

Observational Cosmology with Imperfect Data

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

of

RHODES UNIVERSITY

by

HERTZOG LANDMAN BESTER

November 2015

Abstract

We develop a formalism suitable to infer the background geometry of a general spherically symmetric dust universe directly from data on the past lightcone. This direct observational approach makes minimal assumptions about inaccessible parts of the Universe. The non-parametric and Bayesian framework we propose provides a very direct way to test one of the most fundamental underlying assumptions of concordance cosmology viz. the Copernican principle. We present the Copernicus algorithm for this purpose. By applying the algorithm to currently available data, we demonstrate that it is not yet possible to confirm or refute the validity of the Copernican principle within the proposed framework. This is followed by an investigation which aims to determine which future data will best be able to test the Copernican principle. Our results on simulated data suggest that, besides the need to improve the current data, it will be important to identify additional model independent observables for this purpose. The main difficulty with current data is their inability to constrain the value of the cosmological constant. We show how redshift drift data could be used to infer its value with minimal assumptions about the nature of the early Universe. We also discuss some alternative applications of the algorithm.

Acknowledgements

Writing my thesis has not been easy but this has largely been my own fault (the rest I owe to my dogs who seem to think that walkies and ball should take precedence over all other things). It can't however be attributed to any of the people in my life who all seem to have a lot more faith in my abilities than I do. This has been especially evident on two fronts. On a personal level, there are three people who deserve a special applause for putting up with all my nonsense. My mom and dad, Koos and Baba Bester, are first among these since they have been at it the longest. Mom and dad, since I am not sure how I could ever repay you for the seemingly endless kindness and support you have shown me over the years, I at least want to say, as I have had many opportunities to over the years, "Baie Dankie". I also owe a very large debt to my girlfriend, Morgan Jander, who took over many of our (by which I mean my) household responsibilities over the last couple of months while I battled with this thesis. Love your support (and obviously the delicious food you make me) means more to me than I like to admit. It is unlikely that my sanity would have remained completely intact (as I hope it has) had you not been there to provide the occasional distraction. Her I think I can make it up to though because I have imminent plans to make her my wife (if she will have me that is). I should also thank Johan, Hazel and Marnus for their continuous support (and occasional criticism) over the years. You guys teach me stuff they don't teach in books. The good stuff. That which we live for. So thanks! I would also like to thank Bongekile Myende who helped us out immensely during the course of this year.

I initially doubted whether I should pursue this degree. Seems like at some stage I thought I knew enough. It took four years to convince me that I know almost nothing at all. A number of people were gracious enough to make me aware of my ignorance. The main contributions came from my supervisors, Julien Larena and Nigel Bishop. It is probably not possible to be supervised from more polar opposite sides of the spectrum. Nigel you must be one of the most patient people I know. Your gentle guidance over the last four years played an instrumental role in the eventual success of this project. People who know Julien probably also know that his passion (for physics in particular but not limited to) does not betray the stereotype of his

countrymen (although I must say he hosts a truly authentic South African braai). I would not have pursued this degree if he did not come to Rhodes when he did. Julien I have really learnt a lot from working with you and I look forward to doing so in the future. All considered, I think I got pretty lucky on the supervision front. Another person with whom I have had many useful discussions is Chris Clarkson. I should also thank him for hosting me during my visits to UCT. I would also like to thank Petrus van der Walt for letting me use his code and taking the time to explain its inner workings to me. Our subsequent discussions have played no small part in the developments that follow. I can say the same thing about the discussions I have had with Denis Pollney. Thanks Denis, although it doesn't say so on paper I feel like you played at least a small supervisory role in the project. I would also like to thank Marina Seikel, Charles Hellaby and Anthony Walters for useful discussions. And last but not least, I would like to thank the usual maths and physics crew who joined me at the Rat & Parrot for many a useful discussion. I think I figured out (and subsequently forgot) most of my research problems while trying to explain it to them in broken mathemanglish. I trust that at least some of the very pertinent ideas we discussed made it into this thesis.

The financial assistance of the South African SKA Project (SKA SA) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the SKA SA. www.ska.ac.za

All the simulations performed in this thesis were run on the Rhodes maths departments GetaFix cluster.

This thesis is dedicated to Cheese

Publications list

This thesis builds on the formalism developed in the following publications:

1. **What's inside the cone? Numerically reconstructing the metric from observations** [1]. The results presented in this publication do not feature in this thesis. It can however be consulted for some of the ideas that led up to the development of the algorithm.
2. **Towards the geometry of the Universe from data** [2]. The results presented in this publication have been extended in this thesis.

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction and review	1
2 General Framework	10
2.1 Geometry, gravity and relativity	10
2.2 Matter and Energy Description	16
2.2.1 Energy momentum tensor	18
2.2.2 The fundamental 4-velocity	19
2.3 Comoving formalism	21
2.3.1 Maximally symmetric spacetimes	21
2.3.2 Spherically symmetric spacetimes	25
2.4 Observational formalism	29
3 Kinematics and observables in Spherical Symmetry	35
3.1 Geometric Optics	35
3.2 Redshift	38
3.2.1 The $v(z)$ relation	39
3.2.2 Redshift drift	40
3.3 Expansion	41
3.3.1 The longitudinal and transverse expansion rates	41
3.3.2 Cosmic chronometers	43
3.4 Distance	44
3.4.1 Flux, luminosity and distances	44
3.4.2 Standard candles	48
3.5 Density	50

3.5.1	Number counts	51
3.5.2	Density from $D(z)$ and $H_{\parallel}(z)$	52
3.6	Age of the Universe	53
4	Numerical ΛLTB solution	56
4.1	Structure of the Equations	57
4.2	Boundary regions	58
4.2.1	Inner boundary ($v \approx 0$)	58
4.2.2	Characteristic cut-off	60
4.3	Coordinate transformation	61
4.4	Numerical Implementation	64
4.4.1	Initial data and radial integration	65
4.4.2	Temporal integration	67
4.5	Diagnostics	68
4.5.1	Stability and convergence	69
4.5.2	Simulating “realistic” test data	71
4.5.3	Three level self convergence test	72
4.5.4	Λ LTB consistency relation	73
5	Statistical methodology	76
5.1	Bayesian statistics	76
5.1.1	Probability	76
5.1.2	The multivariate Gaussian distribution	77
5.1.3	Markov Chain Monte Carlo	78
5.2	Gaussian processes	85
5.2.1	Gaussian process regression	85
5.2.2	Marginalisation and the choice of covariance function	87
5.2.3	Derivative observations and sampling	89
5.3	Inference in a non-parameteric Λ LTB universe	91
5.3.1	Bayesian model	91
5.3.2	MCMC for functional spaces	94
5.3.3	The Copernicus algorithm	95
6	The metric from observations	98
6.1	Simulations and verification	98
6.2	Status of the Copernican principle	102

