ plane by adding arbitrarily large numbers of nodes. The Bayes factor penalises over-complex models, identifying how much complexity the data is able to support. Model selection techniques can thus be used to solve questions on the constraining power of the data in cosmological applications (Vázquez et al. 2012b; Vázquez et al. 2012c; Aslanyan et al. 2014; Planck Collaboration et al. 2016f; Hee et al. 2015).

We apply this reconstruction to $w(z)$. The models we consider, along with their priors are detailed in table 5.3. The previous work using WMAP satellite era data by Vázquez et al. (2012c) found that $\Lambda$CDM was favoured, whilst 2CDM had the second largest evidence, pointing to structure in $w(z)$ that could not be captured by a constant equation of state parameter $w$CDM, or
### 5.2. Datasets and Computation

<table>
<thead>
<tr>
<th>Model name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda$CDM</td>
<td>$w = -1$</td>
</tr>
<tr>
<td>$\omega$CDM</td>
<td>$w$ constant in $z$, but allowed to vary</td>
</tr>
<tr>
<td>$\tau$CDM</td>
<td>tilted spectrum: two fixed-position nodes at $z = 0, 3$</td>
</tr>
<tr>
<td>1CDM</td>
<td>One internal node</td>
</tr>
<tr>
<td>2CDM</td>
<td>Two internal nodes</td>
</tr>
<tr>
<td>3CDM</td>
<td>Three internal nodes</td>
</tr>
</tbody>
</table>

Table 5.3: The six models we consider. Priors on each $w$ parameter are uniform on the range $[-2, 0]$, and were chosen to be conservative (Vázquez et al. 2012c). Priors on each $z$ parameter are uniform on $[0, 3]$ and sorted, such that for more than one internal node we have $z_i < z_{i+1}$ (i.e. sorted uniform priors).

even the 1 internal node model. The work with Planck 2013 era in chapter 4 showed that $\Lambda$CDM was again favoured, and that each model of increasing complexity was more disfavoured than the last. We now investigate this more fully with Planck 2015 era datasets, the addition of Ly$\alpha$ data and further dataset analysis tools. The dark energy equation of state results supersede the analysis in chapter 4 in dataset constraining power and depth of analysis.

#### 5.2.4 Kullback-Leibler divergence and dataset analysis

We expand on the model selection complexity analysis through the use of the Kullback-Leibler (KL) divergence. The KL divergence of $P$ from $Q$ is defined as

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \ln \left[ \frac{p(x)}{q(x)} \right] dx = \int \ln \left[ \frac{dP}{dQ} \right] dP,$$  \hspace{1cm} (5.1)

where $p(x)$ and $q(x)$ are the probability density functions of probability distributions $P$ and $Q$. Evaluating the KL divergence (5.1) of a posterior distribution from its prior provides a measure of information gained from the data (Kullback & Leibler 1951; Trotta et al. 2008; Bridges et al. 2009; Seehars et al. 2014, 2016; Grandis et al. 2016).

We wish to restrict our analysis to the constraining power of the datasets on $w(z)$, and not the other cosmological and nuisance parameters as a whole. First, we can calculate the KL divergence of the marginalised posterior $\Pr(w|z)$ from the marginalised prior $\pi(w|z)$ for $w$ at each $z$ to obtain a function:

$$D_{KL}(z) = \int \Pr(w|z) \ln \left[ \frac{\Pr(w|z)}{\pi(w|z)} \right] dw$$  \hspace{1cm} (5.2)

which defines the gain in information on $w$ at each $z$. Second, we calculate the $D_{KL}$ for the whole plane by using the function $\Pr(w, z)$ and its prior, which can be written as

$$D_{KL} = \int D_{KL}(z) \Pr(z)dz$$  \hspace{1cm} (5.3)
where $P(z)$ is flat (as $z$ is not constrained by the analysis given that every posterior sample for a nodal reconstruction passes through every point in $z$). Note that it is also possible to integrate over $da$ or $d \log(a)$ to compress higher redshifts, however $dz$ is more natural here given how we have defined our reconstruction. Together the two values allow us to analyse the gain in information due to different datasets using $D_{\text{KL}}$ as well as to understand where each dataset provides the greatest gains in information using $D_{\text{KL}}(z)$. We obtain the posterior plane reconstructions from POLYCHORD and the prior distributions based on $\pi(z_i)$ and $\pi(w_i)$ together with the physical restrictions imposed by CosmoMC.

Typically, a gain in information can occur for two reasons: either due to an increase in parameter constraints, or due to a shift in the position of the peak from prior to posterior (Trotta et al. 2008; Seehars et al. 2014, 2016; Grandis et al. 2016). It is not yet possible to differentiate between the two cases for non-Gaussian distributions. In order to identify the constraining power of the data, we supplement our analysis by calculating the $D_{\text{KL}}$ and $D_{\text{KL}}(z)$ when moving from a completely flat prior on $w(z)$ to the posterior. As there is no peak to shift from for a flat posterior, this measure only identifies how tightly constrained the posterior is, due both to the priors and data. In cases where the CosmoMC prior divergences are larger than those from the flat prior we can deduce that a significant shift is present.

### 5.3 Results: dark energy equation of state reconstruction

The columns in Figure 5.2 show from left to right the prior, posterior and marginalised 1D and 2D posteriors for the $w(z)$ plane reconstructions alongside the Bayes factors for the 5 model extensions compared to $\Lambda$CDM. $\Lambda$CDM is the favoured model in the Bayesian model selection analysis. $w$CDM is disfavoured by more than 2 log-units, a slight disfavouring on the Jeffreys scale, whilst all other models are significantly disfavoured at beyond 2.5 log-units. We conclude that the additional flexibility in capturing $w(z)$ features provided by additional parameters does not produce favourable Bayes factors. This is consistent with the previous results obtained with Planck 2013 data in section 4.4. The systematic dropping off in Bayes factors for models with increasing numbers of parameters used for defining $w(z)$ suggests that there is not sufficient time dependence in the true equation of state function to overcome the Occam factor associated with the additional parameters (MacKay 2003). Specifically, one can estimate the evidence integral using a Laplace approximation to obtain an Occam factor given by $\sigma_{\theta|D}/\sigma_{\theta}$ (MacKay 2003, page 349), where $\sigma_{\theta|D}$ is the width around the peak of the posterior and $\sigma_{\theta}$ is the same for the prior, and use this to determine the size of the Occam factor between models. See section 3.1.3 for details. When moving from 2CDM to 3CDM we obtain an Occam factor of approximately 0.72, where we assume the posterior on the additional nodal position parameter is equal to the
5.3. Results: dark energy equation of state reconstruction

\[ B_{Aw} = -2.34 \pm 0.29 \]

\[ B_{Al} = -2.68 \pm 0.29 \]

\[ B_{A1} = -2.68 \pm 0.29 \]

\[ B_{A2} = -2.82 \pm 0.29 \]

\[ B_{A3} = -3.36 \pm 0.29 \]

Figure 5.2: The \( w(z) \) priors, \( w(z) \) reconstructions and parameter constraints for each of the 5 model extensions beyond \( \Lambda \)CDM. The leftmost plots are the prior space on the function \( w(z) \) as a result of our uniform nodal reconstruction parameters and CosmoMC’s sampling, and the central plots show the constraints on \( w(z) \) as a result of the data. These plots show the posterior probability \( \Pr(w(z)) \): the probability of \( w \) as normalised in each slice of constant \( z \), with colour scale in credible region values. The \( 1\sigma \) and \( 2\sigma \) credible regions are plotted as black lines. Note that the sigma-deviations are plotted assuming a central value such that a flat prior would not have a uniform colour, thus interpreting regions of the posterior space that are highly unconstrained is more difficult, such as when interpreting the lower bounds of \( w \) at high redshifts. Reviewing priors we see a slight favouring in \( w(z) \) of the central values, as expected when calculating priors analytically and given that CosmoMC restricts the permissible parameter space. The posteriors show that the data constrains \( w(z) \) strongly compared to our priors. Rightmost are the 1D and 2D marginalised posteriors of the additional model parameters. Marginalised plots were produced using GetDist and \( w(z) \) reconstructions were produced in python with the cubehelix colour scheme by Green (2011) for linearity in grey scale.
Figure 5.3: Summarising the DE model extension results for the constraints on the $w(z)$ plane. All models are weighted by their evidences to give a model averaged plane reconstruction (Parkinson & Liddle 2013; Planck Collaboration et al. 2016f; Hee et al. 2015), and plotted as in Figure 5.2. The three plots show the prior space (top left) contracting down to the posterior odds ratio averaged $w(z)$ plane reconstruction for all of the model extensions beyond $\Lambda$CDM (top right) and for all of the models including $\Lambda$CDM (bottom). For the model extension averaged reconstruction it is clear that there is one solution around $w=-1$ and another favouring a supernegative equation of state. When including $\Lambda$CDM the significance of the supernegative solution wanes due to the associated large Bayes factor for the $w=-1$ equation.
Results: dark energy equation of state reconstruction

prior, as there is little additional structural information, and have taken the average full width
half max value of the five 3CDM amplitude parameters to estimate the effect of adding the
additional node (the prior is flat so $\sigma_\theta$ is the width, 2). This shows that the observed Bayes
factor drop of 0.54 (with errors on order 0.29) is comparable to the Occam factor and therefore
the information gained from the additional node, which should compensate the effect, is small.

The plane reconstructions show clear constraining power compared to the priors. In all
cases that allow for time dependence there is the suggestion that a supernegative equation of
state fits the data best at higher redshifts. Specifically, the $\omega$CDM model deviates from $\Lambda$CDM
by $1\sigma$ already before $z=1$ whilst the models with internal nodes, which are able to identify more
flexibly where deviations occur, suggest a $1\sigma$ deviation around $z=1.5$. No model deviates at
$2\sigma$ however. It should also be noted that the tightest constraints on the EoS are around redshift
$z=0.1$–$0.5$, and all models tend to $\Lambda$CDM in this region. This suggests that conclusions are
still data limited but that time dependent behaviour of a supernegative EoS is hinted at by the
combinations of Planck + BAO + JLA + Lyα.

We can look at the marginalised posteriors of nodes and amplitude parameters to gather
further insights. Interestingly, the 1D marginalised posteriors on the $w(z=0)$ parameters of
the models seem to favour $w>1$, whilst the $\omega$CDM model does not specifically as the single
amplitude parameter has simultaneously to model the late time behaviour. This suggests that
using $\omega$CDM simplifies the dark energy problem in a way that can obscure underlying dark
energy physics. Given that the difference in Bayes factors between $\omega$CDM and any of the
more flexible models is indistinguishable on the Jeffreys scale, using a more flexible model is
statistically valid and therefore advisable if wishing to analyse $w(z)$.

Looking at the 2D marginalised node positions in the $w(z)$ plane it is clear that in all cases
the lowest redshift node is well defined as agreeing with $\Lambda$CDM. In 1CDM, where there is only
1 internal node, the plane reconstruction takes a very similar form to $\omega$CDM as a result. For the
2CDM and 3CDM models, the additional nodes then have considerable freedom and the plane
reconstruction shape at higher redshifts reflects this via a more constant value of $w$ from about
redshift 2 onwards. The last node for both the 2CDM and 3CDM models is largely consistent
with $\Lambda$CDM as the amplitude is poorly constrained beyond $z=2$, whilst in the range $1.5<z<2.0$
it deviates by $1\sigma$, as consistent with the plane reconstructions. Generally, we conclude that
all the amplitude parameters are in good agreement with $\Lambda$CDM, which is why the additional
parameters do not generate Bayes factors that favour the models over $\Lambda$CDM.

Reviewing the model averaged plane reconstructions shown in Figure 5.3 we observe the
conclusions noted above quite clearly in the bifurcation of probabilities on $w(z)$. In the central
plot averaging over all models that allow for deviation from $\Lambda$CDM, a supernegative solution
creates a second peak in the posterior of $w$ for $z>1.5$. As the reconstruction colour represents
posteriors on \( w \) in constant slices of \( z \) measured by 1σ credible region intervals with respect to the maximum, the dual peak structure defined by the 1σ contour suggests that the data is sufficiently powerful to resolve a distinct supernegative solution. This supernegative structure is well within the 1σ credible region of the posterior distribution, fitting the data well, whilst \( w(z) = -1 \) creates the peak probability that defines the 0σ credible region. When including \( \Lambda \)CDM in the model averaging, to produce the bottom plot, again the statistical significance and consistent identification of deviations away from \( \Lambda \)CDM in the reconstructions identifies the alternative supernegative equation of state structure. However, the significant Bayes factor favouring of \( \Lambda \)CDM ensures that the functional reconstruction heavily favours \( w = -1 \) for all redshifts. When including \( \Lambda \)CDM in the model averaging, we conclude that a supernegative equation of state fits the observed data at best to within the 1.5σ credible region interval. It should be noted that the model averaging has been done over 4 models with very similar features identified, which no doubt adds to the strength of the bifurcation when averaging.

5.4 Results: Kullback-Leibler divergence and dataset analysis

To understand how the various datasets constrain the \( w(z) \) equation of state we analyse every combination of the datasets using the 2CDM model and the Kullback-Leibler divergence (\( D_{KL} \)). We chose the 2CDM model for its flexibility to capture features whilst not being as computationally demanding as 3CDM. For each combination we present the \( w(z) \) plane reconstruction to identify features visually, the single value \( D_{KL} \) to understand the total information gained and dataset constraining power, and the distribution \( D_{KL}(z) \) to localise these effects as a function of redshift. As discussed in section 5.2.4, the \( D_{KL} \) values and \( D_{KL}(z) \) functions are presented for each dataset using both the CosmoMC priors to calculate the \( D_{KL} \), which reflect the dataset information content when updating our knowledge from prior to posterior, and also using a flat prior when calculating the \( D_{KL} \) to quantify only the strength of the posterior distribution constraints.

Figure 5.4 shows the plane reconstructions and plane \( D_{KL} \) for each dataset combination in a grid of Ly\( \alpha \) versus non-Ly\( \alpha \) datasets. The \( D_{KL} \) values in brackets show the constraining power only, whilst the \( D_{KL} \) values not in brackets show the information content of the datasets. Note that for the top row, containing only Planck and Ly\( \alpha \) dataset combinations, the information content is larger than the constraining power. As discussed in section 5.2.4, this happens when the posterior peak shifts from the prior peak, and the \( D_{KL} \) analysis therefore is consistent with the observed posterior reconstruction shift to a supernegative equation of state for these four dataset combinations. From reviewing the constraining power in plane reconstructions along each row (where the combinations vary in use of Ly\( \alpha \) datasets), it appears that Ly\( \alpha \) datasets do
5.4. Results: Kullback-Leibler divergence and dataset analysis

Figure 5.4: Plane reconstructions of $w(z)$ using the 2CDM model for Planck data with each possible combination of the $A_{Ly\alpha}$, $B_{Ly\alpha}$, $BAO$ and $JLA$ datasets (abbreviated to $P$, $a$, $b$, $B$ and $S$ respectively). Results are laid out in a grid with columns of Ly$\alpha$ combinations (without any, with $a$, with $b$, and with both) against rows of $BAO$ and $JLA$ combinations (without either, $B$, $S$, and both). $D_{KL}$ values for the $w(z)$ plane reconstructions, from 2CDM prior to each given posterior, are stated next to each dataset combination to quantify the information gained when moving from prior to posterior due to the data. In brackets are the $D_{KL}$ values when moving from a flat prior to the posterior, which capture the overall constraining of the posterior whilst ignore any shifts between prior and posterior peaks. Reviewing each row from left to right shows that the Ly$\alpha$ datasets add only some constraining power, whilst reviewing each column from top to bottom shows that $BAO$ and $JLA$ datasets are both numerically and graphically significant. The axes span the prior range defined in table 6.2.
Chapter 5. Time dependence in dark energy equation of state

Figure 5.5: $D_{KL}(z)$ for all combinations of datasets, laid out as in Figure 5.4, quantifying the constraining power observed qualitatively in the plane reconstructions. The solid lines use the CosmoMC priors when computing $D_{KL}(z)$ and demonstrate the additional information gained by using the data in updating our knowledge from the CosmoMC priors to the posteriors. The dashed lines use flat priors across the $w(z)$ plane when calculating $D_{KL}(z)$ and quantify more intuitively how constrained the plane appears visually, without including the effect of the posterior shifting from the CosmoMC prior peaks. Using the CosmoMC priors shows that the Lyα datasets add much information due to this shift, whilst the posteriors themselves are less tightly constrained than when using BAO and JLA data.

not strongly affect the constraining power despite their large information content. Comparing $P$ with $Pab$ we observe an increase in constraining power of 0.14nats only. When reviewing plots along the columns (where the combinations vary in use of BAO and JLA), we visually notice a more pronounced constraint on $w(z)$ and an increase of 0.49nats when comparing $P$ with $PBS$. In general comparing $PBS$ and $PabBS$, on either measure of information content or constraining power, shows an increase of 0.1nats, which suggests that the Lyα datasets can complement the analysis even if not significantly changing the constraining power.