6.3	Forecasts	107
6.4	Discussion	112
7	Concluding remarks	116
A	Spherically symmetric models	120
A.1	Comoving dust and radiation	120
A.2	Single spherically symmetric fluid	123

1

Introduction and review

In this thesis we further explore ideas developed in [3] and adopt an observational approach to cosmology. The scientific endeavour behind cosmology is to provide a model of the large scale structure of the Universe. A complete model of the Universe should be able to explain its origin, evolution and eventual fate, while accurately accounting for astronomical observations. The peculiarity of the cosmological experiment makes this a very challenging task which, in general, requires a number of assumptions. It will be useful to review some ideas behind the scientific method, in particular its application to physics, before elaborating on this statement. At the most fundamental level, the aim of theoretical physics is to provide an objective description of the Universe as a whole, at every conceivable scale. However, since we cannot observe the Universe on all scales, we have to content ourselves with an incomplete information set. For a complete model of the Universe the laws of physics must, using the language of mathematics, be extrapolated to these inaccessible scales. A common strategy is to formulate a mathematical model for a problem, test it at accessible intermediate scales, and then deduce information about the Universe at inaccessible scales by looking at the limiting behaviour of the model. This process has been very successful in establishing the local laws of physics i.e. at spatial scales ranging, roughly, between the size of a proton and the size of the solar system, and time scales ranging between approximately 10^{-15}s - 10^{10}s . It often happens that we can confirm the compatibility of a particular model when, with advances in technology and experimental techniques, we obtain data at scales that were hitherto inaccessible. Of course, because data are never perfect, it is practically impossible to confirm that any particular model is in fact the only correct model. The more common situation is to have a number of models which give degenerate predictions within the range of scales at which data are available. A field of research can be considered to be in the precision phase when the objective becomes to improve the quality of the data until, for all practical purposes, the only surviving models give virtually

indistinguishable predictions. In accordance with Occam's razor, we then select the simplest model which gives an adequate fit to the data. This concept plays a very important role in scientific reasoning. Note that, on small spatial and short temporal scales, our experiments are usually repeatable under a wide variety of environmental circumstances and with different initial conditions. Since astrophysical observations are made from what is effectively a single space-time point (i.e. here and now), this is not possible on cosmological scales. One way to achieve a deeper level of scrutiny in cosmology is to supplement the top-down approach with a bottom-up methodology, a distinction which we will now clarify¹.

In a top-down approach we start with an intuitive description of the cosmological model (or models). A convenient way to achieve this is to invoke certain gauge choices and to parametrise the model. In particular, the parametrisation can be chosen to express prior beliefs which we might want to confront with data. It is not surprising that this is the most common approach to cosmology since it allows us to gain an intuitive understanding of the model. The bottom-up approach, on the other hand, aims to be as general and rigorous as possible. In particular, it invokes the minimum number of assumptions required to generate data for the class of models under consideration. It can therefore be employed to test the assumptions that go into model specification during the top-down approach. The observational approach to cosmology (see section 2.4) aims to achieve this by relaxing certain symmetry assumptions and working backwards from the data towards the class of allowable cosmological models. The concordance model of cosmology relies on a number of assumptions, not all of which can be tested at the same time. The aim of this thesis is to develop a framework that invokes minimal assumptions to reconstruct the background geometry of the Universe. We should therefore motivate our reasons for wanting to do so.

It is often said that, because of the drastic improvement in the availability and quality of data, cosmology has entered an era of precision science. Indeed the concordance model of cosmology is an extremely good model which can accurately account for most, if not all, observations. However, there are certain aspects of the model that we do not fully understand. Examples include the flatness, horizon, relic and cosmological constant problems (see §4 of [5] for example). Our current understanding is that the Universe has been expanding, from a very hot and dense state, for approximately 14 billion years i.e. the standard Big Bang model. Evidence² for the expansion came from observations made by Hubble and Humason between 1929 and 1931 [8, 9]. By measuring the recession velocities of galaxies as a function of distance (i.e. the

¹Note that the terms bottom-up and top-down can have different meanings depending on the context.

²Note that Georges Lemaître proposed an expanding universe model two years before Hubble in 1927 [6]. He was also the first to put forth the Big Bang hypothesis [7].

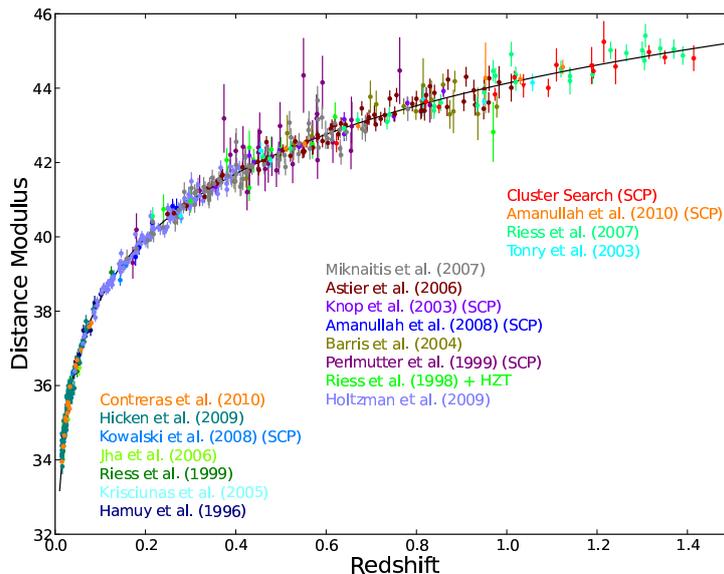


Figure 1.1: Union 2.1 distance modulus vs redshift [4]. This figure has taken from the Supernova Cosmology Project’s website <http://supernova.lbl.gov/union/>. Note that the bibliography of [4] should be consulted for the references shown in the figure.

famous Hubble diagram), they were able to conclude that the Universe must be expanding. Even though, at $H_0 \approx 500 \frac{\text{km}}{\text{sMpc}}$, their result was off by about an order of magnitude, it placed the expansion of the Universe on firm observational grounds. This led to widespread adoption of the (homogeneous and isotropic) expanding models of Friedmann, Lemaître, Robertson and Walker (FLRW), which supported the Big Bang hypothesis. For a long time it was believed that the background dynamics of the late-time Universe could be sufficiently well described by the Einstein-de Sitter (EdS) (i.e. flat, dust dominated with no cosmological constant) subclass of FLRW models. The revolution came from observations of distant supernovae nearly seventy years later. When interpreted within an FLRW context, these observations tend to indicate that the expansion of the Universe recently (since $z \approx 1$) started accelerating. The most likely interpretation of this result is that the cosmological constant must be non-zero [10, 11, 12] and dominates the dynamics of the Universe at late times. A wealth of follow up supernovae observations have since solidified this conclusion. In Figure 1.1 we show the Union 2.1 [4] supernovae data (see section 3.4). Note that, when accounting for systematics, these data alone do not place very tight constraints on the cosmological parameters. As shown in Figure 1.2, the tightest constraints currently come from combining the measured 2-point correlation functions (equivalently power spectra) of galaxies and the cosmic microwave background (CMB).

The CMB is one of the strongest observations in support of the Big Bang hypothesis. Its existence was confirmed, quite accidentally, in 1964 when Wilson and Penzias [13] noticed that

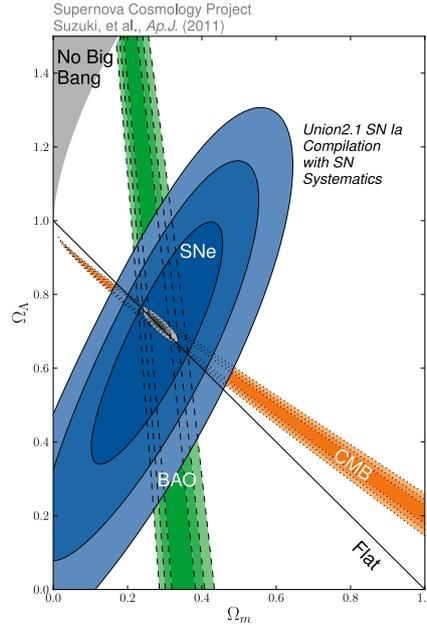


Figure 1.2: Combined constraints on Ω_m and Ω_Λ [4]. This figure has been taken from the Supernova Cosmology Project’s website <http://supernova.lbl.gov/union/>.

the effective noise temperature of the Holmdel Horn Antenna was higher than expected. The discrepancy could be explained by the existence of the CMB. The CMB is a uniform bath of electromagnetic radiation with a spectrum very near to that of a black-body. Today this radiation is detected in the microwave region and is thought to have been emitted, almost simultaneously, once the Universe had cooled (and expanded) enough for photons to decouple from baryonic matter. With this view it would necessarily permeate throughout the entire universe and, in FLRW models, be very homogeneous. Of course in practice, since our observations are made from a single vantage point, it is only possible to measure its departure from isotropy. After subtracting the dipole term, which we believe originates from our peculiar velocity with respect to the fundamental four velocity (see section 2.2), the temperature of the CMB is indeed remarkably isotropic. Observations from WMAP [14] and the Planck satellite [15] indicate that the CMB temperature varies by a maximum of 10^{-5}K over the entire sky. The temperature fluctuations are thought to correspond to tiny inhomogeneities in the gravitational potential at the time of decoupling. In the inflationary paradigm (see §8 of [5] for example) these inhomogeneities are generated by quantum fluctuations of the inflaton field and seed structure formation later on. One of the greatest successes of the standard model is accurately accounting for the spectrum of observed temperature (and polarization) fluctuations over a large range of scales. This spectrum, shown in Figure 1.3, is measured by constructing the two point correlation function of the CMB temperature. The shape of the spectrum can be

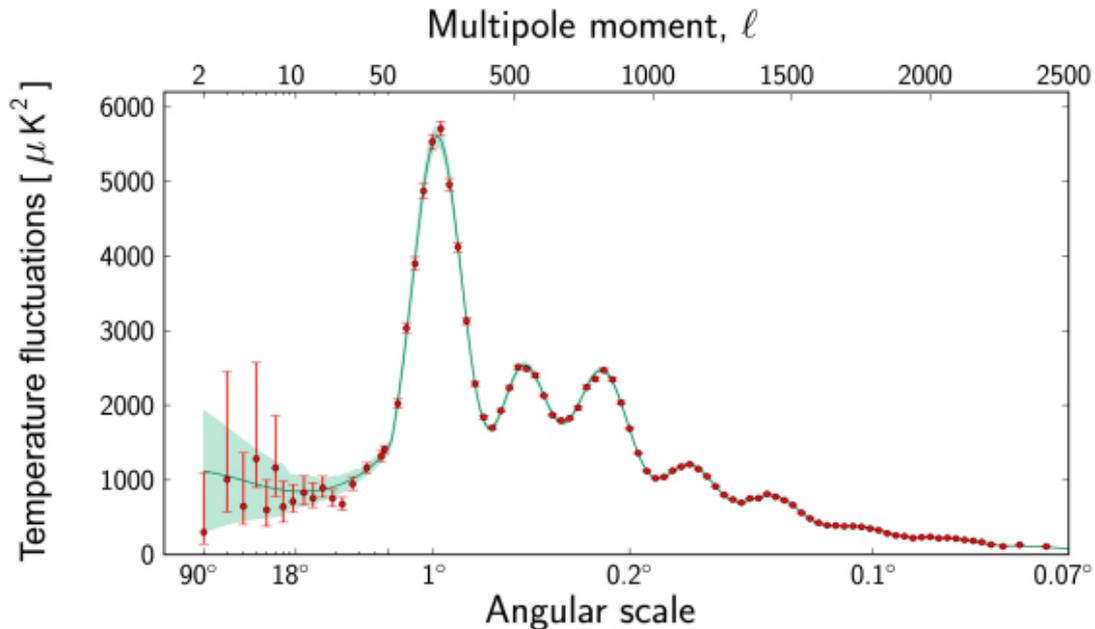


Figure 1.3: CMB power spectrum measured by the Planck collaboration [15]. This figure has been taken from the European Space Agency’s website <http://sci.esa.int/jump.cfm?oid=51555>.

predicted in linearly perturbed FLRW models and is very sensitive (albeit slightly degenerately so) to the cosmological parameters. As can be seen from Figure 1.3 the standard model fits the data remarkably well. When combined with supernovae observations these data already place impressive constraints on the cosmological parameters (see Figure 1.2 for the joint constraints on Ω_m and Ω_Λ for example). These can be tightened even further by also computing the galaxy power spectrum.

The above view supports the idea that the Universe was highly homogeneous at early times with galaxies and galaxy clusters emerging at late times as structures grow by gravitational collapse. Galaxy surveys such as the Sloan Digital Sky survey (SDSS) [16] map the distribution of galaxies about our space-time location. By averaging the number of galaxies observed behind a certain patch of sky (within a specific redshift bin), and constructing the galaxy correlation function, we can therefore also measure the power spectrum of density fluctuations. The observed power spectrum of density fluctuations can be reproduced in a perturbed FLRW model and seems to support our current understanding of structure formation (see §5 of [5] for example). A particular feature of this spectrum is the presence of baryon acoustic oscillations (BAO) (see [17] for example).