Figure 5.5 shows the $D_{KL}$ as a function of redshift, calculated again using both the CosmoMC priors (solid lines; information content) and flat priors (dashed lines; constraining power). Comparing $PB$ and $PS$ we observe that the peak information content of the BAO dataset is significantly smaller than the JLA peak information content. Specifically, the $PB$ dataset has a peak of 0.6nats at $z=0.3$ which is of lower magnitude but later redshift than the
PS peak of 1.2nats around $z=0.2$. The constraining power functions show that this information content is largely due to tightening posterior constraints, and we conclude that the JLA dataset is more powerful in constraining the dark energy equation of state than BAO. Reviewing the $D_{KL}(z)$ information content for the Lyα dataset combination $Pab$ shows a large and broad peak of almost 2nats at redshift 0.4, suggesting that the Lyα dataset contains significantly more information than both the BAO or JLA datasets. However this is due to a shift, and the constraining power has a significantly lower peak of only 0.6nats but over a broad redshift range.

When analysing which datasets may primarily support deviations from ΛCDM, it is interesting to note that the addition of the Lyα datasets pushes the high redshift constraints away from $w=-1$ further towards the supernegative. The PB combination plane reconstruction shows that $w=-1$ is on the 1σ contour over the range $1.5 < z < 2.0$, whilst the $PabB$ combination disfavours $w=-1$ at more than 1σ for $z > 1.5$. This is similar for the PS and $PabS$ comparison. Generally though, the plane reconstructions of most combinations either favour or approach a supernegative $w(z)$ for $z > 1.5$ at a 1σ level even without the Lyα datasets and the constraints often broaden out for $z > 2$ to be consistent with ΛCDM due to a lack of data (as can be observed by the trailing off in the $D_{KL}(z)$ plots at higher redshift). Therefore we do not attribute supernegative behaviour strongly to any single dataset when combining them. Another deviation from $w=-1$ can be observed in the combination $PaB$ at low redshift, where this time $w > -1$ is favoured. Generally, the BAO dataset seems to favour a less negative equation of state for $z < 0.5$, whilst JLA is consistent with $w=-1$ at the same period and the Lyα datasets favour a supernegative $w$-value at all redshifts (which Planck does too).

Generally, from the dashed $D_{KL}$ plots, we conclude that for the Lyα datasets a broad but small peak in $D_{KL}(z)$ at around $z=1$ can be observed to complement the BAO and JLA datasets (when comparing PB with $PabB$, PS with $PabS$ and PBS with $PabBS$) by increasing $D_{KL}(z)$ for $z > 1.5$\(^a\). Comparing the $PabBS$ plane reconstruction figure (or any dataset combination) with the corresponding $D_{KL}(z)$ plot shows good agreement with the qualitative conclusion that the datasets provide the most constraining power at redshift 0.1–0.5, and now provide a clear quantification of this effect together with a more precise conclusion: the constraining power for the $PabBS$ dataset and CosmoMC prior combination peaks at redshift 0.25 at 2.1nats whilst the dataset maximises information gain at redshift 0.2 with 1.5nats.

\(^a\)Note that taking the difference of two $D_{KL}(z)$ graphs does not represent the information gained or lost between combinations, but the observed change in shape is what we are commenting on. The addition of $ab$ raises $D_{KL}(1.5 < z < 2)$ slightly and tightens the plane reconstruction contours for higher redshifts.
5.5 Conclusions

We have presented a detailed Bayesian model selection analysis applied to the nodal reconstruction of $w(z)$, concluding that the Bayes factors on the Jeffreys scale ‘slightly favour’ $\Lambda$CDM when compared to $w$CDM and ‘significantly disfavour’ the $\tau$CDM, $1$CDM, $2$CDM and $3$CDM models, with an error on the Bayes factors of around 0.29. Despite this favouring, a model averaging approach presents a bifurcation of the $P(w|z)$ plane reconstruction space which shows that, whilst $w=-1$ for all redshift is strongly favoured, a supernegative $w(z)$ equation of state at redshift $z>1.5$ within the $1.5\sigma$ credible region of the posterior on $w(z)$ is supported by the data.

To understand this possible deviation we analysed the constraining power of the datasets using the Kullback-Leibler divergence ($D_{KL}$). We calculated a novel function $D_{KL}(z)$ to analyse the information gained when moving from the prior distribution of $w(z)$ to the posterior distribution, in slices of constant $z$, as well as a single $D_{KL}$-value for the whole plane. For each we used both CosmoMC priors and flat priors to observe information gain due to the data and the overall constraining power respectively, and we analysed each permutation of datasets using the 2CDM model. We observed that the $BAO$ and $JLA$ datasets constrained the $w(z)$ plane much more strongly than the Ly$\alpha$ datasets used. These two datasets had a strong peak at redshifts $<0.5$ whilst the Ly$\alpha$ datasets peaked more broadly at $z=1$. As expected, the combination of all datasets had the greatest constraining power, specifically the $Planck$ dataset alone had $D_{KL}=0.33$nats, the combination with $BAO$ and $JLA$ datasets had $D_{KL}=0.82$nats and the combination $Planck + BAO + JLA + Ly\alpha$ had $D_{KL}=0.91$nats. The same dataset combination had a maximum information gain at redshift 0.2 of 1.5nats. Reviewing the plane reconstructions and $D_{KL}(z)$ functions showed that the Ly$\alpha$ datasets provided additional constraints at $z>1.5$ that favours a supernegative equation of state, with $\Lambda$CDM outside of the $1\sigma$ credible region.

Generally, many of the dataset combinations have $\Lambda$CDM outside of the $1\sigma$ credible region around $1.5<z<2$, with higher redshifts being too poorly constrained to draw conclusions. For redshifts below 1.5, the Ly$\alpha$ datasets favoured a supernegative $w(z)$, the $JLA$ dataset typically agrees with $\Lambda$CDM and the $BAO$ dataset tends towards $w>-1$ values (around $1\sigma$ significance at $z=0.25$). Concluding on the higher redshift deviations, we do not attribute this supernegative favouring to a particular dataset, but note that the inclusion of Ly$\alpha$ data adds prominence as it provides a small amount of much needed constraining power over that range.

In the future, the conclusions of an analysis with these techniques will strengthen as data quality improves. The nodal reconstruction has again been shown to be useful in constraining cosmological models and developing a model independent data driven analysis (Vázquez et al. 2012b; Vázquez et al. 2012c; Aslanyan et al. 2014; Planck Collaboration et al. 2016f; Hee et al. 2015). In addition, the novel formalism introduced here of the Kullback-Leibler divergence as
a function of redshift provides a quantitative analysis of dataset information content applied to specific cosmological problems. Future applications of this method with upcoming mission and survey data or for forecasting with mock-data will provide useful insights into the value of datasets in constraining our cosmological models.
Double dark energy model investigation

Continuing to investigate dark energy, this chapter investigates two specific dark energy models which are alternatives to the concordance $\Lambda$CDM model. The first introduces a ‘missing’ matter component with equation of state $w = -2/3$ and the second studies a ‘double dark energy’ model. Neither is favoured over $\Lambda$CDM, nor can they be ruled out due to insufficient data quality and high levels of degeneracy in the model.

The missing matter model is motivated by a conformal time symmetry which requires the specific equation of state value chosen. The double dark energy is a natural extension of the missing matter model where the equation of state is allowed to vary, rather than fixing it to the required symmetry value. This chapter is adapted from a paper in preparation which updates an old but similar analysis. My contribution to this paper in preparation is specifically the model selection and posterior analysis which defines the slight disfavouring of the model, as well as describing the effects of these components on the observable power spectra and luminosity distance. I contribute only briefly to the theoretical framework but it is reviewed here for context.

This chapter presents a model-driven analysis of the dark energy equation of state whilst chapter 5 presents a data-driven analysis. This chapter concludes the contributions of this thesis to the study of dark energy equation of state behaviour. The contribution can be summarised broadly as finding agreement with $\Lambda$CDM where ever we look, though many interesting details are worth further consideration (such as phantom dark energy favouring and the potential existence of double dark energy).
6.1 Introduction

Over the past decade cosmological observations have confirmed that the background expansion of the universe is accelerating Riess et al. (1998); Perlmutter et al. (1999). This remarkable phenomenon is usually explained by a single dark energy component modelled as a perfect fluid with a potentially time varying equation of state parameter $w(z)$, such as the phenomenological analysis in chapters 4 and 5. The simplest form of dark energy is a cosmological constant $\Lambda$, which corresponds to a constant equation of state $w = -1$. Together with cold dark matter, which is key to explaining the evolution of structure in the universe, the cosmological constant gives rise to the standard $\Lambda$CDM model. This standard model fits well with existing cosmological observations, as discussed in more detail in section 2.1. There have been a large number of other exotic forms of matter proposed to provide alternative explanations for the current accelerating universal expansion Copeland et al. (2006); Durrer & Maartens (2008), including, for example, topological defects Vilenkin (1985). Some additional models are discussed in section 2.2, and we note that generally the $\Lambda$CDM model is favoured over any such extensions.

In this chapter, we remain focussed on the $\Lambda$CDM model, but with the inclusion of a second, additional, dark energy component. This additional component is termed the “missing matter” component denoted by subscripts $X$ and different to the cosmological constant due to having a different equation of state parameter. One of the motivations for exploring such a possibility arises from a symmetry of the Friedmann equations when written in terms of conformal time. In an (unpublished) paper (Vazquez et al. 2012a) the symmetry was introduced and discussed in terms of the invariance of the development of the scale factor under the swapping of roles of several of the energy densities. In this revised version, an alternative way of viewing the symmetry is presented which is simpler to explain and avoids having to think about replacing one type of physical component of the universe with another. We will discuss this revised form of symmetry shortly but, to set the scene, let us first discuss the background to the original suggestion about why a particular type of physical component might be ‘missing’.

For a homogeneous and isotropic universe described by the Friedmann–Lemaître–Robertson–Walker (FRW) metric with curvature parameter $k$, the Friedmann equation describing the dynamical evolution of the scale factor $a(t)$ can be written as

$$\left( \frac{H}{H_0} \right)^2 = \sum_i \Omega_{i,0} a^{-3(1+w_i)},$$

(6.1)

where $H = \dot{a}/a$ is the Hubble parameter (the dot denotes differentiation with respect to cosmic time $t$), and the energy density $\rho_i$ of each of the constituent components of the universe is taken into account through a corresponding density parameter $\Omega_{i,0} = 8\pi G \rho_{i,0}/(3H_0^2)$. Subscript $0$ refers to evaluation at the time $t_0$ at which $a(t) = 1$ (typically chosen to be the current time now),
but that there is no necessary link with the present day — $t_0$ is just some reference time. The equation of state parameters are $w_i$, which we will assume throughout to be time independent. The summation in equation (6.1) also includes the curvature density parameter $\Omega_k$, so that $\sum_i \Omega_{i,0} = 1$.

In the $\Lambda$CDM model, the total density parameter usually consists of contributions from radiation ($w = \frac{1}{3}$), matter (typically modelled as dust with $w = 0$), curvature ($w = -\frac{1}{3}$), and the cosmological constant ($w = -1$). These are listed in table 6.1, in which one can see an obvious ‘gap’ that we term ‘missing matter’ with $w = -\frac{2}{3}$. Interestingly, forms of matter have been proposed for which $w = -\frac{2}{3}$, such as domain walls Conversi et al. (2004); Battye et al. (1999); Mithani & Vilenkin (2012), or particular scalar field models Caldwell et al. (1998).

The above observation is only suggestive of a neglected additional component. It can be shown, however, that the existence of an additional component with $w = -\frac{2}{3}$ is required if the universe is to have a particular symmetric development in conformal time. We introduce this symmetry by looking briefly at the case of a radiation-filled universe with cosmological constant, and then pass to the more general case with matter included below.

### 6.2 Conformal time development of a flat-$\Lambda$ radiation-filled universe

A universe with just radiation and a cosmological constant may seem artificial, but it in fact corresponds well to the initial and final stages of a real universe containing matter (since radiation dominates at the beginning, and $\Lambda$ at the end). The solutions for the time development of the main parameters of such a universe can be expressed simply in terms of cosmic time $t$,
using the definition $H_\infty^2 = \Lambda/3$, as

$$a(t) = a_{eq} \sinh^{1/2}(2H_{eq}t),$$

$$a_{eq}^4 = \frac{8\pi G \rho_{r,0}}{3H_\infty^2},$$

$$\rho_r(t) = \frac{\rho_{r,0}}{a^4(t)},$$

$$H(t) = H_\infty \coth(2H_{eq}t).$$

Subscripts $eq$ refer to the instant $t_{eq}$ when the radiation energy density $\rho_r$ is equal to the vacuum energy density $\Lambda/(8\pi G)$, and as above, $\rho$ refers to the time $t_0$ when $a = 1$.

We could also write these solutions in terms of conformal time $\eta$, related to cosmic time by $d\eta = dt/a$. As discussed in Ibison (2011), the motivation for working in terms of $\eta$ is that, for currently accepted values of the density parameters $\Omega_i$, the conformal time intervals since the Big Bang ($a = 0$) and until the conformal singularity ($a = +\infty$) are both finite. Although the cosmic time since the Big Bang is finite, the future singularity occurs at $t = \infty$ such that the same cannot be said for the cosmic time formalism. This asymmetry means that it is more natural to work in terms of conformal time if one is to realise scenarios such as Penrose’s Cycles of Time model (Penrose 2010). Moreover, it is worth noting that, like cosmic time which corresponds to the proper time of comoving observers, conformal time also has a clear operational definition: the time kept by a clock whose ‘tick’ is the bounce of a light pulse confined to a pair of parallel comoving mirrors separating with the Hubble flow.

Making the transformation to conformal time for equations (6.2) results in solutions in terms of elliptic functions. Note that it is possible to show (see section 6.3 below) that the ‘epoch of equality’ $\eta_{eq}$ always occurs exactly half way through the total conformal time evolution from the big bang to the future singularity. Moreover, the development after equality is identical to that before equality if we work in terms of a reciprocal scale factor defined by $a' = a_{eq}^2/a$. This equivalence is illustrated in figure 6.1.

All radiation-filled flat-$\Lambda$ universes have this same basic symmetry: the development of the scale factor after the mid-point in conformal time evolution is the reciprocal (up to an overall multiplicative constant) of the development up to the mid-point. The value chosen for $a_{eq}$ is in fact arbitrary, and just determines the units of conformal time once $\Lambda$ has been specified. It is sensible to use $a_{eq} = 1$ in this case, so that the reciprocal relation is just $a' = 1/a$.

### 6.3 Inclusion of matter

The symmetry is interesting in connection with attempts, such as the Penrose Conformal Cyclic model, to relate the final singularity in conformal time to the big bang. We have shown above
6.3. Inclusion of matter

Figure 6.1: The red solid curve shows the evolution of the natural logarithm of the scale factor in a flat radiation-filled universe with $\Lambda$ given by recent estimates (specifically we take $\Omega_\Lambda = 0.7$ and $H_0 = 72 \text{ km s}^{-1} \text{ Mpc}^{-1}$), and with the unit of time and space given by 1 Mpc. As an example, we have arbitrarily taken $a_{eq} = 2^{1/4}$ and in these units, the future conformal boundary is at $\eta_{tot} = 7759.25$. The blue dashed curve is the red curve reflected left-right about $\eta = \eta_{tot}/2$. These curves are symmetrical not just left-right, but top-bottom if the line of reflection is taken through the value of $a$ at the mid-point, i.e. where $a = 2^{1/4}$. We can thus put the curves on top of one another if we use the reciprocal, $a' = a_{eq}^2/a = \sqrt{2}/a$. The blue curve is then flipped and slid up the right amount to lie on top of the red curve.

that in a radiation-only universe with $\Lambda$, the future conformal singularity is approached in a manner identical (as a function of $1/a$) to the way the big bang is exited as a function of $a$.

Here we show that the symmetry can survive if matter is included, provided a suitable amount of the component labelled “missing matter” in table 6.1 is present.

Making the change of variable $d\eta = dt/a$ in the Friedmann equation (6.1), and including an additional missing matter component $X$, one obtains

$$
\frac{1}{H_0^2} \left( \frac{da}{d\eta} \right)^2 = \Omega_{r,0} + \Omega_{m,0}a + \Omega_{k,0}a^2 + \Omega_{X,0}a^3 + \Omega_{\Lambda,0}a^4, \quad (6.3)
$$

where we note that the right-hand side is simply a fourth-degree polynomial in $a$. Guided by what we have just seen in the radiation case, we make the change of variable $\tilde{a}(\eta) = c^2/a(\eta)$, where $c$ is a constant. This immediately yields

$$
\frac{1}{H_0^2} \left( \frac{d\tilde{a}}{d\eta} \right)^2 = c^4 \Omega_{r,0} + c^2 \Omega_{m,0} \tilde{a} + \Omega_{k,0} \tilde{a}^2 + \frac{1}{c^2} \Omega_{m,0} \tilde{a}^3 + \frac{1}{c^4} \Omega_{r,0} \tilde{a}^4. \quad (6.4)
$$
We thus obtain an identical equation in the new variable, $\tilde{a}$, if the densities are related by

$$\Omega_{m,0} = c^2 \Omega_{X,0}, \quad \text{and} \quad \Omega_{r,0} = c^4 \Omega_{\Lambda,0}. \tag{6.5}$$

Note that the left hand sides of equation (6.3) and (6.4) are invariant under $\eta \mapsto -\eta$, and that the right hand sides do not contain $\eta$ explicitly. If the conditions in equation (6.5) are satisfied, and if we measure $\eta$ from the point where $\tilde{a} = a$, i.e. where $a^2 = c^2$, then for general $\eta$ we will have $a(\eta)a(-\eta) = c^2$. The relevance of satisfying equation (6.5) is that this leads to the derivatives of $a$ and $\tilde{a}$ matching at this point, which is of course necessary if the function is to go smoothly through the point, whilst at the same time tracing out the reciprocal behaviour. We note this behaviour will be obtained even with curvature included, since this does not require any special values of $\Omega_{k,0}$ for the symmetry to work. As a concrete example of this behaviour, we show in figure 6.2 the evolution of the energy densities of the components as a function of both conformal time and cosmic time, in a flat ($\Omega_{k,0} = 0$) case where equation (6.5) is satisfied, with $c^2 = 10$. Specifically, in this illustrative case, we have chosen

$$\Omega_{m,0} = 100 \Omega_{\Lambda,0}, \quad \Omega_{r,0} = 100 \Omega_{\Lambda,0} \quad \text{and} \quad \Omega_{X,0} = 10 \Omega_{\Lambda,0}. \tag{6.6}$$

These particular values mean that, for example, the radiation and matter densities should be equal at $a = 1$, and the ‘missing matter’ and vacuum energy densities should be equal at $a = 10$, both of which can be verified easily from the third plot of figure 6.2.

We see in this case that we have indeed obtained symmetry in the density parameters about the mid-point in conformal time, and moreover that the $a(t)$ plot is again symmetric under flipping about the horizontal axis going through the value at the mid-point ($a = \sqrt{10}$). Therefore the solution is symmetric in the inverse scale factor in the same way as for the radiation-only case discussed in section 6.2. Note that it is easy to extend this example to include curvature, with results unchanged.

A caveat to what has been said so far, is that it should be noted that the true equation of state parameter $w_i$ for each component will, in general, differ from the canonical values listed in table 6.1 (although these values are assumed in most cosmological analyses). For example, non-relativistic matter does not have exactly zero pressure ($w = 0$), but a pressure proportional to $(\nu/c)^2$. Similarly, relativistic particles such as massive neutrinos have an equation of state parameter slightly less than $w = \frac{1}{3}$, which changes with cosmic epoch. As a consequence, the right-hand side of equation (6.3) will not, in general, be a fourth-degree polynomial, in which case it no longer has the opportunity to remain form-invariant under the reciprocity transformation $\tilde{a}(\eta) = c^2/a(\eta)$.

Nevertheless, the basic notion of symmetric behaviour at the Big Bang and conformal singularity remains valid, and particularly so since at these extremes only massless particles are
6.3. Inclusion of matter

<table>
<thead>
<tr>
<th>$\Omega_r$</th>
<th>$\Omega_m$</th>
<th>$\Omega_X$</th>
<th>$\Omega_\Lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a=1$</td>
<td>$a=\sqrt{10}$</td>
<td>$a=10$</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 6.2:** Evolution with matter and ‘missing matter’ present in the proportions discussed in the text. Top figure shows evolution of radiation, matter, missing matter and vacuum energy densities as a function of conformal time $\eta$, while second panel is the same but in cosmic time $t$. The final panel is the evolution of the natural logarithm of the scale factor versus conformal time over the same period.
likely to be present (as argued by Penrose in the CCC model). We thus think it is interesting to explore the symmetry discussed here as a possible approximate symmetry of our universe. The key to this working at all is the existence of the ‘missing matter’ component, which has to be present in the proportion discussed in equation (6.5). This model can be implemented in CAMB, and so this possibility seems worth testing.

Once the possibility of adding an extra component to the energy content of the universe is tested, it is natural to extend the investigation by allowing its equation of state parameter to vary freely (rather than fixing it to \( w = -\frac{2}{3} \)). This more generic ‘double dark energy’ model comes at the cost of breaking the reciprocity invariance of the Friedmann equation, even if the equation of state parameters of the other components are assumed to have their canonical values. This model is also of interest in its own right since the observed acceleration of the universal expansion may be driven by more than just a single dark energy component. We note that a generic two component model of dark energy has previously been considered in Gong & Chen (2007).

The structure of the rest of this chapter is therefore as follows. In section 6.4, we give a brief summary of the phenomenology of an additional missing matter component with \( w = -\frac{2}{3} \) by investigating its effect on the expansion history of the universe. In particular effects are discussed for the distance-redshift relation, and on the evolution of perturbations measured by the cosmic microwave background (CMB) and matter power spectra. In section 6.5, we describe briefly the Bayesian parameter estimation and model selection analysis methodology (as a recap of section 3.1) and the cosmological data sets used to set constraints on our ‘missing matter’ and ‘double dark energy’ models. The results of these analyses are given in section 6.6 and our conclusions are presented in section 6.7.

### 6.4 Phenomenology

It is of interest to investigate the phenomenology of a cosmological model containing a second component \( X \) with negative pressure (in the event that the energy density is positive), in addition to a cosmological constant. Since our missing matter model (for which \( w_X = -\frac{2}{3} \)) is just a special case of our double dark energy model, we will focus here on the former as being a representative example of the latter.

The effect of the additional component \( X \) on the global expansion history of the universe depends only on the equation of state parameter \( w_X \), whereas its effect on the evolution of perturbations will also depend on the nature of the component \( X \), in particular its assumed dynamical properties. We therefore consider these two issues separately.
6.4. Phenomenology

![Figure 6.3: Dimensionless luminosity distance $H_0 d_L(z)$ as a function of redshift $z$ for a concordance $\Lambda$CDM cosmology with an additional component $X$ with equation of state parameter $w_X = -\frac{2}{3}$, for different values of $\Omega_{X,0}$ (and adjusted $\Omega_{\Lambda,0}$).](image)

6.4.1 Background evolution

The global expansion history of the cosmological model is most conveniently represented through the distance-redshift relation. Indeed, comparing the predicted relation between the luminosity distance $d_L$ and redshift $z$ of an object with observations of astronomical ‘standard candles’, such as Type Ia supernovae, has provided the most direct and convincing evidence that the expansion of the universe is accelerating.

The luminosity distance to an object at redshift $z$ is given by

$$d_L(z) = (1 + z) \frac{S_k(\sqrt{|\Omega_{k,0}|} \chi(z))}{\sqrt{|\Omega_{k,0}|}},$$

where $S_k(x) = \sinh x$, $x$, $\sin x$ for spatial curvature parameter $k = -1$, 0, +1 respectively, and the comoving radial coordinate $\chi(z)$ is determined by the expansion history:

$$\chi(z) = \int_0^z \frac{dz}{H(z)},$$

where $H(z)$ is obtained from the Friedmann equation (6.1). The inclusion of $\Omega_{X,0}$ into equation (6.1) therefore directly affects the expansion history embodied in $H(z)$, and hence can serve either to increase or decrease the luminosity distance $d_L(z)$ to an object at redshift $z$. Figure 6.3 illustrates this effect for a few representative values of $\Omega_{X,0}$. If $\Omega_{X,0} > 0$, the apparent luminosity is increased and hence the luminosity distance is reduced compared to the standard $\Lambda$CDM model. The opposite effect occurs for $\Omega_{X,0} < 0$.

The power of the luminosity distance as a cosmological probe resides in the fact that it can be simply related to apparent brightness $m(z)$ obtained directly from a set of standard candles,
each (assumed to be) of absolute magnitude $M$, namely

$$m(z) = M + 5 \log_{10} \left[ \frac{d_L(z)}{\text{Mpc}} \right] + 25, \quad (6.9)$$

where the constant offset ensures the usual convention that $m = M$ for an object at $d_L = 10$ pc. Type Ia supernovae constitute a set of ‘standardizable candles’ that can be used to constrain cosmological models in this way Amanullah et al. (2010).

It should be pointed out that, for the background evolution, the combination of a cosmological constant with $w = -1$ and an additional component $X$ with constant $w_X$ is equivalent, under certain conditions outlined below, to a single dark energy component with a time varying equation of state parameter $w_{\text{eff}}(a)$ given by the ratio of the combined pressure of the two components to their combined density Gong & Chen (2007), namely

$$w_{\text{eff}}(a) = \frac{-\Omega_{A,0} + w_X \Omega_{X,0} a^{-1}}{\Omega_{A,0} + \Omega_{X,0} a^{-1}}. \quad (6.10)$$

Examples of such models have been studied extensively Chevallier & Polarski (2001); Jassal et al. (2004); Sendra & Lazkoz (2012); Rubin et al. (2009); Akarsu et al. (2015), albeit not with the particular form of $w_{\text{eff}}(a)$ given above. Plotting $w_{\text{eff}}(a)$, assuming possible values of $\Omega_{X,0}$ and $\Omega_{A,0}$, shows that the variation with either $a$ or redshift $z$ is non-linear, so $w_{\text{eff}}(a)$ is not contained within either of the common $w(z) = w_0 + w_1 z$ or $w(a) = w_0 + w_a (1 - a)$ parameterisations. More importantly, it should be noted that if $\Omega_{A,0}$ and $\Omega_{X,0}$ have different signs, as we allow in our analysis in section 6.5, then $w_{\text{eff}}(a)$ becomes singular at $a = |\Omega_{X,0} / \Omega_{A,0}|$. Therefore, if $\Omega_{A,0}$ or $\Omega_{X,0}$ (or both) are allowed to take positive and negative values, then our missing matter (or double dark energy) model is not well described by a single time varying dark energy component.

### 6.4.2 Evolution of perturbations

An additional component $X$ will affect the growth of perturbations through its contribution to $H(z)$ and the evolution of the matter density. Moreover, we assume here that $X$ has the same dynamical behaviour as that usually assumed for a generic dark energy component. In particular, we use the CAMB Lewis et al. (2000) dark energy module developed by Fang et al. (2008), in which dark energy is assumed itself to exhibit Gaussian adiabatic perturbations. As the equation of state parameter approaches $w = -1$, the effects of the dark energy perturbations disappear, as one would expect for a pure cosmological constant.\footnote{It should be borne in mind, however, that a possible physical instantiation of an additional component $X$ with $w_X = -\frac{3}{2}$ could be in the form of domain-wall topological defects, for example, in which case the effect on the generation and evolution of perturbations may be very different to that assumed here.} We modified the CAMB...
code to include our additional component and calculated the predicted power spectra of cosmic microwave background (CMB) anisotropies and matter perturbations.

We plot the CMB power spectra in figure 6.4 for several choices of $\Omega_{X,0}$ with remaining cosmological parameters set to their concordance values (with $\Omega_{A,0}$ varying accordingly to ensure $\sum_i \Omega_i = 1$). From figure 6.4 we see that, as one might expect, the main effect of a non-zero $\Omega_{X,0}$ is to shift the positions of the acoustic peaks which are sensitive to the spatial geometry of the universe. Therefore one would expect constraints on $\Omega_{X,0}$ from CMB observations to be degenerate with the constraints on $\Omega_A$ and $\Omega_k$. For positive values of $\Omega_{X,0}$, we also see an enhancement of power on the largest scales from the late time ISW effect.

In figure 6.5, we plot the predicted matter power spectra for different values of $\Omega_{X,0}$; again the other parameters are set to their concordance values with $\Omega_{A,0}$ varied to incorporate the missing matter density. We see that the dominant effect of the additional component is on the normalisation of the matter power spectrum. The amplitude of fluctuations is suppressed for $\Omega_{X,0} > 0$ and enhanced for $\Omega_{X,0} < 0$, whilst the positions of the acoustic oscillations, which depend on the matter density, are unaffected by the introduction of the additional component.

It is worth noting that, although the background evolution of the universe is identical for our missing matter (or double dark energy) model and for a model with a single time varying dark energy component defined by equation (6.10) (provided $\Omega_{A,0}$ and $\Omega_{X,0}$ have the same sign), the evolution of perturbations is generally different. This is true even in the simplest case where one assumes the same dynamical behaviour for the generic dark energy components in the two models, namely that they exhibit Gaussian adiabatic perturbations. This is illustrated in figure 6.6, in which we plot the CMB and matter power spectra for a specific example of each
model. We reiterate our earlier comment that the many previous studies of models containing a single time varying dark energy component are therefore not equivalent to the study presented here.

6.5 Analysis

We now perform a Bayesian parameter estimation and model comparison analysis of our ‘missing matter’ and ‘double dark energy’ models, using recent cosmological observations. In particular, we use the Planck 2015 data release temperature measurements Planck Collaboration et al. (2016b) and lensing data Planck Collaboration et al. (2016e). In addition to CMB data, we include distance measurements of 740 Supernovae Ia from the SNLS-SDSS collaborative effort called the joint light-curve analysis (JLA; Betoule et al. (2014)) and several baryon acoustic oscillation (BAO; Anderson et al. (2014); Beutler et al. (2011); Ross et al. (2015); Delubac et al. (2015); Font-Ribera et al. (2014)) measurements of distance.

Throughout the analysis we consider purely Gaussian adiabatic scalar perturbations and neglect tensor contributions. We assume a modified \( \Lambda \)CDM model specified by the following parameters: the physical baryon density \( \Omega_b h^2 \) and CDM density \( \Omega_{DM} h^2 \), where \( h \) is the dimensionless Hubble parameter such that \( H_0 = 100h \text{ km s}^{-1}\text{Mpc}^{-1} \); the curvature density \( \Omega_{k,0} \) of the universe; \( \theta \), which is 100 times the ratio of the sound horizon to angular diameter distance at last scattering surface; the optical depth \( \tau \) at reionisation; and the amplitude \( A_s \) and spectral index \( n_s \) of the primordial perturbation spectrum measured at the pivot scale \( k_0 = 0.05 \text{ Mpc}^{-1} \). These are as described in section 3.3. We also include 17 nuisance parameters associated with the Planck and JLA datasets. The ranges of the uniform priors assumed on the

---

Figure 6.5: Matter power spectra for a concordance \( \Lambda \)CDM model with an additional component \( X \), with equation of state parameter \( w_X = -\frac{2}{3} \), for several values of \( \Omega_{X,0} \).
standard ΛCDM parameters are listed in table 6.2, with nuisance parameter priors set to the advised values. Our hypothetical additional component is characterised by its density parameter \( \Omega_{X,0} \) and equation of state parameter \( w_X \). We assume a uniform prior on \( \Omega_{X,0} \) in the range \([-1, 2]\) throughout. For the missing energy model, we have \( w_X = -\frac{2}{3} \), and for the double dark energy model we assume the uniform prior \( w_X = [-\frac{3}{2}, -\frac{1}{2}] \).

To carry out the exploration of the parameter space, we first incorporate the extra component into the standard cosmological equations, by performing the minor modifications to the \textsc{camb} code \cite{Lewis2000} described in section 6.4.2 (which implement a parameterised post-Friedmann (PPF) prescription for the dark energy perturbations \cite{Fang2008}). We then include into the \textsc{cosmomc} code \cite{Lewis2002} a fully-parallelised version of the
### Table 6.2: Ranges of the uniform priors assumed on the standard ΛCDM parameters in the Bayesian analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_{b,0}h^2$</td>
<td>[0.019, 0.025]</td>
</tr>
<tr>
<td>$\Omega_{\text{dm},0}h^2$</td>
<td>[0.095, 0.145]</td>
</tr>
<tr>
<td>$\Omega_{k,0}$</td>
<td>[−0.05, 0.05]</td>
</tr>
<tr>
<td>$\theta$</td>
<td>[1.03, 1.05]</td>
</tr>
<tr>
<td>$\tau$</td>
<td>[0.01, 0.4]</td>
</tr>
<tr>
<td>$n_s$</td>
<td>[0.9, 1.1]</td>
</tr>
<tr>
<td>$\ln[10^{10} A_s]$</td>
<td>[2.7, 4.0]</td>
</tr>
</tbody>
</table>

Nested sampling algorithm PolyChord Handley et al. (2015a,b), which significantly increases the efficiency of calculating the Bayesian evidence and also reliably produces posterior samples even from distributions with multiple modes and/or high dimensionality. A suitable guideline for making qualitative conclusions has been laid out by Jeffreys (1961): if $B_{ij} < 1$ model $i$ should not be favoured over model $j$, $1 < B_{ij} < 2.5$ constitutes significant evidence, $2.5 < B_{ij} < 5$ is strong evidence, while $B_{ij} > 5$ would be considered decisive.