The contribution to the gravitational potential from matter is believed to be almost completely dominated by cold dark matter with baryons accounting for only about 10-15% of the total mat-

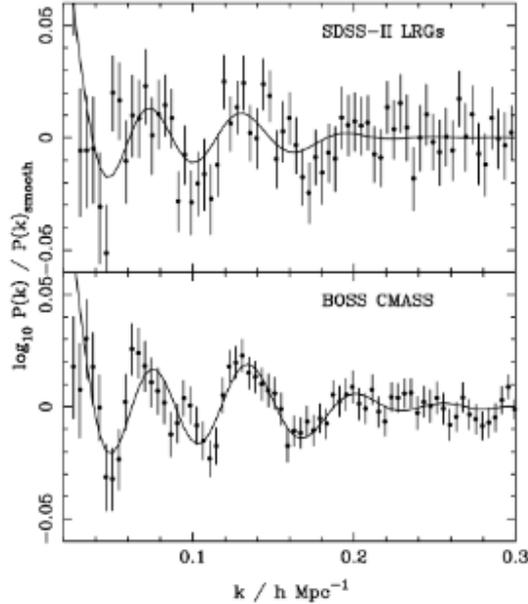


Figure 1.4: BAO power spectrum measured by the BOSS collaboration [18]. This figure has been taken from the SDSSIII website <https://www.sdss3.org/surveys/boss.php>.

ter in the Universe. Of these two only baryons are expected to interact with radiation. Up until decoupling, the photon-baryon plasma oscillates under the competing effects of pressure and gravitational collapse. These acoustic oscillations can be shown to propagate at a characteristic speed (c_s) proportional to the baryon to photon ratio. As photons decouple from the plasma they start free streaming and the pressure on baryons suddenly drops to effectively zero. The baryon acoustic waves are then frozen, and leave an imprint on the baryons at a characteristic scale proportional to the distance travelled by the sound wave up until decoupling. This feature is observed as a peak in the galaxy correlation function at comoving scale of approximately 150 Mpc (see [18] for example). Figure 1.4 shows the relative amplitude of the measured power spectrum (normalised by the power spectrum obtained in the absence of baryons). The shape of this feature is also sensitive to the values of the cosmological parameters and can therefore be used to constrain the model. As shown in Figure 1.2, impressive constraints can be derived on Ω_m and Ω_Λ by combining supernovae, CMB and BAO data (see [4] for constraints on some of the other parameters). We could keep adding more data (e.g. weak lensing [19], redshift-space distortions [20], galaxy clusters [21]) and systematically tighten the constraints until we are “certain” that we have the correct values of the cosmological parameters. Doing so we would find, with rather high certainty, that the Universe currently contains about 70% dark energy in the form of a cosmological constant (or vacuum energy), approximately 25% dark matter and only about 5% ordinary matter which adheres to the standard model of particle physics. This is

the best fit Λ CDM model. However it should be kept in mind that there are many assumptions involved in gathering these data, many of which are only valid if the background dynamics of the Universe can indeed be described by an FLRW model. These data therefore cannot be employed when using a bottom up approach to cosmology. Before getting into the details of this approach, it will be useful to review the main assumptions that go into the standard model. The two assumptions at the heart of FLRW cosmology are:

1. that gravity is sufficiently well described by general relativity (GR) on cosmological scales,
2. that on large enough scales the spatial sections of the Universe are statistically homogeneous and isotropic i.e. the cosmological principle.

When either of these assumptions are relaxed the data can be interpreted with vastly different outcomes. It is known for example that certain modified theories of gravity (see [22] for example) have the potential to account for the apparent late-time acceleration of the Universe without the need to introduce dark energy. Modified theories of gravity can also describe galactic rotation curves without the need to introduce dark matter. However, such theories must also be compatible with all other available constraints, in particular those on violations of the equivalence principle. The advent of precision cosmology has ruled out many exotic models that were hitherto considered plausible. With GR being the simplest classical theory of gravity, the scientific method also dictates that, before any of the survivors can be adopted by the scientific community, they would have to give even better predictions than GR. In this thesis we do not question its validity. It is, after all, only 100 years young. Instead our focus will be on cosmological models which relax the cosmological principle.

It is important to realise that any cosmological model which relaxes the cosmological principle will give degenerate predictions with the standard model (as long as they allow for a cosmological constant). Our faith in the standard model stems from the fact that it is the simplest model that can account for all currently available data. In this thesis we will scrutinise the symmetry assumptions of the standard model using an observational methodology. Note that the observational formalism is not completely without assumptions. These are discussed in detail in §2 (see §2.4 in particular), where we present the theoretical foundations on which the remainder of the thesis relies. However, even under these assumptions, we currently do not have access to enough model independent data to completely relax the symmetry assumptions of the standard model. Indeed, there are very few (if any) observables that can be obtained without presupposing a cosmological model. This is the topic of §3. As we discuss there, current data are barely sufficient to constrain spherically symmetric dust models of the Universe. Note that it

is crucial to allow for a cosmological constant within our proposed framework. If we are to test the foundations of concordance cosmology our framework must generalise the standard Λ CDM model. This makes it more difficult to find analytic solutions for the model, especially in observational coordinates. As a result we have to resort to numerical solutions of the Einstein field equations. This is the topic of §4, where we develop a method to specify the model by setting initial data on the observer’s past lightcone. The ability to specify initial data in the form of observables has some major advantages. Note, in particular, that it uniquely fixes the gauge freedom in our coordinates. However, since the numerical integration scheme requires smooth functions as input, it also introduces some additional challenges. Observables are inevitably reported as discrete data. These data therefore need to be smoothed before they can be used as input to the integration scheme. Although it is possible to smooth data by parametrising the input functions, we prefer not to parametrise the model because no single parametrisation, or gauge choice for that matter, can cover the entire space of models considered. We further elaborate on this in §5, the chapter dedicated to our statistical methodology. In this chapter we write the Λ LTB model in a non-parametric form and present an algorithm, dubbed the Copernicus algorithm, which can infer the geometry of the Universe directly from discrete and imperfect data. Our main results are presented in §6 where we apply the algorithm to currently available data and test the Copernican principle (CP). We conclude in §7 with a discussion on possible extensions and applications of the algorithm. There is also an appendix, § A, in which we derive some supporting results pertaining to spherically symmetric models of the Universe. Our notation is, for the most part, fairly standard. First of all we employ the Einstein summation convention in which any index appearing once as a superscript and once as a subscript implies summation over that index. Lower case Latin indices from the beginning of the alphabet (a to f) run through 0 to 3. Latin indices from the middle of the alphabet (i to p) run through 1 to 3 and upper case Latin indices only take the values 2 and 3. Employing the convention in which the metric has signature $(-, +, +, +)$, we will reserve the index zero to denote the temporal coordinate and the rest for the spatial coordinates. Indices enclosed in round braces denote symmetrisation over these indices whereas indices enclosed in square brackets imply anti-symmetrisation. With the exception of §2.3, we will use a prime to refer to partial derivatives with respect to v (i.e. the radial coordinate down the past lightcone) and an overdot to refer to the partial derivative with respect to w (i.e. the coordinate which defines the lightcones of events along the observer’s worldline). In §2.3 a prime and overdot will refer, respectively, to partial derivatives with respect to the comoving radial coordinate r and the time coordinate t . This should not cause any ambiguity, an overdot refers to temporal derivatives and a prime to