### 6.6 Results

For comparison purposes, we first assume no additional component $X$, in order to determine the constraints imposed by the current combined data sets on the standard ΛCDM model. In particular, we find the data indicate the dominance of dark energy in the form of a cosmological constant with $\Omega_{\Lambda,0} = 0.696 \pm 0.007$, followed by matter density (dark matter + baryons) $\Omega_{m,0} = 0.305 \pm 0.007$, and an almost negligible spatial curvature $\Omega_{k,0} = -0.0013 \pm 0.0024$. We also obtain the present Hubble parameter $H_0 = 67.78 \pm 0.70$. The constraints on the other parameters $\{\theta, \tau, A_s, n_s\}$ remain essentially unaffected by the introduction below of our additional component $X$, and so we do not consider them further.

#### 6.6.1 Missing matter model

The inclusion of a missing matter component $X$ with $w_X = -\frac{2}{3}$ considerably broadens the parameter constraints. In particular, we find: $\Omega_{\Lambda,0} = 0.734 \pm 0.083$, which constitutes an order of magnitude increase in the error bars as compared with the standard ΛCDM model, $\Omega_{m,0} = 0.302 \pm 0.010$, $\Omega_{k,0} = -0.0023 \pm 0.0029$ and $H_0 = 68.10 \pm 1.04$. Figure 6.7 shows 1D and 2D marginalised posterior distributions for the density parameters (note that $\Omega_{m,0} = 1 - \Omega_{\Lambda,0} - \Omega_{k,0} - \Omega_{X,0}$). As expected, we observe a clear degeneracy between $\Omega_{X,0}$ and $\Omega_{\Lambda,0}$.
6.6. Results

Figure 6.7: 1D and 2D marginalised posterior distributions for density parameters in the missing matter model (note that $\Omega_{m,0} = 1 - \Omega_{\Lambda,0} - \Omega_{k,0} - \Omega_{X,0}$). The 2D constraints are plotted with 1σ and 2σ credible regions contours and the cubehelix colour map Green (2011).

and slight degeneracy between $\Omega_{X,0}$ and $\Omega_{k,0}$. The 1D constraint on the density parameter of missing matter is $\Omega_{X,0} = -0.034\pm0.75$. The current data prefer a slightly negative value, which is difficult to interpret physically, but the errors suggest this not to be a significant favouring. The 1D marginal shows moderate relative probability even for $\Omega_{X,0} \sim 0.1$, and so the presence of an appreciable missing matter component cannot be ruled out. Our results are, however, still consistent with a standard $\Lambda$CDM model.

This view is supported by our Bayesian model comparison. We find that the log-evidence difference (or Bayes factor) between the missing matter model and the standard $\Lambda$CDM model is $\mathcal{B}_{\Lambda+X,\Lambda} = -1.12 \pm 0.53$. According to Jeffreys guideline Jeffreys (1961); Vázquez et al.
Figure 6.8: 1D and 2D marginalised posterior distributions for density parameters in the double dark energy model (note that $\Omega_{m,0} = 1 - \Omega_{\Lambda,0} - \Omega_{X,0}$). The 2D constraints are plotted with 1σ and 2σ credible region contours. The top-right panel shows the 3D posterior distribution in the $(w_X, \Omega_{X,0}, \Omega_{\Lambda,0})$ subspace, where the colour code indicates the value of $\Omega_{\Lambda,0}$ using the cubehelix colour map Green (2011).

(2012c), the inclusion of the missing matter component is therefore slightly disfavoured, but almost indistinguishable, from a model perspective given current cosmological data.

### 6.6.2 Double dark energy model

We now allow for the equation of state parameter $w_X$ for our additional component to be a free parameter (albeit still independent of redshift), for which we assume a uniform prior in the range $w_X = [-\frac{3}{2}, -\frac{1}{2}]$. We thus allow for the possibility that this second dark energy component

could be a form of phantom energy with \( w_X < -1 \) (Vázquez et al. 2012c). Figure 6.8 shows the resulting 1D and 2D marginalised posterior distributions for \( w_X \) and the density parameters in the model (once again, note that \( \Omega_{m,0} = 1 - \Omega_{A,0} - \Omega_{k,0} - \Omega_{X,0} \)). At the top-right of the figure we also give a representation of the 3D posterior in the \((w_X, \Omega_{X,0}, \Omega_{A,0})\) subspace, where the colour indicates the value of \( \Omega_{A,0} \).

The 1D constraints on the standard parameters are as follows: \( \Omega_{A,0} = 0.797 \pm 0.556, \Omega_{m,0} = 0.305 \pm 0.009, \Omega_{k,0} = -0.0015 \pm 0.0024, H_0 = 67.86 \pm 1.01 \). The constraints on the parameters describing the additional second dark energy component may be given as \( w_X = -1.01 \pm 0.16 \) and \( \Omega_{X,0} = -0.101 \pm 0.557 \), although these numbers obscure the nature of the marginal \((w_X, \Omega_{X,0})\)-space and \((w_X, \Omega_{A,0})\)-space distributions slightly. These results are clearly consistent with a standard \( \Lambda \)CDM model, although the inclusion of the additional dark energy component has again resulted in the uncertainties in the constraints on the standard parameters being much larger than those obtained assuming a \( \Lambda \)CDM model. Indeed, the 1D marginal for \( \Omega_{X,0} \) shows moderate relative probability even for \( \Omega_{X,0} \approx 0.25 \), although this is likely due to a value of \( w = -1 \) simply reproducing the \( \Lambda \)CDM model.

Moreover, the 2D and 3D marginal distributions in figure 6.8 have interesting features that are worth noting. As might be expected, we again see a pronounced degeneracy between \( \Omega_{A,0} \) and \( \Omega_{X,0} \). The marginal distribution in \((\Omega_{X,0}, \Omega_{A,0})\) subspace shows a strong correlation between these energy densities that would imply the potential for a trade-off between them. One might be concerned, however, that the marginal distribution plotted is strongly dominated by the contribution (after marginalising over \( w_X \)) from near \( w_X = -1 \). If so, one could then not infer the potential of a trade-off between these two energy densities at (any) other values of \( w_X \). To investigate this possibility, we also calculated the conditional distributions in \((\Omega_{X,0}, \Omega_{A,0})\) subspace for a small set of fixed \( w_X \)-values in the range \([-0.7, -1.3] \). The resulting distributions were, however, very similar to that plotted in figure 6.8, and so indicating that the two energy densities can indeed be traded-off against one another.

Also of interest is our Bayesian model comparison, which finds that the log-evidence difference (Bayes factor) between the double dark energy model and standard \( \Lambda \)CDM is \( \mathcal{B}_{\Lambda \Lambda} = -0.43 \pm 0.43 \). This shows that neither model is preferred over the other with any significance; indeed they are in the indistinguishable range of Jeffreys guideline and identical within \( 1\sigma \) of the error on the evidence calculation. Thus, the two additional parameters \( \Omega_{X,0} \) and \( w_X \) in the double dark energy model allow it the freedom to fit the data sufficiently better than \( \Lambda \)CDM to compensate for the corresponding increase in the prior volume, and hence the model is not penalised by the evidence. The Bayes factor stated is likely also aided by the broadening of posteriors on some of the parameters, as this implies a lower Occam factor associated with those parameters.
6.7 Discussion and Conclusions

We have investigated the possibility that there exist two dark energy components in the universe: a cosmological constant, with \( w = -1 \) and an additional component \( X \) with equation of state parameter \( w_X \). First, we fix the equation of state parameter of \( X \) to the value \( w_X = -\frac{2}{3} \). This ‘missing matter’ model corresponds to the special case in which the additional component is required for the Friedmann equation written in terms of conformal time \( \eta \) to be form invariant under the reciprocity transformation \( a(\eta) \to 1/a(\eta) \). Foregoing this requirement, we then consider the more general ‘double dark energy’ model, in which \( w_X \) is a free parameter assumed to have uniform prior in the range \( w_X = [-\frac{3}{2}, -\frac{1}{2}] \). For both models, we perform a Bayesian parameter estimation and model selection analysis, relative to standard \( \Lambda \)CDM, using recent cosmological observations of cosmic microwave background anisotropies, Type Ia supernovae and large scale structure.

For the missing matter model, the introduction of the additional component \( X \) significantly broadens the constraints on the standard parameters in the \( \Lambda \)CDM model, but leaves their best-fit values largely unchanged. The 1D marginalised constraint on the missing matter density parameter is \( \Omega_{X,0} = -0.034 \pm 0.075 \). Thus, current cosmological observations prefer a slightly negative value, which is difficult to interpret physically, but the posterior on this parameter is sufficiently broad that significant relative probability exits even for \( \Omega_{X,0} \sim 0.1 \), and so the presence of a missing matter component cannot be ruled out. To support this conclusion, our results are consistent with \( \Lambda \)CDM and our Bayesian model selection analysis suggests the missing matter model to be almost indistinguishable from \( \Lambda \)CDM, with a Bayes factor of \(-1.12 \pm 0.53 \) log-units of evidence.

For the double dark energy model, the constraints on standard \( \Lambda \)CDM parameters are again considerably broadened. The 1D marginalised constraints on the vacuum and second dark energy component are \( \Omega_{\Lambda,0} = 0.797 \pm 0.556 \) and \( \Omega_{X,0} = -0.101 \pm 0.557 \) (with \( w_X = -1.01 \pm 0.16 \)), respectively, which are again consistent with \( \Lambda \)CDM. Once more, however, the 1D marginalised posterior on \( \Omega_{X,0} \) is sufficiently broad that even \( \Omega_{X,0} \sim \pm 0.25 \) is not ruled out. We also find that the double dark energy model has a similar Bayesian evidence to \( \Lambda \)CDM, and hence neither model is preferred over the other.
This chapter returns to the new method developed in chapter 4. The H3L method is applied to tests of general relativity using gravitational waves. The test studied is particularly amenable to this new method as it requires the computation of 16 evidences which are summed into one final Bayes factor (or posterior odds ratio). The H3L method is combined with a nested sampling parameter error estimation technique which greatly improves the method, allowing the H3L method to compute errors on Bayes factors from a single run. The H3L method is found to be 24 times faster on a toy model, and estimated to be ≈ 100 times faster for a Kerr likelihood with a standard general relativity gravitational wave tested.

The tests using the H3L method are all performed using toy data (constructed from the models themselves) but are in principle easy to implement using real data. Noting that Abbott et al. (2016c) compute errors on Bayes factors for similar tests by repetitions of a nested sampling algorithm, we see strong potential for significant time savings in real world GW problems. This chapter is based on a paper in preparation, with the likelihood codes for the Kerr likelihood being supplied by a collaborator. My contributions have focussed on testing the new method and this chapter focusses more on that work rather than the Kerr models themselves. As the H3L method is general, the analysis can be carried over to any waveform model.
Chapter 7. H3L method acceleration in a gravitational wave test of GR

7.1 Introduction

The recent detection of gravitational waves (GWs) (Abbott et al. 2016b,d) is a historic step towards an exciting era of gravitational wave astronomy (Riles 2013; Blair et al. 2015; Vitale 2016). Space based detectors (Sesana 2016; Ni 2016; Amaro-Seoane & et al. 2017), pulsar timing array data (Hobbs & et al. 2010; Middleton et al. 2016; Babak et al. 2016; Taylor & et al. 2016) and advanced ground based detectors (Punturo et al. 2010; Somiya 2012) continue to accelerate future data gathering capabilities, vital for statistically significant astrophysical and cosmological analysis. Alongside probing fascinating astrophysical systems, gravitational wave astronomy also grants access to new tests of general relativity (GR) by comparing the detected signals to GR predictions, often in regimes inaccessible to existing tests. Tests of GR are discussed in more detail in section 2.3.

For the GW150914 and GW151226 signals, the LIGO Scientific Collaboration and the Virgo Collaboration have investigated and confirmed that the signals are consistent with general relativity to within statistical uncertainties (Abbott et al. 2016a,c). These tests involved showing that the residual signal, after subtracting the most probable GR waveform model, was consistent with noise; confirming that the inspiral and the merger-ringdown parameters are in agreement with GR to the statistical precision available; and testing phenomenological deviations in phase coefficients of the various coalescence phases, such as PN coefficients. Other authors have expanded on this analysis (Yunes et al. 2016) without finding evidence for deviations from GR.

This chapter looks to expand on such tests of GR which use gravitational waves. Specifically, the GR test of Li et al. (2012) is used, in which a Bayesian framework is developed for detecting deviations from GR in waveforms. Their GR test does not rely on any specific alternative theory of GR, but instead tests the consistency of waveform coefficients with GR predictions. It therefore provides a data driven approach to detecting potential deviations from GR, similar to the work in chapter 5 constraining the dark energy equation of state. The similarity continues as the original GR test work uses nested sampling to compute Bayes factors and answers the model selection question of whether GR or a generic phase modification is the favoured model.

This chapter aims to drastically reduce the number of likelihood calculations required to carry out this gravitational wave GR test defined in Li et al. (2012) by using the method developed in chapter 4 (which we will refer to as the H3L method). The H3L method will be used to compute Bayes factors, and in turn posterior odds ratios, and their errors will be computed using the parameter error estimation technique described by Higson et al. (2017). The efficacy will be demonstrated by implementing this combined approach for two different waveform models: firstly a simple toy sinusoidal waveform model is used, and secondly a more gravitationally accurate Kerr likelihood waveform model is tested.
In the first model, the toy model, testing for deviations from the waveform data without deformation finds that the new technique requires 24 times fewer likelihood evaluations to carry out the GR test. When a small deformation is present in the toy waveform data the efficiency gain is closer to a factor of 9, with this reduction in improvement shown to be due to the construction of the GR test rather than a weakness in the H3L method. The H3L method is then applied to a second model: that of a Kerr waveform. For the Kerr waveform a similar efficiency gain is observed at a factor of over 100 times fewer likelihood calculations required (though the Kerr likelihood is more computationally expensive and thus we cannot probe this factor as comprehensively as in the toy model).

This chapter specifically demonstrates an improvement for the GR test by Li et al. (2012), but we note that the efficacy of the H3L method is general enough that it could be applied widely and beyond the field of cosmology. Staying within the GW field, we note that the LIGO collaboration GR tests used nested sampling and computed errors on Bayes factors for the PN coefficient (and other) models by repetition of the evidence calculation (Abbott et al. 2016c). Firstly an efficiency gain would be delivered in computing errors from a single nested sampling run (a significant improvement if they carried out a statistically significant number of repetitions) and secondly an efficiency gain may be realised due to the mechanisms discussed in this chapter.

In the following sections we describe the method and the two models tested, highlighting the efficiency gains and how the method is applicable more widely to Bayes factor estimation problems. In section 7.2 we provide a summary of the statistical framework of the waveform likelihood and summarise briefly the three papers that combine for this analysis. Section 7.3 presents the new technique applied to a generic toy model which deforms a sinusoidal waveform, whilst section 7.4 applies this to the Kerr GW model. Section 7.5 summarises the efficiency gains for the two models.

## 7.2 Statistical framework

This chapter presents an amalgamation of the GW GR tests presented in Li et al. (2012), the method presented in chapter 4, and the error calculating technique by Higson et al. (2017). To show how these combine, section 7.2.1 summarises the GW GR tests presented in Li et al. (2012) and section 7.2.2 describes how the H3L method and error calculation of Higson et al. (2017) will be used.
7.2.1 The waveform likelihood and $P_{\text{modGR}}^{\text{GR}}$

In order to do both parameter estimation and model selection, a likelihood needs to be defined for the model. In gravitational wave signal processing this can be done through matched filtering on a time or frequency signal for GW strain. The likelihood is created by taking the noise weighted inner product of an injection signal $h_{\text{inj}}(t)$ and a signal generated by a specific waveform model $h(t|\theta)$:

$$\log \mathcal{L}(\theta) = -\frac{1}{2} < h_{\text{inj}}(t) - h(t|\theta) | h_{\text{inj}}(t) - h(t|\theta) >$$  \hspace{1cm} (7.1)

where $\theta$ are the model parameters that define the waveform, $t$ is the time, and $<a|b> = (a|b)/\sigma_{\text{ref}}^2$ is the noise weighted inner product. The inner product for two time series $a$ and $b$ is defined as $(a|b)=\int_0^T a(t)b(t)dt$ and approximated numerically as $(a|b)=\sum_i (a(t)\sim b(t))^2\delta t$. The noise $\sigma_{\text{ref}}$ defines a scale for deviations from the injection data and we set it to 1 unless otherwise stated. For a given waveform, the signal to noise ratio can be calculated as $\text{SNR} = <h|h>^{1/2}$, note that this takes into account the reference noise. $T$ defines the length of the time series, with a sampling rate of $\delta t$.

The waveform model to be used is defined in each of the results sections. Both of the waveform models are designed to implement the general relativity test presented by Li et al. (2012). They each have a set of underlying core parameters which define the base model, $\phi$, and a set of deformation parameters which deviate the waveform time series from the base model, $\epsilon_i|_{i=1,...,I}$, where we use $I=4$ deformation parameters throughout.