spatial derivatives, it will always be clear which coordinate system these derivatives refer to. To further distinguish between comoving and observational coordinates we will sometimes use $\stackrel{\tau}{=}$ to emphasise that a relation is only valid in comoving coordinates and $\stackrel{w}{=}$ to emphasise that a relation holds in observational coordinates. Similarly, a subscript zero can refer to either a quantity evaluated on the current past lightcone or on the current time slice of the observer. Again the meaning should be unambiguous from the context. The current past lightcone of the observer will be abbreviated by PLC0 whereas we use the notation Σ_{t_0} to refer to the current time slice. A comma or a ∂ sign followed by an index means partial derivative with respect to that index whereas a semicolon or a ∇ sign refers to the covariant derivative. Angle brackets $\langle \cdot \rangle$ are used to denote ensemble averages. Unless otherwise stated units in which $c = G = 1$ are employed.

2

General Framework

The purpose of this chapter is to introduce some fundamental concepts in cosmology. None of the material presented here is entirely original so, to be as concise as possible, the details behind most derivations are omitted. This chapter borrows extensively from [23, 24, 25, 5, 26, 27].

The general covariance of the Einstein field equations (EFE) allows for an arbitrary choice of coordinates. However, cosmology aims at modelling the large scale structure of the Universe and is best understood by choosing coordinate systems that manifest the laws of physics in an intuitive way. The question of how such a coordinate system should be chosen is therefore fundamental. Two popular coordinate choices in cosmology are reviewed. The first, comoving coordinates, is motivated by the particularly transparent 1+3 split of space-time that they result in. The covariant 1+3 Ehlers-Ellis formalism ([28],[29]) is well known by now (see for example [25, 30, 31]) and will not be exploited in great depth.

The observational approach of Ellis and co-workers (initiated in [3] with further developments in [32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42]) plays an integral role throughout this thesis. This topic is covered in §2.4. We start this chapter with a brief review of some concepts in relativity.

2.1 Geometry, gravity and relativity

Our end goal is to determine the geometry of the Universe directly from data. Since this amounts to finding the background cosmological metric of the Universe, the metric tensor seems an adequate starting point. Adopting general relativity (GR) as the theory of gravity we assume that spacetime¹ can be modelled as a smooth (pseudo-Riemannian) manifold \mathcal{M} of dimension $n = 4$. The manifold is endowed with a type $(0, 2)$ tensor field g , called the metric,

¹Note that by invoking this assumption we associate no energy to the concepts of space and time, they are simply coordinates on the manifold.

which at each point p in \mathcal{M} is a bilinear, symmetric and nondegenerate map defining the inner product between vectors in the tangent space at p viz. $T_p(\mathcal{M})$. The component view of the metric tensor g_{ab} is its representation in some arbitrary basis. A local coordinate system x^a on an open set U in \mathcal{M} induces a basis $e_a = \frac{\partial}{\partial x^a}$ in which case

$$g_{ab} = g(e_a, e_b) = g\left(\frac{\partial}{\partial x^a}, \frac{\partial}{\partial x^b}\right) \equiv \frac{\partial}{\partial x^a} \cdot \frac{\partial}{\partial x^b}. \quad (2.1)$$

This view of the metric lends interpretation to its components in any particular coordinate system. In any basis however, the metric provides the quadratic form

$$ds^2 = g_{ab}dx^a dx^b, \quad (2.2)$$

where ds^2 is an invariant measure of (spacetime) displacement in the manifold. With our specified signature (i.e. $+2$), this displacement is spacelike, timelike or null depending on whether $ds^2 > 0$, $ds^2 < 0$ or $ds^2 = 0$, respectively. The sets of timelike and non-zero null vectors each have two disjoint subsets viz. those that are future pointing and those that are past pointing. We assume that the manifold is time orientable so that the future direction can be chosen consistently over the whole manifold. Furthermore, the metric tensor provides the map between vectors and 1-forms (dual vectors) which, in the case of a non-degenerate pseudo-Riemannian metric, is 1-1 and there exists a symmetric type $(2, 0)$ tensor g^{ab} such that

$$g_{ab}g^{bc} = \delta_a^c.$$

The covariant g_{ab} and contravariant g^{ab} forms of the metric can then be used to raise and lower tensor indices.

In general, since tensors do not live in \mathcal{M} , we need some additional structure to compare tensors defined at different points in the manifold. This is done by dragging them along a curve connecting points in \mathcal{M} and gives rise to two notions of differentiation. The first, called the covariant derivative and denoted interchangeably by $A^a{}_{;b}$ or $\nabla_b A^a$, requires introducing some (in general arbitrary) non-tensorial field called the connection $\Gamma^a{}_{bc}$. The change in the vector A^a along a curve parametrised by λ is then

$$DA^a = A^a{}_{;b}dx^b + \Gamma^a{}_{bc}A^c dx^b, \quad \text{or} \quad \frac{DA^a}{d\lambda} = u^b \nabla_b A^a,$$

where $u^b = \frac{dx^b}{d\lambda}$ is the tangent vector to the curve. The covariant derivative of a general tensor may then be deduced by demanding that the operator D obeys the product rule for differentiation. A vector A^a (or indeed any tensor field) is said to be parallelly transported along a curve if its covariant derivative along the curve is zero i.e.

$$\frac{DA^a}{d\lambda} = u^b \nabla_b A^a = u^b (A^a{}_{;b} + \Gamma^a{}_{bc}A^c) = 0.$$