For a full description of the method, we refer the reader to Li et al. (2012). Continuing the summary of their work, these $I$ models create $2^I$ models which either do or do not include the various deformation parameters. The model with all parameters off is the base model, or the GR model in the gravitational wave testing scenario. The remaining $2^I-1$ models are non-GR models which include the presence of a deformation parameter. The test of whether the waveform data agrees well with GR is to compute the posterior odds ratio and address the model selection question of whether GR or a modification to GR (mod-GR) is favoured. The posterior odds ratio for this is given by

$$P_{\text{modGR}}^{\text{GR}} = \frac{\Pr(M_{\text{modGR}}|D)}{\Pr(M_{\text{GR}}|D)} = \sum_i \frac{\Pr(M_i|D)}{\Pr(M_{\text{GR}}|D)},$$  \hspace{1cm} (7.2)

where the probability of the mod-GR model is defined as $M_{\text{modGR}} = \bigvee_i M_i$ (‘$\bigvee$’ is the ‘or’ operator). Li et al. (2012) show that the individual models are logically disjoint, such that $\bigvee_i M_i = \sum_i M_i$. Hence $P_{\text{modGR}}^{\text{GR}}$ simplifies to the sum given in equation (7.2). As the posterior odds ratio includes the prior, we note that the prior is defined such that $\pi(M_{\text{modGR}}) = \pi(M_{\text{GR}})$. Specifically we observe that $\pi(M_i)|_{i\neq\text{GR}} = \pi(M_{\text{GR}})/(2^I-1)$ for each individual mod-GR model.
Li et al. (2012) proceed to expand the above posterior odds ratio test to take account of multiple injection waveforms to facilitate testing GR against many detected signals, ready for the age of gravitational wave astronomy. It amounts to averaging over multiple such posterior odds ratio. To demonstrate the message of the current chapter, the above GR test for a single injection suffices as the method improves the calculation of $P_{GR}^{modGR}$.

The posterior odds ratio of equation (7.2) quantifies the question “is the mod-GR model favoured over GR?”, and the method we propose decreases the number of likelihood calculations in the nested sampling algorithm required to compute $P_{GR}^{modGR}$ to a given error accuracy.

### 7.2.2 The H3L method and errors on posteriors

The H3L method presented in chapter 4 computes Bayes factors and posterior odds ratios without calculating evidences by sampling over a model selection parameter $n$. The posterior on $n$ therefore becomes the probability of the model: $Pr(M_i|D) = Pr(n = M_i|D)$. From equation (3.6) we see that the posterior odds ratio is computed from such posteriors, where normally this model posterior is calculated from the evidence in equation (3.4). The H3L method simplifies this step by computing the posterior directly using nested sampling parameterisation. The trick is in the implementation details.

Chapter 4 provides a more thorough description, but here we will overview the implementation specific details and discuss the potential for speeding up Bayes factor and posterior odds ratio calculations. In order to sample over $n$, the models for which we wish to compute posterior odds ratios are combined along the parameter direction $n$ in the likelihood space to form a hyper-model. This is a simple procedure. If we had 2 models, $M_1$ and $M_2$, that are each defined by their 2 parameters $(x_1, y_1)$ and $(x_2, y_2)$ for models 1 and 2 respectively, then the hyper-model $M_H$ will have a model space of $(x_H, y_H, n)$. When $n = M_1$, the $x$ and $y$ parameters of the hyper model are the $x_1$ and $x_2$ parameters, such that the likelihood $L_1$ is returned for the given parameter set, and similarly for $n = M_2$. Therefore one must construct a hyper-likelihood $L_H$ which returns the usual model likelihoods $L_1$ and $L_2$ depending on the value of $n$. If the priors on $x_i$ and $y_i$ are the same for each $i$, the hyper-likelihood simply passes the parameters to the specific model likelihoods. This will be the case for all the likelihoods reviewed in the chapter. If the priors are different then a suitable conversion may be required (as nested sampling algorithms often explore in the ‘unit hypercube’ (Handley et al. 2015b) this is either trivial or automatic). If one model ($M_f$) has more parameters than another ($M_i$), the superfluous parameter of model $M_f$ are simply not included in the $M_{i,f}$ likelihood such that their posteriors are unconstrained (and marginalise out to unity such that they do not affect the posterior of $n = M_f$).
Chapter 7. H3L method acceleration in a gravitational wave test of GR

The two model scenario generalises for any number of models, each partitioned along the new parameter direction $n$. Although $n$ is discrete it can be sampled over as a normal continuous parameter to avoid any alteration to the nested sampler used, where the hyper-likelihood chooses the likelihood value to return based on a binning of the continuous $n$ value (with equal bin width per model, for simplicity). The posterior on $n$ is then calculated via the usual sum of importance weights ($p_i$ in section 3.2.3 equation (3.18)) for which the continuous parameter $n$ was within the binning range defined by the hyper-likelihood. If we use $n$ to define the various models with and without deformation parameters, then the GR test POR $P^{\text{modGR}}_{\text{GR}}$ can be calculated easily as

$$P^{\text{modGR}}_{\text{GR}} = \log \left( \frac{\sum_{i \in \text{GR}} p_i}{\sum_{j \not\in \text{GR}} p_j} \right).$$

(7.3)

$i \neq \text{GR}$ implies the values of $p_i$ where $n$ is not in the range defined in the hyper-likelihood to return the GR likelihood value.

This summarises the H3L method. PORs calculated with the new method will be compared to PORs calculated by evaluating evidences for each individual model to obtain $P^{\text{modGR}}_{\text{GR}}$. The latter method of obtaining posterior odds ratios will be referred to as the vanilla method. The vanilla method computes errors on $P^{\text{modGR}}_{\text{GR}}$ automatically as the log-evidence errors propagate for two models in quadrature. This is not the case for the H3L method, and we require the analysis of Higson et al. (2017) on calculating errors on parameter functions in nested sampling.

Let us quickly summarise the method used to obtain parameter errors, for a detailed description please see Higson et al. (2017). Skilling (2006) showed that two nested sampling runs with $N^a_{\text{live}}$ and $N^b_{\text{live}}$ can be combined to form a new nested sampling run of $N_{\text{live}} = N^a_{\text{live}} + N^b_{\text{live}}$ live points. The nested sampling weights can be suitably adjusted in post-processing, as might be expected as each of the new live points chosen in each run was sampled uniformly from an iso-likelihood contour. Imagining that we had two single $N_{\text{live}}$ runs and we wished to combine them, the ordered set of combined likelihood calculations is a set of samples taken uniformly from within iso-likelihood contours. We can start a 2 live point run but instead of computing new likelihood values with the condition of uniformly sampling within the iso-likelihood contour of the lowest likelihood point (the only required condition of nested sampling), we can recognise that we already have that information. Choosing the likelihood calculations from the ordered combined set of the two single live point runs creates the required $L(\mathcal{X})$ curve for integration by quadrature, after adjusting the shrinkage approximation to reflect that there are now two points (possible as the prior mass is not calculated in the runs but only approximated).

Higson et al. (2017) realised that this process works in reverse, as long as you have information on the birth and death order of points. Adjusting the nested sampling algorithm to output the birth number of live points (the death order is usually contained in the order
of addition to the output file) is all that is required. From a single $N_{\text{live}}$ run we are able to deconstruct $N_{\text{live}}$ single live point runs. These are termed *threads*. These threads can be combined via bootstrapping to create new valid nested sampling runs. Any such new run contains only information contained in the original run. Combining a set of new realisations of this data and performing standard mean and variance analysis on the posteriors completes the bootstrapping algorithm and produces a measure of the variance within the original $N_{\text{live}}$ run for any chosen function of the parameters. We refer to this process as *rethreading*. Higson et al. (2017) demonstrate this algorithm in greater detail.

The H3L method can utilise this to compute standard deviation errors on $\Pr(n|\mathcal{D})$ and also on $\mathcal{P}_{\text{mod GR}}$ (where we use a bootstrapping sample size of 1000 rethreaded realisations of the parameters). This allows the H3L method to compute errors on Bayes factors and posterior odds ratios from a single run. Specifically for the GW GR test in the chapter, we can use the bootstrap to compute 1000 realisations of $\mathcal{P}_{\text{mod GR}}$ from a single run and compute the error on this directly. This parameter error will be used to compare the accuracy of the H3L method to the vanilla method. To ascertain the efficiency of a run we compare both the error and the number of likelihood calculations for the H3L and vanilla methods.

To ensure that this comparison of errors is accurate, for the toy model we ensure that the H3L method and vanilla method quoted errors are correct. To check for correctness of a single $\mathcal{P}_{\text{mod GR}}$ calculation we compare the single run result to the scattering error computed from multiple repetitions of the entire algorithm (where we repeat each $\mathcal{P}_{\text{mod GR}}$ calculation 50 times). For the Kerr model we cannot apply such a rigorous test as the likelihoods are prohibitively expensive, instead we observe for convergence of the H3L errors using chi-squared analysis whilst running PolyChord with increasingly expensive runtime parameters. For the vanilla method we extend the courtesy of assumed accuracy as PolyChord will deal well with single model likelihood runs if it can compute the more complex hyper-model likelihood of the H3L method.

For the vanilla method, a PolyChord run is needed per model to compute an evidence, whilst for the H3L run the hyper-model structure computes $\mathcal{P}_{\text{mod GR}}$ from a single run. One can envisage immediately that this can save computational cost. It is expected however that the more complex hyper-model parameter space requires more runtime to explore effectively, such that differences in efficiency cannot be simply calculated from the number of models present. An additional difference in efficiency may result from the fundamentally different propagation of errors from shrinkage factor approximations (and approximations of the iso-likelihood contour) described in section 3.2.3. We note that Higson et al. (2017) found that errors on parameter estimation were typically smaller than the errors on the evidence calculations.

We note that the H3L method computes posterior odds ratios if the model prior is included
in the likelihood code and Bayes factors if the prior on $n$ is flat. Throughout this work, the model priors are related simply by a constant multiplicative factor between GR model and mod-GR models such that the Bayes factor and posterior odds ratio are trivially related. The model prior is set flat in the likelihood implementations of this chapter and the factor for the prior adjustment is included post-processing to compute posterior odds ratios.

### 7.3 Results: sine wave toy model

A toy likelihood model is tested to show the efficiency gains in obtaining posterior odds ratios when using the H3L method over computing evidences for individual models. The model is a sine wave with a polynomial frequency component that is defined by its deformation parameters, with the ‘GR’ model being a sine wave with no time varying frequency. The 4 deformation parameters, $\epsilon_{li}=1,\ldots,4$ to be defined below, can each be turned on and off to create the $2^N - 1$ mod-GR models, as described in section 7.2.1. An injection waveform, created using a specific parameter set and model, represents our input mock-GW signal. Using the waveform likelihood, nested sampling computes the posterior odds ratio $\mathcal{P}_{\text{modGR}}$ to identify whether the GW signal deviates from the base ‘GR’ model. This captures the main features of a gravitational waveform likelihood whilst benefiting from conceptual simplicity and quick computation which facilitates thorough testing.

Section 3.1.2 shows that one can obtain $\mathcal{P}_{\text{modGR}}$ by calculating evidences to obtain Bayes factors, the ‘vanilla’ method. Section 7.2.2 shows how Bayes factors and posterior odds ratios can also be computed using the H3L method. Both of these produce the GR posterior odds ratio with an error bar. For the H3L method this is obtained from a single nested sampling run through the hyper-model likelihood, whilst for the vanilla method each model needs to be sampled over to obtain an evidence. We will refer to a ‘run’ of each method as the passing over the model space that calculates $\mathcal{P}_{\text{modGR}}$ and its error. To compare the efficiency of these two methods we quote the error on $\mathcal{P}_{\text{modGR}}$ and compare that to the number of likelihood calls that the run took to obtain the result. In order to obtain varying degrees of accuracy the POLYCHORD runtime parameter $N_{\text{live}}$ is varied. This will identify the number of likelihood calls it takes for the vanilla and H3L methods to achieve comparable accuracies, and this accuracy comparison is the principle purpose of this section. The other POLYCHORD runtime parameter is set as $N_{\text{rep}}=30$, which is shown to be sufficient at the end of section 7.3.2. Note that in later sections we calculate the error on $\mathcal{P}_{\text{modGR}}$ by repeating the H3L calculation a number of times and taking the sample standard deviation of the posterior odds ratio: the repetitions are not required for the method, but are to further validate that the H3L method and error calculation from a single run is working as expected.
7.3. Results: sine wave toy model

In practical GW applications it is sensible to assume that the GR signal is favoured, either as the physical phenomenon shows no signs of deviation or because the signal to noise ratio is too low to identify deviations. To identify efficiency gains when using H3L in this scenario, a GR injection model is tested in this toy model. For completeness a signal with $\mathcal{P}_{\text{GR}} \sim 1$ is tested to highlight the efficiency gains when a weak but present deviation is observed. Section 7.3.1 defines the likelihood and related subject matters in greater detail whilst section 7.3.2 defines the efficiency gains in detail for the GR and weak non-GR injection waveforms. Finally, section 7.3.3 discusses the implications for the more physical gravitational waveform tested in section 7.4.

7.3.1 Waveform and likelihood setup

The mock likelihood uses a sine wave of the form,

$$h(t) = A \sin(\Omega t + \varepsilon_2 t^2 + \varepsilon_3 t^3 + \varepsilon_4 t^4 + \varepsilon_5 t^5)).$$  \hspace{1cm} (7.4)

The likelihood is then constructed as discussed in section 7.2.1 via equation (7.1), where $\theta=\{A, \Omega, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5\}$ are the model parameters. In this construction, $A$ is the amplitude, $\Omega$ provides a scale to the frequency whilst the $\varepsilon_i$ parameters define any time dependence in the frequency. The signal we generate is defined over a specific time interval $T_{\text{obs}}=100$ with $10k$ samples taken to construct the waveform, unless otherwise specified.

The waveform model produces a sine wave with up to four ‘deformation parameters’. We define 16 different models using the notation $M_{dcba}$ where $dcba$ is a binary number that indicates which deformation parameter is included for the model: $d=1$ implies $\varepsilon_5$ is included, $c=1$ implies $\varepsilon_4$ is included, $b=1$ implies $\varepsilon_3$ is included, and $a=1$ implies $\varepsilon_2$ is included. Therefore the model $M_{0000}$ is the pure sign wave without deformation, $M_{1111}$ includes all possible deformation parameters and, for example, $M_{1010}$ includes deformations of the $\varepsilon_3$ and $\varepsilon_4$ forms whilst the $\varepsilon_4$ and $\varepsilon_2$ deformations are not included. When implementing the waveform and likelihood in code form we reparameterise the waveform:

$$h(t) = A \sin(\Omega \times t(1 + \varepsilon_2(t/\tau) + \varepsilon_3(t/\tau)^2 + \varepsilon_4(t/\tau)^3 + \varepsilon_5(t/\tau)^4)) \hspace{1cm} (7.5)$$

where $t \in [0, T_{\text{obs}}]$ defines the number of oscillations in the wave’s ‘signal’, $\Omega=1$ implies a period of $2\pi$, $\tau=2T_{\text{obs}}$ scales the deformation parameters, and the deformation parameters are sampled over in log-space. Numerical values of $\varepsilon_i$ parameters will be given in log-units as default.

We use two different injection waveforms to test the H3L efficiency: injection model $M_{0000}^{\text{inj}}$ with parameters $\theta=\{A=1, \Omega=1\}$ and no deformation parameters, which has a purely sinusoidal signal; and injection model $M_{0010}^{\text{inj}}$ with $\theta=\{A=1, \Omega=1, \varepsilon_3=1.9\}$ to simulate a weak deformation.
Chapter 7. H3L method acceleration in a gravitational wave test of GR

Figure 7.1: The waveforms $h(t)$ for the two injection models (top), alongside 2D slices of the log-likelihood this creates for the $M_{0000}^{\text{inj}}$ (bottom left) and $M_{0010}^{\text{inj}}$ (bottom right) injection data. The $M_{0000}^{\text{inj}}$ log-likelihood slice shows the $M_{0010}$ model likelihood varying in $\epsilon_3$ and $\Omega$, with $A = 1$. The $M_{0010}^{\text{inj}}$ log-likelihood slice shows the $M_{1110}$ model’s superfluous deformation parameters varying with the actual deformation $\epsilon_3$ off (with $A = \Omega = 1$). The waveforms create highly degenerate and multi modal likelihoods that present a difficult space to sample from.

signal in the waveform. We refer to the former as the base or ‘GR’ injection and the latter as the deformed or ‘detection’ injection.