In GR, the principle of equivalence enforces that the connections be symmetric (i.e. torsion free $\Gamma^a_{bc} = \Gamma^a_{cb}$) and metric compatible ($\nabla_c g_{ab} = 0$) in which case the connections are called Christoffel symbols and they are explicitly given in terms of the metric by

$$\Gamma^a_{bc} = \frac{1}{2} g^{ad} (g_{db,c} + g_{dc,b} - g_{bc,d}).$$

Another useful way to compare tensors defined at different points on a manifold is the Lie derivative. Suppose the curve γ parametrised by λ passes through points p and q an infinitesimal distance, ϵ say, apart. The Lie derivative of X^a at the point p along a curve with tangent vector $\xi^a = \frac{dx^a}{d\lambda}$ is defined as

$$\mathcal{L}_\xi X^a(p) := \lim_{\epsilon \rightarrow 0} \frac{X'^a(q) - X^a(p)}{\epsilon}, \quad (2.3)$$

where $X'^a(q)$ is the value of the dragged along vector at the point q . Now if p has coordinates x^a then q has coordinates $x^a + \epsilon \xi^a$. To evaluate $X'^a(q)$ we view the dragging along operation as the coordinate transformation $x'^a = x^a - \epsilon \xi^a$. Applying the tensor transformation law shows that the definition (2.3) is equivalent to

$$\mathcal{L}_\xi X^a = X^a_{;b} \xi^b - \xi^a_{;b} X^b = X^a_{;b} \xi^b - \xi^a_{;b} X^b,$$

where in the last step we have used the fact that the manifold is torsion free. This definition extends to arbitrary tensors in a straightforward manner. Furthermore, we say that the vector ξ^a is a Killing vector if

$$\mathcal{L}_\xi g_{ab} = \xi_{a;b} + \xi_{b;a} = 0,$$

showing that $\xi_{a;b}$ is antisymmetric if ξ^a is a Killing vector. Killing vectors correspond to the continuous symmetries of the spacetime and can be used in the classification of cosmological models (see [43] for example).

In GR, gravity is not considered to be a force in the classical sense but rather the manifestation of curvature. This is captured by the Ricci identities

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) A_c = R_{dabc} A^d, \quad (2.4)$$

where A^a is an arbitrary vector field and R_{abcd} is known as the Riemann tensor. Curvature is caused by the presence of matter, the physical properties of which can be described by an energy momentum tensor (EMT) T_{ab} , to be defined shortly. This is captured by the Einstein field equations

$$G_{ab} + \Lambda g_{ab} = \kappa T_{ab}, \quad (2.5)$$

where $G_{ab} = R_{ab} - \frac{1}{2} R g_{ab}$ is the Einstein tensor defined in terms of the Ricci tensor $R_{ab} = R^c_{acb}$ and Ricci scalar $R = R^a_a$; Λ is the cosmological constant and, working in units where $G = 1 = c$,

$\kappa = 8\pi$ is the coupling constant. The EFEs can be written in the equivalent form

$$R_{ab} + \Lambda g_{ab} = \kappa(T_{ab} - \frac{1}{2}Tg_{ab}), \quad (2.6)$$

where $T = T^a_a$ is the trace of the EMT. The Bianchi identities

$$\nabla_e R_{abcd} + \nabla_d R_{abec} + \nabla_c R_{abde} = 0, \quad (2.7)$$

can be contracted (first on ac then on bd) to establish that the divergence of the Einstein tensor vanishes $\nabla_b G^{ab} = 0$. By (2.5) this also implies

$$\nabla_b T^{ab} = 0, \quad (2.8)$$

which is a statement of local conservation of energy and momentum. This four dimensional form of the theory is not particularly illuminating from an observer's point of view. We can better grasp its meaning by embedding lower dimensional submanifolds to which our intuition lends itself more easily. In particular, this allows us to construct physically transparent coordinate systems for observers in the spacetime.

One of the fundamental insights in relativity is that, in the absence of non-gravitational forces, massive particles travel along timelike geodesics while massless particles propagate along null geodesics. The inner product operation provided by the metric tensor allows for convenient computation of a spacetime's geodesics (see [23] for example). Regardless of the type of geodesics (spacelike, timelike or null) we can always find an affine parameter λ such that

$$\frac{d^2 x^a}{d\lambda^2} + \Gamma^a_{bc} \frac{dx^b}{d\lambda} \frac{dx^c}{d\lambda} = 0, \quad \text{or} \quad u^b \nabla_b u^a = 0, \quad (2.9)$$

where $u^a = \frac{dx^a}{d\lambda}$ is tangent to the geodesics. Next we consider two affinely parametrised neighbouring geodesics γ_1 and γ_2 and imagine a family of interpolating geodesics between them. The whole family of geodesics can then be described collectively with relations of the form $x^a(s, \lambda)$, where $s \in [0, 1]$ labels the geodesics (i.e. $s = 0$ at γ_1 and $s = 1$ at γ_2) and λ is an affine parametrisation (for timelike, spacelike or null geodesics). Then the vector field $u^a = \frac{\partial x^a}{\partial \lambda}$ is everywhere tangent to the geodesics satisfying $u^b \nabla_b u^a = 0$. We may consider another family of curves obtained by varying s while keeping λ fixed. These curves will have tangent vector $\xi^a = \frac{\partial x^a}{\partial s}$. We now define the connecting vector as $\xi^a(0, \lambda)$ (i.e. ξ^a evaluated on γ_1). Evaluating the Lie derivatives, and using the fact that the partial derivatives commute, we find

$$\mathcal{L}_u \xi^a = \mathcal{L}_\xi u^a = 0, \quad \Rightarrow \quad u^b \nabla_b \xi^a = \xi^b \nabla_b u^a. \quad (2.10)$$

This is an important result, when combined with the geodesic equation $u^b \nabla_b u^a = 0$, it shows that $\xi^a u_a$ is constant along γ_1 . The parametrisation can be chosen so that $\xi^a u_a = 0$ initially,

in which case they remain orthogonal along the geodesic. The acceleration of the connecting vector is described by the geodesic deviation equations

$$\frac{D^2 \xi^a}{d\lambda^2} = -R^a{}_{bcd} u^b \xi^c u^d. \quad (2.11)$$

We now turn very briefly to a concept that stems from adopting the (weak) equivalence principle in the form “free falling observers cannot distinguish the local effects of gravity (curvature) from acceleration in flat spacetime”. As a result it is always possible to find a coordinate system at any given point p of the spacetime such that

$$g_{ab}(p) = \eta_{ab}, \quad \text{and} \quad \Gamma^a{}_{bc}(p) = 0, \quad (2.12)$$

where $\eta_{ab} = \text{diag}[-1, 1, 1, 1]$ is the Minkowski metric in rectangular coordinates. This idea can be extended to an entire geodesic (see [44, 24] for example). It is therefore always possible to introduce local coordinates $x^a = (t, x^i)$ such that the connections vanish and the metric is Minkowski on the geodesic. Such coordinates are called Fermi normal coordinates (FNCs). The implication is that a free falling observer is always able to construct a tetrad basis \mathbf{e}_μ i.e. four mutually orthogonal basis vectors, the components of which can be written in terms of a local coordinate basis. Denoting the FNCs attached to an observer by x^a , we find the tetrad components $e_\mu{}^a(x^b)$ which must satisfy

$$\mathbf{e}_\mu = e_\mu{}^a(x^b) \frac{\partial}{\partial x^a} \quad \Leftrightarrow \quad \mathbf{e}_\mu(f) = e_\mu{}^a(f) \frac{\partial f}{\partial x^a}, \quad e_\mu^i = \mathbf{e}_\mu(x^i). \quad (2.13)$$