Figure 7.1 shows the two waveforms, where the deformation from $M_{0010}^{\text{inj}}$ is clear to observe near the end of the time range. Also shown are a 2D slice of the model $M_{0010}$ log-likelihood for the base injection, as well as another 2D slice of the $M_{1110}$ model tested against the $M_{0010}^{\text{inj}}$.
injection. The slice of the $\mathcal{L}_{\text{inj}}^{0000}$ injection likelihood shows the curved and multi-modal nature of the parameter space that is explored for the deformation parameters that are not used to generate the injection data. We note that whilst $\Omega$ has a clear peak (as has $A$, though not shown), the $e_3$ parameter exhibits a ‘cut-off’ shape where the likelihood is unaffected by the parameter up to a point at which the likelihood declines sharply. This is due to a low $e_3$ value not deforming the waveform as the frequency shift is too small, whilst as soon as the effect does become significant it deviates away from the correct waveform and the likelihood declines accordingly. This is observed for any $e_i$ parameter in the base model because any deviation can only make the waveform fit worse.

We can calculate the expected Occam factor for such superfluous parameters, and therefore the posterior odds ratios of models. Each parameter will incur an Occam penalty, which we can approximate as the log-ratio of the prior and posterior width which is well defined by the cut-off point, see section 3.1.3 for a discussion of the Occam factor. Due to this cut-off, models with more parameters will have lower posterior odds ratios. However, for a cut-off value around $-4$, with a prior range of 10, we only expect an Occam factor on order of $\log(6/10) = -0.5$. The various deformation parameters for the toy model are observed to cut off around $3-4$, so that we expect $\mathcal{P}_{\text{GR}}$ to be on order $-1$ for the deformed models tested against the $\mathcal{M}_{\text{inj}}^{0000}$ injection signal described.

For the parameters superfluous to an injection model with a deformation there exists parameter degeneracy which complicates this behaviour. In figure 7.1, the 2D log-likelihood for the $e_4$ and $e_5$ parameters show that, although the injection data was generated using only a $e_3$ deformation, the two $e$ parameters can approximate the deformation well (log-likelihoods close to zero). This degeneracy means that posterior odds ratios for individual models will be non-trivial when deformations exist in the injection data, for example models without the $e_3$ parameter are not necessarily disfavoured. We note that such degeneracy is less likely to exist in the gravitational wave likelihood we test later, but that the model is still fit for purpose as a test of identifying that deformations exist in waveform data.

### 7.3.2 Efficiency gain

Figure 7.2 shows the $\mathcal{P}_{\text{GR}}^{\text{mod}}$ calculations for each of the vanilla and H3L method runs across the set of $N_{\text{live}}=[100, 200, \ldots, 1000, 1500, 2000, 2500]$ (increasing in steps of 100 up to 1000, then in steps of 500). For both injection models we see very clearly that the H3L method is more effective: for a given number of likelihood calculations it produces a smaller error on $\mathcal{P}_{\text{GR}}^{\text{mod}}$, or alternatively it takes fewer likelihood calls to produce $\mathcal{P}_{\text{GR}}^{\text{mod}}$ calculations to a given stated accuracy.
Figure 7.2: The $\mathcal{P}_{GR}^{mod}$ values obtained with various $N_{live}$, and therefore numbers of likelihood calculations, for the vanilla and H3L runs. Tables 7.1 and 7.2 show the numerical results in detail. It is clear that all runs converge well on the same value whilst the H3L method achieves smaller error bars with less likelihood calculations.
### Results: sine wave toy model

#### Table 7.1: Injection model $M_{\text{inj}}^{0000}$ results for several vanilla method and H3L method runs. The columns show the $N_{\text{live}}$ PolyChord sampling parameter used to produce a run, the number of total likelihood calculations which represents the computational cost, the posterior odds ratio $P_{\text{mod}GR}$ and its error, as well as the average error $\langle \sigma_{P_{ij}} \rangle$ on individual model posterior odds ratios. Both of the measures of errors on posterior odds ratios show a significant reduction in likelihood evaluations when using the H3L method instead of the vanilla method.

<table>
<thead>
<tr>
<th>$N_{\text{live}}$</th>
<th>Vanilla method</th>
<th></th>
<th>H3L method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L$-calls</td>
<td>$P_{\text{mod}GR}^{\text{GR}}$</td>
<td>$\langle \sigma_{P_{ij}} \rangle$</td>
<td>$L$-calls</td>
</tr>
<tr>
<td>100</td>
<td>1,696,687</td>
<td>-0.644 ± 0.231</td>
<td>0.323</td>
<td>109,580</td>
</tr>
<tr>
<td>200</td>
<td>3,366,575</td>
<td>-0.464 ± 0.164</td>
<td>0.229</td>
<td>218,022</td>
</tr>
<tr>
<td>300</td>
<td>5,047,836</td>
<td>-0.291 ± 0.136</td>
<td>0.188</td>
<td>323,612</td>
</tr>
<tr>
<td>400</td>
<td>6,854,588</td>
<td>-0.255 ± 0.118</td>
<td>0.164</td>
<td>398,944</td>
</tr>
<tr>
<td>500</td>
<td>8,378,919</td>
<td>-0.360 ± 0.105</td>
<td>0.146</td>
<td>540,622</td>
</tr>
<tr>
<td>600</td>
<td>10,070,279</td>
<td>-0.076 ± 0.098</td>
<td>0.134</td>
<td>650,034</td>
</tr>
<tr>
<td>700</td>
<td>11,693,841</td>
<td>-0.086 ± 0.091</td>
<td>0.124</td>
<td>755,582</td>
</tr>
<tr>
<td>800</td>
<td>13,354,365</td>
<td>-0.336 ± 0.083</td>
<td>0.115</td>
<td>861,864</td>
</tr>
<tr>
<td>900</td>
<td>15,075,375</td>
<td>-0.235 ± 0.079</td>
<td>0.109</td>
<td>1,031,331</td>
</tr>
<tr>
<td>1000</td>
<td>16,740,132</td>
<td>-0.297 ± 0.075</td>
<td>0.103</td>
<td>1,077,525</td>
</tr>
<tr>
<td>1500</td>
<td>24,954,613</td>
<td>-0.416 ± 0.061</td>
<td>0.084</td>
<td>1,615,063</td>
</tr>
<tr>
<td>2000</td>
<td>33,405,713</td>
<td>-0.187 ± 0.053</td>
<td>0.073</td>
<td>2,145,196</td>
</tr>
<tr>
<td>2500</td>
<td>41,785,904</td>
<td>-0.184 ± 0.048</td>
<td>0.066</td>
<td>2,683,016</td>
</tr>
</tbody>
</table>

#### Table 7.2: Injection model $M_{\text{inj}}^{0010}$ posterior odds ratios $P_{\text{mod}GR}^{\text{GR}}$, with columns as in table 7.1. The errors on posterior odds ratios again show a significant reduction in likelihood evaluations when using the H3L method instead of the vanilla method, though this time the gain is lower.

<table>
<thead>
<tr>
<th>$N_{\text{live}}$</th>
<th>Vanilla method</th>
<th></th>
<th>H3L method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L$-calls</td>
<td>$P_{\text{mod}GR}^{\text{GR}}$</td>
<td>$\langle \sigma_{P_{ij}} \rangle$</td>
<td>$L$-calls</td>
</tr>
<tr>
<td>100</td>
<td>2,678,107</td>
<td>0.887 ± 0.226</td>
<td>0.308</td>
<td>205,264</td>
</tr>
<tr>
<td>200</td>
<td>5,252,470</td>
<td>1.146 ± 0.167</td>
<td>0.221</td>
<td>411,135</td>
</tr>
<tr>
<td>300</td>
<td>7,928,647</td>
<td>1.159 ± 0.136</td>
<td>0.181</td>
<td>602,933</td>
</tr>
<tr>
<td>400</td>
<td>10,471,295</td>
<td>1.027 ± 0.118</td>
<td>0.157</td>
<td>821,168</td>
</tr>
<tr>
<td>500</td>
<td>12,980,485</td>
<td>1.083 ± 0.106</td>
<td>0.141</td>
<td>986,366</td>
</tr>
<tr>
<td>600</td>
<td>15,690,311</td>
<td>1.135 ± 0.097</td>
<td>0.128</td>
<td>1,198,236</td>
</tr>
<tr>
<td>700</td>
<td>18,295,355</td>
<td>1.075 ± 0.089</td>
<td>0.118</td>
<td>1,382,609</td>
</tr>
<tr>
<td>800</td>
<td>20,345,182</td>
<td>1.180 ± 0.083</td>
<td>0.111</td>
<td>1,621,884</td>
</tr>
<tr>
<td>900</td>
<td>23,360,394</td>
<td>1.134 ± 0.078</td>
<td>0.104</td>
<td>1,786,966</td>
</tr>
<tr>
<td>1000</td>
<td>25,880,921</td>
<td>1.167 ± 0.074</td>
<td>0.099</td>
<td>1,977,586</td>
</tr>
<tr>
<td>1500</td>
<td>38,460,250</td>
<td>1.231 ± 0.062</td>
<td>0.081</td>
<td>2,971,334</td>
</tr>
<tr>
<td>2000</td>
<td>51,838,181</td>
<td>1.127 ± 0.053</td>
<td>0.070</td>
<td>3,952,188</td>
</tr>
<tr>
<td>2500</td>
<td>64,914,620</td>
<td>1.048 ± 0.047</td>
<td>0.062</td>
<td>4,986,665</td>
</tr>
</tbody>
</table>
Table 7.1 and table 7.2 show the numerical data for the $M_{0000}$ and $M_{0010}^{inj}$ injections respectively. For each injection, when increasing $N_{\text{live}}$ we see that the number of likelihood calls required for either vanilla or H3L method increases linearly, for example comparing the $M_{0000}^{inj}$ injection $N_{\text{live}} = 100$ and $N_{\text{live}} = 1000$ vanilla runs shows it took $10 \times$ the likelihood calculations for $10 \times$ the live points. Similarly, within the results for each method and injection, we observe that more likelihood evaluations in a run produce smaller errors on $P_{\text{modGR}}$ as expected for the nested sampling algorithm. The $(\sigma_{p_{ij}})$ column shows the average error on individual posterior odds ratios produced by a run, and follows the same pattern with respect to likelihood calculations as the errors on $P_{\text{modGR}}$.

Note that the errors on $P_{\text{modGR}}$ are smaller for each run than the corresponding $(\sigma_{p_{ij}})$, where $P_{\text{modGR}}$ includes the model prior such that $\pi(M_{GR})=\sum_i \pi(M_i)|_{i\neq GR}$ whilst the individual model posterior odds ratios do not. This is as expected when considering that the individual model posterior odds ratios are essentially averaged to obtain $P_{\text{modGR}}$.

To compare efficiency between the vanilla and H3L methods we can choose a vanilla run $P_{\text{modGR}}$ error bar size, find a similar error accuracy for the H3L run and compare the ratio of likelihood calculations that the runs took. Choosing runs where the H3L error is lower ensures we do not overestimate the efficiency gain, but instead are underestimating it. Doing this for the $M_{0000}^{inj}$ injection waveform with the vanilla $N_{\text{live}} = 2500$ and H3L $N_{\text{live}} = 1500$ (which we will label (V2500, H1500)), with errors 0.048 and 0.046 respectively, produces a likelihood calls ratio of 26. Hence it took the vanilla method of running individual models 26 times as many likelihood calls to obtain that level of $P_{\text{modGR}}$ accuracy. Similarly (V1500, H1000)=23, (V1000, H600)=26.

A more holistic approach is to plot the errors against the number of likelihood calculations for each method. As errors on evidences in nested sampling are proportional to $1/\sqrt{N_{\text{live}}}$ (Skilling 2006), the ratio of likelihood calculations at comparable error accuracy may not be constant if the H3L errors exhibit alternative behaviour. Figure 7.3 shows a linear interpolation of the errors against likelihood calls for each method. The large window on the left shows the error behaviour where we observe that both methods produce similar behaviour. Note that the dashed line is the error from repetition to be discussed at the end of this section, at this point it suffices to say the dashed and solid lines agree well. The small window on the right shows the multiplicative shift between the likelihood calculations at each error accuracy, between the vanilla and H3L run. When averaged across the range of overlapping data, this gives a shift of 24.3 for the $M_{0000}^{inj}$ injection model. Hence we conclude that, for the $M_{0000}^{inj}$ injection model, the H3L method takes around 24 times fewer likelihood calculations than the vanilla method.

Understanding this efficiency gain when using the H3L method is best done by considering the number of likelihood calculations and the error on $P_{\text{modGR}}$ separately. Firstly, it is clear to
Figure 7.3: Errors on the final posterior odds ratio calculation versus the number of likelihood calculations required to achieve the stated accuracy, for the various methods. $\mathcal{M}_{i0000}^{(m)}$ results (top) and $\mathcal{M}_{i0010}^{(m)}$ results (bottom) are shown. Each graph has two panels: left shows the error accuracy achieved by the vanilla method (green crosses) and H3L method (blue plus markers). The dashed line is the computed error from 50 repetitions of the H3L method to check that the single-run quoted error is accurate. The right panel shows the average multiplicative difference (between vanilla and H3L) in the number of likelihood calculations required to achieve a given error accuracy (corresponding to a shift in the log-x-axes of the left panel).
see that the H3L method will take fewer likelihood calculations for a given \( N_{\text{live}} \) as the vanilla run needs to be repeated for each of the 16 models; the H3L method is repeated only once on the hyper-model. This gives an immediate factor of 16 reduction in the number of likelihood calculations. In practice, the H3L method adds an additional parameter and involves a more complex potentially multi-modal parameter space as a result, which means that the single H3L run will take more likelihood evaluations than a single vanilla model test. We observe this by noting that the vanilla run for a given \( N_{\text{live}} \) does not take 16 times more likelihood calculations, but for the \( M_{\text{inj}}^{\text{inj}} \) injection data runs with 1000, 1500, 2000 and 2500 live points it is 15.5, 15.5, 15.6 and 15.6, respectively. As the favoured model has no deformations the parameter space is simple and the likelihood gain is very close to 16.

This reduction in likelihood calculations does not matter, however, if the error on \( P_{\text{modGR}}^{\text{modGR}} \) is consequently higher. Each of the models in the H3L hyper-model will have to effectively share the live points: a direct evidence calculation for each model with fewer live points would suffer from a larger error value. However, the error on the posterior odds ratio calculated using the H3L method is computed as a parameter estimation problem with fundamentally different error properties. Specifically, the error on an evidence calculation in the nested sampling algorithm is an aggregation from the individual errors on the approximation of nested sampling weights each time a new live point is chosen from the likelihood, whilst the errors on a posterior depends on a different aggregation of these weights (Skilling 2006; Higson et al. 2017).

As discussed in section 3.2.1, a weight is calculated assuming that the prior mass has shrunk exponentially with respect to the step number (where a step is a new live point being chosen), but this assumption has an error associated with the shrinkage factor approximation. As the evidence sums over all these weights, the evidence error depends on absolute weight errors due to this stepping estimation. For parameter estimation, however, we compute an integral of the posterior weighted by these nested sampling weights. Specifically for the H3L method’s \( P_{\text{modGR}}^{\text{modGR}} \), or any other posterior odds ratio calculation, we take the ratio of posterior values. Now the error in our posterior odds ratio calculation depends on the relative errors of these estimated steps and shrinkage factor approximation errors from early in the run are down weighted to be negligible. Errors cancelling and taking relative errors generally produces lower errors on the parameter estimation calculation than the evidence calculation. Further detail is contained in the Skilling (2006) papers, with direct application to posteriors and further analysis found in Higson et al. (2017). For our discussion it implies that the H3L posterior odds ratio errors are distinct from the errors on the evidence. Depending on the likelihood, the posterior errors are typically lower than the evidence errors (which is shown explicitly in Higson et al. (2017)), such that the H3L method can accurately compute posterior odds ratios despite a lower \( N_{\text{live}} \) per mode in the hyper-model. In fact, for the \( M_{\text{inj}}^{\text{inj}} \) injection likelihood at a given \( N_{\text{live}} \), the
H3L method produces a more accurate posterior odds ratio calculation.

For the $\mathcal{M}_{0010}^{\text{ini}}$ injection model we obtain a smaller efficiency gain when using the H3L method: $(V1500,H2500)=7.7$, $(V1000,H1500)=8.7$, and $(V700,H900)=10.2$. With the error-likelihoods analysis in figure 7.3 for the $\mathcal{M}_{0010}^{\text{ini}}$ showing an average multiplicative likelihood calls shift of 8.8. We conclude that the H3L method takes approximately 9 times fewer likelihood calculations than the vanilla method for the $\mathcal{M}_{0010}^{\text{ini}}$ injection likelihood.

This is somewhat lower than for the $\mathcal{M}_{0000}^{\text{ini}}$ injection data. Again this is a function of the likelihood calculations and posterior errors behaviour. The likelihood calculations gain with $1000$, $1500$, $2000$ and $2500$ live points this time is only $13.1$, $12.9$, $13.1$ and $13.0$, respectively. For the $\mathcal{M}_{0010}^{\text{ini}}$ injection data, this suggests that the introduction of the $N$ model parameter dimension for the H3L hyper-model, compared to the vanilla method, creates a more complex likelihood than it did previously for the $\mathcal{M}_{0000}^{\text{ini}}$ injection data. This is as expected given that the injection data with $e_3$ introduces peaks in the $8 \mathcal{M}_{d=1,a,c,d=1,2}$ models which have that deformation parameter, as well as the degeneracies discussed in section 7.3.1 for models without it. It leads to a more pronounced distribution on $N$ with multiple modes defined across $N$ as well as the frequency parameters. A nested sampling algorithm will take more time to find a new live point within a complex iso-likelihood contour than a simple one, as observed too in the number of likelihood calculations that the vanilla method takes for the two different sets of $N_{\text{live}}$ for both injection models despite almost equivalent error accuracy. Alongside the change in likelihood calls at a given $N_{\text{live}}$, a significant change is observed in the error on $P_{\text{GR}}^{\text{mod}}$, too: for the $\mathcal{M}_{0010}^{\text{ini}}$ injection data the error on $P_{\text{GR}}^{\text{mod}}$ is now larger for the H3L method runs than for the vanilla method runs at a given $N_{\text{live}}$.

The difference in $P_{\text{GR}}^{\text{mod}}$ error accuracy between the two injection models can best be understood by considering how the H3L method explores the parameter space: the nested sampling algorithm steps exponentially through prior mass so that low likelihood regions are explored less thoroughly than higher likelihood regions, where a model with a low posterior odds ratio is such a lower likelihood region. In the $\mathcal{M}_{0010}^{\text{ini}}$ injection model case, the $\mathcal{M}_{0000}$ model in the H3L hyper model has a lower probability and would be explored less thoroughly than a more favourable model. As $P_{\text{GR}}^{\text{mod}}$ is a posterior odds ratio with respect to $\mathcal{M}_{0000}$, its error is dependent on the $\text{Post}(n = \mathcal{M}_{0000})$ error. Specifically the error of the logarithm of the posterior goes as $\delta \log(x)=\delta x/x$. Sampling the posterior space of the $\mathcal{M}_{0000}$ mode less thoroughly increases this percentage error and hence the final $P_{\text{GR}}^{\text{mod}}$ calculation will be less accurate. Figure 7.4 shows this difference between the two injection models explicitly for the H3L method run with $N_{\text{live}} = 2500$: the posterior on model $\mathcal{M}_{0000}$ is smaller for the deformation injection. Its percentage error is 0.057, which is large compared to the average percentage error for the other models of 0.030 ± 0.004. Compared to the $\mathcal{M}_{0000}^{\text{ini}}$ injection data
run this is approximately a factor of two increase: the percentage error on Post($N=M_{0000}$) is 0.032 whilst the average percentage error for the other models is $0.039 \pm 0.004$. Therefore, we observe that the H3L method has not somehow intrinsically performed a worse calculation for the $M_{0010}^{\text{inj}}$ injection data, but rather that the construction of the $P_{\text{modGR}}^\text{GR}$ posterior odds ratio itself leads to this behaviour.

The errors from vanilla and H3L methods presented are robust upon repetition. Figure 7.5 shows 50 repetitions of the H3L method for each of the runtime parameters used in figure 7.2 (top) and their averages (bottom). The averages in the bottom figure are computed such that the black error bars are the standard deviation of the 50 samples at a given set of runtime parameters (the error from repetition) and the grey error bars are the average of the 50 stated errors calculated per H3L run using the rethreading method. The blue and green error bars (with white circles) show the H3L and vanilla results in figure 7.2, respectively, for comparison. If the black error bars are larger than the grey, the H3L method is underestimating errors. We observe that the H3L errors very accurately reflect the error from repetition and conclude that they are accurate. Arguably the computed single run errors show less random variation than the errors from repetition in figure 7.3, where the grey error values are much more variable than the smooth H3L single run errors which match well with the vanilla error behaviour (a steady decline with increasing total likelihood calls). This makes sense as the errors from the H3L method single run data have been computed from 1000 rethreadings, whilst the repetition errors are only from 50 (where both rethreading and repetition produces new distinct calculations of $P_{\text{modGR}}^\text{GR}$).

For the vanilla method we observe that individual model evidences converge for each model. Additionally we computed evidences for a single vanilla run, varying all permutations of $N_{\text{live}} = \{200, 500, 1000\}$ with $N_{\text{rep}} = \{10, 20, 30, 40, 50\}$ and 50 repetitions of each runtime parameter set. The PolyChord evidences agree with the scatter from repetitions, though for $N_{\text{rep}} = 10$ there was some underestimating. From this analysis we chose $N_{\text{rep}} = 30$ for ensured accuracy.

### 7.3.3 Conclusions from the toy model

The number of likelihood calculations to achieve a given error accuracy on $P_{\text{modGR}}^\text{GR}$ is significantly lower for the H3L method than the vanilla method. For the ‘GR’ injection waveform with model $M_{0000}^{\text{inj}}$ (no deformation parameters) an average factor of 24.3 times fewer likelihood calculations is required for the H3L run. For the deformation injection waveform with model $M_{0010}^{\text{inj}}$ ($\epsilon_3 = 1.9$) an average factor of 8.8 times fewer likelihood calculations is achieved for the H3L run. The main difference in efficiency gain is because the relative errors on the GR model $M_{0000}$ posterior are increased when it is a less favoured model.
The efficiency gain when using the H3L method is conjectured to come from a mix of two factors: taking fewer likelihood calculations as all models are sampled at once, and errors in the shrinkage factor propagating favourably. The first point benefits from a simple parameter space, where we observe that computing the $M_{0010}^{\text{inj}}$ data posterior odds ratio takes proportionately more likelihood calls than the $M_{0000}^{\text{inj}}$ due to the more complicated structure. Generally though the key effect is that the parameter estimation calculation produces smaller errors. This is consistent with results in the literature (Higson et al. 2017) and is believed to be due to the difference in how errors propagate through to the evidence calculation of the vanilla method and the parameter estimation of the posterior odds ratio for the H3L method: evidences sum over shrinkage factors whilst parameter estimation takes relative shrinkage factors (relative to the evidence) and benefits from cancellations.

The relation between parameter estimation error and likelihood space is expected to be complicated, such that this result cannot be guaranteed for all likelihoods. We now investigate the Kerr waveform likelihood with deformation parameters to test whether the H3L method presents a method for accelerating the gravitational wave test of GR presented by Li et al. (2012).
Figure 7.5: Plots of the $P_{\text{modGR}}^\text{GR}$ computation versus the number of likelihood calculations for repetitions of the H3L method. The vanilla $P_{\text{modGR}}^\text{GR}$ are plotted for reference. The top plot shows all 50 H3L repetitions for each of the $N_{\text{live}}$ settings discussed in the text (blue error bars with white filled circles). Their averages for a given $N_{\text{live}}$ setting are plotted as the grey filled in region (interpolated between the average at a given $N_{\text{live}}$ setting). The bottom figure shows the results of figure 7.2 plotted alongside the averages of the top figure (per $N_{\text{live}}$ setting). Specifically, the light grey "+" marks the average of the posterior odds ratios and the average of the nested sampling quoted error for each individual run. The dark grey "+" marks the average of the posterior odds ratios and the population standard deviation error bars as a measure of the variance of results. Agreement between the dark and light grey error bars shows that the single run estimates (deduced from rethreading) are in agreement with the actual errors (deduced from repetition).
7.4 Results: Kerr waveform model test of GR

In this section we present results for a gravitational waveform model based on the Kerr metric. Deformation parameters are introduced into this model as described in section 7.2.1 to enable the model to test a wide range of deviations from the GR Kerr model. The aim of this section is to demonstrate the increased accuracy obtained when using the H3L method over the vanilla method of computing evidences. The H3L method will utilise the nested sampling rethreading technique described in section 7.2.2 and Higson et al. (2017) in order to compute errors on posterior odds ratios from a single run across the hyper-model of all models.

The Kerr waveform outputs the $h_I$ and $h_{II}$ signals that a detector can measure, as described briefly in section 2.3. There are 14 parameters that define the complete likelihood, but in our analysis we restrict ourselves to only using the parameters for the masses of the binary objects. Spins, eccentricity and parameters associated with detection timings are set constant in our models, such that the injection waveform is always matched by the model waveform in these parameters. The waveform was supplied by a collaborator, and to allow us to focus on the H3L implementation we limit our analysis here.

We note that the Kerr geometry presents a likelihood space that is known to be extremely difficult to sample well. The aim of this section is to present the usefulness of the H3L method and rethreading technique in an application that is more academically interesting than the specific toy likelihood studied in section 7.3. Therefore, although we could use the full likelihood and 14 dimensional parameter space and ensure that sufficient time is allocated to the computation using nested sampling, we proceed to fix all but 2 of the 14 parameters in the space to the default values of the injection model. Additionally, we introduce the 4 deformation parameters and the model selection parameter for the H3L method, creating at most a 7 parameter space.

7.4.1 Efficiency gain

Figure 7.6 shows the $\mathcal{F}_{\text{mod}GR}^{\text{GR}}$ calculations for each of the vanilla and H3L method runs across the set of $N_{\text{live}}=[500, 600, 700, 800, 900, 1000]$ (with $N_{\text{rep}} = 30$). The injection waveform was produced without deformation parameters, so that the results show the efficiency gain when applying the GR test to a waveform that agrees with GR, as could be done with GW150914 and GW151226. The SNR of this waveform is 10, which is a low signal strength. For this injection model we see very clearly that the H3L method is more effective: the errors are tighter and at a lower total number of likelihood calculations for the H3L method.

As the Kerr likelihood is significantly more expensive to compute than the sinusoidal toy model, a similar analysis as that of the toy model is not possible for the gravitational wave
Chapter 7. H3L method acceleration in a gravitational wave test of GR

Figure 7.6: The $P_{\text{modGR}}^{\text{GR}}$ values for the Kerr model GR injection. Results are obtained with various $N_{\text{live}}$, and therefore various numbers of likelihood calculations, for the vanilla and H3L runs. Table 7.3 shows the numerical results in detail. Although there are limited data points, the H3L method appears to converge well on the same value as the vanilla run, whilst the H3L runs take fewer likelihood calls to achieve a greater error accuracy.

<table>
<thead>
<tr>
<th>$N_{\text{live}}$</th>
<th>$\mathcal{L}$-calls</th>
<th>$P_{\text{modGR}}^{\text{GR}}$</th>
<th>$\sigma_{P_{ij}}$</th>
<th>$\langle \sigma_{P_{ij}} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>779,799</td>
<td>-0.982 ± 0.063</td>
<td>0.114</td>
<td></td>
</tr>
<tr>
<td>600</td>
<td>919,378</td>
<td>-0.737 ± 0.064</td>
<td>0.106</td>
<td></td>
</tr>
<tr>
<td>700</td>
<td>1,016,938</td>
<td>-0.717 ± 0.054</td>
<td>0.095</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>1,136,821</td>
<td>-0.784 ± 0.054</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>900</td>
<td>1,283,680</td>
<td>-0.762 ± 0.049</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>1,373,243</td>
<td>-0.838 ± 0.046</td>
<td>0.080</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3: Injection model $M_{0000}^{\text{inj}}$ results for several H3L method runs. The columns are as those used in table 7.1, but with only H3L data shown. The vanilla method run with $N_{\text{live}} = 500$ took 11,953,136 likelihood calls to compute a posterior odds ratio of $P_{\text{modGR}}^{\text{GR}} = -0.707 ± 0.138$. The average error on individual posterior odds ratios was $\langle \sigma_{P_{ij}} \rangle = 0.192$. 
likelihood. Therefore there is not as rich a source of data for analysis, and it is not possible to directly calculate a shift factor. A different approach is to use the relation that the error is proportional to $1/\sqrt{N_{\text{live}}}$. Comparing the vanilla $N_{\text{live}} = 500$ run with the H3L $N_{\text{live}} = 1000$ run we observe that the H3L error is a factor of 4.2 smaller, which would require 17.4 times as many live points for the vanilla method to achieve (due to the error scaling with live points). As the number of likelihood calculations a run takes to complete is proportional to the number of live points, we can approximate that a vanilla run with $11.9M \times 17.4 = 208.2M$ likelihood calls would achieve the same error accuracy as the H3L run. This produces an approximate shift factor of 150. Although this number seems excessively large we note that the $N_{\text{live}} = 1000$ run took almost 9 times fewer likelihood calculations to obtain an error bar that is 4.2 times tighter.

From this we generally conclude that the H3L run has potential to greatly speed up the calculation of posterior odds ratios (or Bayes factors) for gravitational wave tests of GR. We now turn to a more interesting injection waveform scenario that arose, one where single run nested sampling exploration of the posterior was not robust but detailed analysis using the H3L method and rethreading technique provided a robust analysis nonetheless.

### 7.4.2 Robustness considerations of vanilla and H3L methods

Another waveform injection was tested with the same GR injection (without parameters) but a much higher SNR. The increased SNR affects the sampling process as the peaks in the likelihood become much tighter and therefore more difficult to explore.

Figure 7.7 presents the posterior odds ratio results from a single vanilla run compared to several H3L runs with varying PolyChord runtime parameters $N_{\text{live}}$ and $N_{\text{rep}}$. Table 7.4 shows the numerical values. From the figure and table it is clear that any single H3L run quotes an error accuracy far superior to the vanilla run whilst taking significantly fewer likelihood calculations to do so.

However, it is also clear that the errors on $P_{\text{modGR}}^{\text{GR}}$ for a given H3L run are under representative of the error within the set of H3L samples. Taking, for example, the H3L run with the largest number of total likelihood calls, we observe deviations from the other H3L run values larger than $1\sigma$ for 10 out of the 14 runs present. This suggests that the errors, acquired from a single H3L run via the rethreading method, are poorly calculated. Alternatively it suggests that PolyChord is not sampling the parameter space consistently between runs, such that a single H3L run has not the information available to accurately state the error.

Figure 7.8a shows a chi-squared ($\chi^2$) analysis on the set of individual posterior odds ratio calculations, and their errors, compared to the mean of the set. We note that the 5 lowest number of likelihood runs have been omitted with the justification that their low sampling
parameters are inadequate to sample the difficult Kerr likelihood (judged by their large error and deviation). The analysis tests the hypothesis that the samples were obtained from the same distribution, which is passed if the $\chi^2$ per degree of freedom is sufficiently close to 1. $\chi^2 = \sum_i ((x_i - \langle x \rangle)^2 / \delta x_i^2)$ gives the underlying test statistic, where dividing by the number of samples (minus one, due to the mean used) adjusts it per degree of freedom. A value of one implies that, on average, the parameters deviate from the mean at the 1σ level of their stated error. A value less than one implies that the errors on a point are over-estimating the common distribution’s standard deviation, whilst a value larger than one implies that the errors on a point are under-estimating this. With a calculated value of just over 5, it is clear that the calculated $P_{\text{modGR}}^{\text{GR}}$ values from individual H3L runs are a significant underestimate.

Figure 7.8b shows another $\chi^2$ analysis but this time with the hypothesis that the set of samples was drawn from two distinct distributions on $P_{\text{modGR}}^{\text{GR}}$. This is testing whether there are two converging values in the likelihood space which PolyChord finds. To be more specific, it is testing whether the sampling algorithm is not able to accurately resolve the entire likelihood
### 7.4. Results: Kerr waveform model test of GR

Table 7.4: Injection model $\mathcal{M}_0^{\text{inj}}$ results with high SNR for several H3L method runs. The columns are as those used in table 7.1, but with only H3L data shown. The vanilla method run with $N_{\text{live}} = 300$ took 11,080,430 likelihood calls to compute a posterior odds ratio of $\mathcal{P}_{GR}^{\text{mod}} = -0.902 \pm 0.226$. The average error on individual posterior odds ratios was $\langle \sigma_{\mathcal{P}_{ij}} \rangle = 0.316$.

<table>
<thead>
<tr>
<th>$N_{\text{rep}}$ = 30</th>
<th>$N_{\text{live}}$</th>
<th>$\mathcal{L}$-calls</th>
<th>$\mathcal{P}_{GR}^{\text{mod}}$</th>
<th>$\langle \sigma_{\mathcal{P}_{ij}} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>409,056</td>
<td>-1.669 ± 0.120</td>
<td>0.338</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>601,137</td>
<td>-0.851 ± 0.105</td>
<td>0.223</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>696,774</td>
<td>-1.385 ± 0.069</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>850,525</td>
<td>-1.389 ± 0.062</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1,056,358</td>
<td>-1.066 ± 0.063</td>
<td>0.119</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>1,172,635</td>
<td>-1.219 ± 0.049</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>1,361,073</td>
<td>-0.946 ± 0.053</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>1,489,213</td>
<td>-1.090 ± 0.050</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>1,657,351</td>
<td>-1.287 ± 0.042</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1,835,173</td>
<td>-1.221 ± 0.041</td>
<td>0.085</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N_{\text{rep}}$ = 60</th>
<th>$N_{\text{live}}$</th>
<th>$\mathcal{L}$-calls</th>
<th>$\mathcal{P}_{GR}^{\text{mod}}$</th>
<th>$\langle \sigma_{\mathcal{P}_{ij}} \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>695,745</td>
<td>-1.623 ± 0.11</td>
<td>0.301</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1,110,342</td>
<td>-1.070 ± 0.08</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>1,445,774</td>
<td>-1.340 ± 0.07</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>1,720,168</td>
<td>-1.248 ± 0.06</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>2,122,617</td>
<td>-1.076 ± 0.05</td>
<td>0.121</td>
</tr>
</tbody>
</table>

Space in a single run with the search parameters used in the analysis. The $\chi^2$ values now show much better agreement for the two sub cluster sets: the errors on each cluster’s points are such that the values calculated are within expected errors of the cluster mean. The suggestion is that there is a problem with the sampling of the complex Kerr likelihood, such that on a single run the posterior odds ratio calculation is biased to either one of two values.