Parallely propagating this tetrad along the geodesic defines an unambiguous coordinate basis for an observer travelling along it. If the observer is subject to non-gravitational forces this coordinate basis might not be well defined throughout the history of the Universe. This can lead to subtle complications that will not be considered here (see [30] for example). We would not gain anything by it because we are going to assume that the observer is located at the origin of a spherically symmetric universe containing only barotropic fluids. This ensures geodesic motion for the observer². Supposing that the observer has 4-velocity u^a , we can define (locally for now) projection tensors that project parallel and orthogonal to u^a . The component of any vector X^a parallel to u^a is

$$X_{\parallel}^a = U_b^a X^b, \quad \text{where} \quad U_b^a := -u^a u_b, \quad U_a^a = 1, \quad U_b^a U_c^b = U_c^a. \quad (2.14)$$

The tensor that projects into the 3-dimensional instantaneous rest spaces, Σ_t say, orthogonal to u^a is defined by

$$h_{ab} = g_{ab} + u_a u_b, \quad h^a{}_b h^b{}_c = h^a{}_c, \quad h^a{}_a = 3, \quad h^a{}_b u^b = 0. \quad (2.15)$$

²Recall that the origin of a spherically symmetric universe is a special point with maximal symmetry. This implies that pressure gradients must vanish there.

The components of any vector X^a orthogonal to u^a follows from the projection

$$X_{\perp}^a = h^a_b X^b.$$

Decomposing the metric tensor into $g_{ab} = h_{ab} + U_{ab}$ we see that an observer moving with u^a perceives an arbitrary displacement $x^a \rightarrow x^a + dx^a$ as being decomposed into a time difference $\delta t^2 = (-u_a dx^a)^2$ and a spatial distance $\delta l^2 = h_{ab} dx^a dx^b$ according to

$$ds^2 = h_{ab} dx^a dx^b - (u_a dx^a)^2 = (\delta l)^2 - (\delta t)^2 \quad (2.16)$$

Attaching FNCs to an observer allows us to conclude that, at any point p along the geodesic, we can find coordinates such that the metric in Σ_t is the three dimensional Euclidean metric at p . This facilitates the introduction of spherical coordinates (r, θ, ϕ) in which the metric at p takes the form

$$\gamma_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2(\theta) \end{pmatrix}. \quad (2.17)$$

We can therefore always construct a spherical basis in Σ_t which, by parallel propagation, is well defined along the worldline of the observer. This is most reconcilable with our physical intuition, almost all of which stems from combining the weak equivalence principle with a covariant formulation of special relativity.

Recall that in special relativity the velocities of two particles moving relative to each other are related by a boost factor γ as

$$\tilde{u}^a = \gamma(u^a + v^a), \quad \text{where} \quad u_a v^a = 0, \quad v_a v^a > 0, \quad \gamma = \frac{1}{\sqrt{1 - v_a v^a}}. \quad (2.18)$$

Here v^a is the relative velocity of the frame \tilde{u}^a as seen by u^a . Decomposing v^a into its magnitude $v = \sqrt{v_a v^a}$ and direction e^a , we can write the four momentum as

$$p^a = m\tilde{u}^a = m\gamma(u^a + v e^a) = E u^a + p e^a, \quad E = \gamma m, \quad p = \gamma m v, \quad (2.19)$$

where $E = -u_a p^a$ is the energy of the particle and p is the magnitude of its three momentum $p^2 = h_{ab} p^a p^b$. Accordingly we find the familiar relation

$$p_a p^a = -m^2 = p^2 - E^2. \quad (2.20)$$

An explicit expression for e^a can now be given in terms of p^a by projecting with (2.15) and normalising

$$e^a = \frac{h^a_b p^b}{\sqrt{h_{ab} p^a p^b}}, \quad e_a e^a = 1. \quad (2.21)$$

When the particle is massless it follows from (2.20) that $E = p$. As a result the decomposition (2.19) can be written as

$$k^a = Eu^a + pe^a = E(u^a + e^a) = (-u_b k^b)(u^a + e^a), \quad \text{and} \quad e^a = \frac{h^a_b k^b}{(-u_c k^c)}, \quad (2.22)$$

where we use a different symbol k^a to distinguish the massless case. The local flatness theorem ensures that these relations always hold locally. However, since they have been written down in covariant form, the above relations must be valid in any coordinate system. We may therefore adopt them in general relativity by using the metric tensor g_{ab} (as opposed to η_{ab}).

2.2 Matter and Energy Description

We usually assume the following constituents for the matter and energy in the Universe:

- Ordinary standard model matter, the vast majority of which are photons and baryons.
- Dark matter.
- Dark energy.

There is strong evidence, both astrophysical (eg. from galactic rotation curves) and cosmological (eg. growth of density perturbations), that there must be an additional form of non-relativistic matter in the Universe that does not interact with radiation, or at best only interacts very weakly with radiation. We call it cold (since it is non-relativistic) dark (because it does not interact with radiation in any way but gravitationally) matter (CDM). It is not yet clear exactly where, or indeed if, it fits into the standard model of particle physics but a number of candidates have been identified [45]. It is also possible that modified theories of gravity (see [22] and [46] for comprehensive reviews) can account for certain observations without the need for dark matter. However, it is very difficult to propose modified theories of gravity which can compete with GR on a large range of scales while also being compatible with certain well tested physical principles (e.g. theories where the metric can couple differently to different types of matter obviously violate the equivalence principle [47]). Even when people agree that there must be some additional form of matter in the Universe, not all of them agree on what it is. A number of experiments attempting to detect it, either directly or indirectly, have been initiated (see [45] for a review).

Dark energy, be it in the form of a cosmological constant, quintessence (see [48] for example) or perfect fluid, is on even less sure footing. The classical evidence for its existence comes mainly from the Hubble diagram which, under the assumption that the Universe is spatially

homogeneous and isotropic, tends to indicate that the expansion of the Universe recently started accelerating. The acceleration can be explained if the Universe undergoes a transition from a matter dominated era to an era dominated by a cosmological constant. Allowing for a non-zero cosmological constant therefore gives a much better fit to the Hubble diagram but raises the question of what the cosmological constant actually is. This is a question which is not easy to answer (see [49, 50, 51] for example) and we will make almost no attempt to do so. Here we will simply note that, in treating space and time simply as coordinates on a manifold, GR almost takes the existence of spacetime for granted. Space (or what we call the vacuum) is not merely the absence of energy. Indeed, quantum theory associates energy to the vacuum, called the vacuum energy ρ_{vac} . Noting that Λ can be interpreted as a perfect fluid with equation of state $w = -1$ and constant energy density, it is perhaps not surprising to find a value of $\Lambda \neq 0$ when fitting a cosmological model that does not properly account for the presence of vacuum energy. The problem, known as the cosmological constant problem, is that quantum theory predicts a value of ρ_{vac} which is between 60 to 120 orders of magnitude larger than the value measured in cosmology. We will make no attempt to resolve this discrepancy. We will, however, suggest a way to obtain its value in the simplest generalisation of Λ CDM models (see §6).