Note that the sample size is very low, so the analysis should not be taken as fact that there are two attractors in this problem. Instead, it highlights that there is a sampling bias at the given resolutions of $N_{\text{live}}$ and $N_{\text{rep}}$. Even from the figure 7.8a analysis it is clear that the sampling is not consistent between runs of the H3L method. Specifically, the rethreading error gives the error on the posterior odds ratio given the information contained in a single nested sampling run. If that run has missed a portion of the parameter space then the error on the posterior may not be representative for another nested sampling run, with different information contained due to different sampling. This could occur if sampling new live points is correlated, breaking the requirement of sampling uniformly from within the iso-likelihood contour, or if the resolution is not large enough to find tightly peaked modes.

In a given run with a specific $N_{\text{live}}^{i1}$, this means that the $N_{\text{live}}^{i1}$ individual threads ($t_{i1}\mid i=1,\ldots,N_{\text{live}}^{i1}$)
of that run are correlated such that as a set they do not sample the whole space. Similarly, for another run with \(N_{\text{live}}^2\), the individual threads \(t_{j2|j=1,\ldots,N_{\text{live}}^2}\) are correlated to similarly tend to an under-explored run. This occurs for each run as a systematic failure of the algorithm leads to biased exploration of the likelihood in each of its threads as they evolve together. However, there will be no correlation between a thread in the first run, and a thread in the second run, or any other run, as the cause of the biased exploration lies in how the threads of a single run are evolved together.

Just as the nested sampling algorithm permits an \(N_{\text{live}}\) run to be broken down into \(N_{\text{live}}\) single live point runs, it is possible to recreate nested sampling runs from individual threads. This is the basic premise that facilitates the rethreading method. Therefore, taking random combinations of threads from different nested sampling runs will prevent a single run from having bias across all of its threads. Figure 7.9 shows this for the Kerr posterior odds ratio data. The 10 individual runs are broken down into their 5900 individual threads (as the runs have varying \(N_{\text{live}}\)) and 10 new nested sampling runs are produced each with \(N_{\text{live}} = 590\). For each of these single H3L runs, the errors from rethreading are shown and now agree very well with the mean and the \(\chi^2\) test supports this further.

Combining them all into a single \(N_{\text{live}} = 5900\) run produces a final H3L calculation of \(P_{\text{modGR}}^{\text{GR}} = -1.155 \pm 0.017\). The error is far lower than any in table 7.4, as expected given the combination of information from all H3L results. The number of likelihood calculations for this \(N_{\text{live}} = 5900\) run is the sum of all the H3L runs (including the ones omitted from the analysis as we did not know to omit them \textit{a priori}): 18,223,941 likelihood calls. Comparing the obtained posterior odds ratio to the vanilla result of \(P_{\text{modGR}}^{\text{GR}} = -0.902 \pm 0.226\) in 11,080,430 likelihood calculations we observe that the H3L method produces an error accuracy of an order of magnitude more with comparable likelihood calculations. Knowing that the vanilla method error scales as \(1/\sqrt{N_{\text{live}}}\), a doubling of the vanilla computational expense would produce an error of approximately \(\pm 0.160\), such that the H3L method error is still almost an order of magnitude smaller. Continuing to approximate using the error scaling, it would take the vanilla method approximately 108 times as many likelihood calls to achieve this accuracy.

In general the analysis highlights the depth of information available in the H3L method approach and enhanced debugging facilities. We were able to carry out an extensive analysis of several calculations of \(P_{\text{modGR}}^{\text{GR}}\) with varying \textsc{Polychord} runtime parameters in a number of likelihood calculations that is comparable in computing time to obtaining a single \(P_{\text{modGR}}^{\text{GR}}\) calculation using the vanilla run. We cannot be sure as to whether the evidence calculations in the single vanilla run are robust and we cannot afford to rerun the computations several times. Additionally, the errors from a single run can help to identify potential exploration issues for a nested sampling run as the errors are accurate to the information that a single run contains.
Discrepancies between the error of repeated calculations and the single run errors suggests that a single run does not fully explore the space. Using this information, and the insights gained from nested sampling, enabled us to correct the problem using the existing information acquired by the runs. The H3L method can therefore provide deeper insight into the robustness of computations than the vanilla method.

### 7.5 Conclusions

The H3L method was applied to the test of general relativity using gravitational waves presented by Li et al. (2012). In combining the analysis on posterior errors by Higson et al. (2017), the H3L method is able to compute the posterior odds ratio (or similarly the Bayes factor) of the GR test significantly faster than when computing evidences using nested sampling for individual models.

Using a sinusoidal toy waveform model, the efficiency gain was on order of 24 when the injection waveform was the base model without deformations and on order 9 for a deformation injected. The difference between these two is understood to be due to the GR test construction: the relative error of the base model increases when it is less favoured in the deformed injection.
example such that the posterior odds ratio with respect to the base model is calculated less accurately. This difference is not due to any shortcoming of the H3L method as it computes individual model posteriors accurately in both injection cases.

The Kerr waveform tested showed results in line with the toy model. We observe a significant decrease in the size of error bars when using the H3L method to compute the final posterior odds ratios. The scale of this efficiency boost is difficult to quantify as the computational time of computing the Kerr likelihoods prohibits an extensive analysis similar to the toy model. Indicative results based on the approximation that nested sampling evidence calculation errors scale as $1/\sqrt{N_{\text{live}}}$ suggests a computational saving on order $\approx 100$ times fewer likelihood
7.5. Conclusions

calculations required for the H3L method.

Alongside providing efficient computation of Bayes factors and posterior odds ratios, the H3L method coupled with the parameter error estimation technique (which we refer to as rethreading (Higson et al. 2017)) provides insightful robustness checks not otherwise available to the nested sampling implementation whereby one computes evidences to obtain Bayes factors and posterior odds ratios. We highlight one scenario where the nested sampling exploration was not effective, due to the difficult shape of the Kerr likelihood (highly degenerate and multimodal with very sharp peaks due to large SNR). In this difficult scenario one can utilise the rethreading analysis to identify that a problem exists and combine the information of several runs in post-processing to obtain accurate results regardless (in practice one can never be certain an exploration is accurate). Such analysis was not attempted with the method of computing evidences as the computational time is prohibitive.
The work presented in this thesis has revolved around 3 themes: dark energy, gravitational waves and Bayesian inference. Both dark energy and gravitational wave physics are not yet well constrained and present challenging arenas for Bayesian inference due to the data limitations. This thesis hopes to have provided novel and useful analysis techniques for these fields, with the techniques becoming increasingly relevant as data quality is improved over the coming decades. Generally the analysis techniques are expected to be relevant for other fields more widely too. Additionally, constraints are presented on the time varying behaviour of the dark energy equation of state. The dark energy constraints suggest that the ΛCDM model is in very good agreement with current datasets, though generally a slight preference for phantom dark energy exists in the datasets tested, to low statistical significance.

The novel method presented in chapter 4, referred to as the H3L method, is a contribution to nested sampling Bayesian model selection theory. It requires the use of nested sampling to perform parameter estimation. Although the method can be implemented using other sampling methods, as the statistical theory is general, the implementation with MCMC techniques has not here been tested and is expected to be difficult due to the inherently multi-modal distributions with sharp jumps in likelihood values across the parameter space.

The H3L method is shown to be accurate using a simple toy model that mimics the phenomenological dark energy investigation. The method also has the potential to reduce the number of likelihood computations required to compute posterior odds ratios as low likelihood models are sampled less thoroughly. Thereafter the H3L method is shown to be useful in scenarios where the likelihood is too complicated for a direct evidence calculation using nested
sampling. The parameter estimation problem is much simpler, giving rise to the strength of the H3L method. The results of the dark energy investigation show that time varying behaviour is in good agreement with the $\Lambda$CDM model. When using the method it is noted that models with very low posterior odds ratios may not be sampled at all, though the PolyChord algorithm largely alleviates such problems by using information from likelihood samples not chosen as new live points to provide a fuller parameter estimation picture.

A detailed analysis of time varying behaviour in the equation of state of dark energy and the constraints from various datasets is presented in chapter 5. The conclusions are that the $\Lambda$CDM equation of state parameter $w = -1$ is in very good agreement with the free form reconstructions at all times. A model selection analysis also favours $\Lambda$CDM over the extension models which facilitate time varying behaviour. A slight preference for phantom dark energy ($w < -1$) is observed at redshifts $z > 2$, whilst the tightest constraints on the equation of state parameter are observed around redshift $z = 0.2$ and agree very well with $\Lambda$CDM.

An analysis of the dataset constraining power shows that Planck data only weakly constrains the dark energy equation of state evolution. Lyman-$\alpha$ data is also shown to provide very limited constraining power on its own. The tightest constraints come from supernovae and BAO datasets. The constraining power of all datasets combined peaks at around 2nats at redshift $z = 0.2$, as expected from the visual $w(z)$ plane reconstructions. Generally the overall constraining power is much lower at around 0.5nats for redshifts $z > 1$.

The dataset constraining power analysis introduces a novel formalism of the Kullback-Leibler divergence. We define the Kullback-Leibler divergence from prior to posterior for each value of $w$ at a given $z$, such that the Kullback-Leibler divergence on the equation of state parameter is a function of redshift. It facilitates an analysis that can identify constraining power across the dynamical range of the function of interest. The technique can be applied to any such functional form: for example it has been used by the CORE mission proposal to identify data constraining power on the primordial power spectrum range (CORE Collaboration et al. 2016).

A model consisting of a second dark energy component with equation of state is tested in chapter 6 to ascertain whether this alternative model fits the data better than $\Lambda$CDM. The model is hypothesised for a universe with a conformal time symmetry in the Friedmann equations. Applying Bayesian model selection to this model compared to $\Lambda$CDM we obtain a Bayes factor of $-1.12 \pm 0.53$ favouring slightly the $\Lambda$CDM model. Generally the parameters of the missing matter model are in agreement with $\Lambda$CDM.

A natural extension of this is to then allow the second dark energy component’s equation of state to vary freely. This investigates the model of a universe with two dark energy components, one of which is vacuum energy and another which we test phenomenologically by allowing the equation of state parameter to vary as a constant. Bayesian model selection compared to
$\Lambda$CDM finds a Bayes factor of $-0.43 \pm 0.43$, making this new model indistinguishable from $\Lambda$CDM in terms of model favouring. This is expected to be largely due to a lack of data allowing for a wide range of models to fit the data. The parameters of this new model are again consistent with $\Lambda$CDM and the favoured solution is to set the equation of state parameter close to $w = -1$ (reproducing the $\Lambda$CDM behaviour). This investigation represents an alternative model hypothesis test for vacuum dark energy and compliments the model-independent tests described in chapters 4 and 5.

The H3L method is applied to a model-independent gravitational wave test of general relativity in chapter 7. The principle aim is to demonstrate a gain in computational efficiency when using the H3L method over computing evidences individually to obtain posterior odds ratios. This work complements the work in chapter 4 by investigating specifically the potential efficiency gains of the H3L method. The work contributes also to the field of gravitational wave tests of GR. Due to the detection of gravitational waves and many more missions aimed at this task, the data available to test GR can be expected to grow. The H3L method facilitates doing such tests much more rapidly.

The H3L model is first applied to a toy model which mimics the waveform matching and deformation behaviour of a physical gravitational wave model. An efficiency gain was obtained of between 9 and 24 times fewer likelihood calculations to achieve a given accuracy of the Bayes factor that defines the test. The difference in efficiency gain is due to whether a deformation of GR is found, with the highest gain obtained for waveforms that do not deviate from GR; desirable as deviations are expected to be only slight (with existing detections showing no deformation).

The analysis was then applied to a Kerr binary coalescence waveform model. An efficiency gain on order 100 times fewer likelihood calculations was obtained, though the analysis could not be as thorough as the Kerr likelihood is computationally expensive. Generally it was found that the H3L method was more efficient and also provided a more complete analysis of the robustness of results. The method should be applicable to the existing LIGO tests of GR which use phenomenological deviation parameters to test GR. More generally, the H3L method is easy to implement and likely would boost many model selection investigations.

Additional work that was not completed in time for inclusion in the thesis included an investigation of the reionisation history of the universe as a function of redshift, similar to the work in chapter 5. A collaboration is ongoing for another reionisation analysis, using the $D_{KL}$ methods developed in chapter 5 to define information content. Generally the tools developed in chapter 5 can be applied to other functions of the CAMB code.

An obvious extension of the work in chapter 7 is to apply the test to the publicly available data for the GW150914 and GW151226 gravitational waves. This would involve installing
the LIGO codes for the waveforms that matched the signal and implementing the deformation analysis there. Other avenues are to test other source types, such as for space based detector sources or pulsar timing array sources, and create testing pipelines using the H3L method. Further work could be done to expand the work to include multiple gravitational wave sources or to investigate a wider range of injection models with the existing code to characterise more precisely the efficiency gains due to the H3L method.

Generally, the H3L method defined in chapter 4 could be applied to any model selection problems within astrophysics, such as identifying the number of exoplanets orbiting stars, or any model selection problems outside of astrophysics. We note that the work presented in this thesis should set the method up on a firm foundation.
Abbott B. P., et al., 2016a, Physical Review X, 6, 041015
Abbott B. P., et al., 2016c, Physical Review Letters, 116, 221101
Armendariz-Picon C., Mukhanov V., Steinhardt P., 2000, Physical Review Letters, 85, 4438
Aubourg É., et al., 2015, Phys. Rev. D, 92, 123516
Bassett B. A., Tsujikawa S., Wands D., 2006, Reviews of Modern Physics, 78, 537
158

Blair D., et al., 2015, Science China Physics, Mechanics, and Astronomy, 58, 5748
Burda P., Gregory R., Moss I. G., 2015, Journal of High Energy Physics, 8, 114
Clark A., 2013, Behavioral and Brain Sciences, 36, 181â€¢$204
Collaboration V., 2015, Classical and Quantum Gravity, 32, 024001
Das S., et al., 2011, Physical Review Letters, 107, 021301
Durrer R., Maartens R., 2008, General Relativity and Gravitation, 40, 301
Engineering, 269, 515
29
Ghosh B., 2016, Pramana, 87, 43
Phys., 5, 034
Green P. J., 1995, Biometrica, 82, 711
Hobbs G., et al. 2010, Classical and Quantum Gravity, 27, 084013
Hu W., Sawicki I., 2007, Physical Review D, 76, 064004
Kawamura S., et al. 2011, Classical and Quantum Gravity, p. 094011
Bibliography

Neal R. M., 2000, ArXiv Physics e-prints
Penrose R., 2010, Cycles of Time. Bodley Head (UK)
Pettorino V., Baccigalupi C., Perrotta F., 2005, Journal of Cosmology and Astroparticle Physics, 2005, 3
Planck Collaboration et al., 2014b, A&A, 571, A15
Planck Collaboration et al., 2016e, A&A, 594, A15
Punturo M., et al., 2010, Classical and Quantum Gravity, 27, 194002
Qian F., Zheng W., 2017, Engineering Structures, 140, 298
Ratra B., Peebles P., 1988, Phys. Rev. D, 37, 3406
Read J. I., 2014, Journal of Physics G Nuclear Physics, 41, 063101
Riles K., 2013, Progress in Particle and Nuclear Physics, 68, 1
Seikel M., Clarkson C., Smith M., 2012, J. Cosmology Astropart. Phys., 6, 036
Skilling J., 2006, Bayesian Analysis, 1, 833
Somiya K., 2012, Classical and Quantum Gravity, 29, 124007
Starobinsky A. A., 2007, JETP Letters, 86, 157
Trotta R., 2008, Contemporary Phys., 49, 71
Tsujikawa S., 2013, Class. Quant. Grav., 30, 214003
Valkenburg W., 2012, Journal of Cosmology and Astroparticle Physics, 2012, 47
Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013,
Yagi K., Stein L. C., 2016, Classical and Quantum Gravity, 33, 054001
Yunes N., Siemens X., 2013, Living Reviews in Relativity, 16, 9
Yunes N., Yagi K., Pretorius F., 2016, Phys. Rev. D, 94, 084002