It is also possible that the Universe is not accelerating at all but that the “acceleration” is a remnant of unjustified symmetry or averaging assumptions. We know for example that the real universe is not homogeneous and isotropic at all scales. On small scales, $\ll 1$ Gpc say, the real universe is clearly lumpy with multiple structures of vastly different sizes. It then becomes a question of how to relate the field equations at two different scales. Clearly we have to rely on some sort of averaging procedure (see [52] for example). More specifically, suppose we are given a metric g_{ab} and energy momentum tensor T_{ab} suitable for a particular scale. We then compute the LHS and the RHS of the field equations (2.5) separately and form the residual $\epsilon_{ab} = G_{ab} - \kappa T_{ab}$. We then repeat this procedure over a number of small regions and find the average of the residual $\langle \epsilon_{ab} \rangle$. The problem is that, because averaging does not in general commute with taking derivatives, the result will be different depending on whether we average before or after forming the residual. The fact that ϵ_{ab} is proportional to Λg_{ab} necessitates caution on our part (see [53, 54] for reviews of the backreaction and coarse graining problems). Inhomogeneities on small scales can also have other interesting consequences (see for example [55, 56]). However, we will be more concerned with the large scale structure of the Universe. Our aim in the remainder of this thesis is to test whether incorrect symmetry assumptions can give rise to “acceleration” from large scale (typically ≥ 1 Gpc) inhomogeneities. To do so we need to take a closer look at the RHS of the field equations (2.5). The following two sections

borrow extensively from the notation and ideas presented in [5] and [25].

2.2.1 Energy momentum tensor

A model describing the energy and matter content of the Universe is one of the essential ingredients in cosmology. Adopting the fluid approximation we assume that the relativistic energy and momentum of the matter field can be sufficiently well described by an energy-momentum tensor T^{ab} . This symmetric rank 2 tensor can be irreducibly split into parts parallel and perpendicular to u^a as

$$T_{ab} = \rho u_a u_b + P h_{ab} + 2q_{(a} u_{b)} + \pi_{ab}. \quad (2.23)$$

The above quantities are defined relative to u^a by

- $\rho = T_{ab} u^a u^b =$ relativistic energy density,
- $P = \frac{1}{3} h^a_b T^b_a =$ isotropic pressure,
- $q_a = -h_a^c T_{cb} u^b =$ relativistic momentum density or energy flux,
- $\pi_{ab} = h_a^c h_b^d T_{cd} - \frac{1}{3} h_{ab} h_{cd} T^{cd} =$ anisotropic pressure or stress.

Because of the high degree of observed isotropy in the Universe, a reasonable assumption in cosmology, especially at late times, is that anisotropic contributions to T_{ab} are negligible i.e. $q^a \approx 0 \approx \pi_{ab}$. The EMT then takes the form of a perfect fluid

$$T_{ab} = \rho u_a u_b + P h_{ab}. \quad (2.24)$$

To complete the specification we also need to prescribe an equation of state for the fluid. In general an equation of state specifies how the fluids thermodynamic variables (i.e. density, pressure, temperature T and entropy S) relate to each other. For a perfect fluid the equation of state can be specified by relations of the form $P = P(T, \rho)$ so that pressure depends on both temperature and density. Note that irreversible flows are still possible for fluids with equation of state $P = P(T, \rho)$ (see § 5.2.1 of [25] for example). In cosmology we generally assume that the fluid flow is reversible and adopt a barotropic equation of state $P = P(\rho)$. This assumption is usually made for simplicity; it is hard to motivate from a fluid dynamics perspective. It can be better motivated using a kinetic theory description. Assuming that the particles of the fluid are in thermodynamic equilibrium (as is expected when the reaction rate is much larger than the expansion rate of the fluid) they can be modelled as perfect Fermi-Dirac or Bose-Einstein gases. The particle distribution function can then be used to define macroscopic quantities such as the particle number density n , energy density ρ and pressure P . We can then get an idea of

what the equation of state should be for the different particle species present in the Universe. It can be shown [57, 5] that in the ultra-relativistic limit $T \gg m$, where m is the mass of the particle, both bosons and fermions satisfy $P = \frac{1}{3}\rho$, whereas in the low energy limit $T \ll m$ they satisfy $P \ll \rho$. To some extent this justifies the simple linear barotropic equations of state

$$P = w\rho \quad \text{where} \quad w = \text{const.} \quad (2.25)$$

usually employed in cosmology. The dynamics of relativistic particles such as photons is then described by $w = \frac{1}{3}$. Both CDM and (the vast majority of) baryons are expected to fit into the non-relativistic regime $w \ll 1$. Furthermore, on cosmological scales, the pressure from baryons is only felt on small scales and is expected to be completely sub-dominant on large scales at late times. As a result it is usually assumed that baryons can be lumped together with cold dark matter and described collectively as non-interacting dust (i.e. $w = 0$). This approximation is expected to be adequate at late times and on large enough scales. It will not be a good approximation on small scales. We will henceforth assume that the linear equation of state is sufficient to describe the late time background dynamics of the Universe on large enough scales. Once an equation of state has been prescribed the conservation equations (2.8) can be used to describe the dynamics of the fluid. In particular the rate of change of relativistic energy ρ along the fluid flow lines is given by

$$u^b \nabla_a T^a_b = 0. \quad (2.26)$$

This is the energy conservation equation (ECE). The equations governing conservation of momentum are found by projecting (2.8) orthogonal to the fluid flow lines

$$h^b_c \nabla_a T^a_b = 0. \quad (2.27)$$

These are known as the momentum conservation equations (MCE). In §4 we are going to use (2.26) and (2.27) to evolve the fluid variables from one PLC to the next. In order to compute them we also need to find the form of u^a .

2.2.2 The fundamental 4-velocity

So far we have defined projection tensors w.r.t. the 4-velocity of a single observer. We now ask whether this idea could be extended to a whole family of observers travelling with a unique normalised average 4-velocity u^a . The sheer diversity of particle species and states present in the real universe makes this a very difficult and scale dependent problem. It turns out that if we are able to identify an observer independent (locally Lorentz invariant) 4-current density, J^a say, then it is possible to define a unique 4-velocity in the direction of J^a if the current is

conserved i.e. $\nabla_a J^a = 0$. Suppose we define a particle number 4-current density N^a . This current will be locally Lorentz invariant because any observer will count the same number of particles regardless of their mass, charge or velocity. Next we may decompose N^a into

$$N^a = nu^a + n^a, \quad n^a u_a = 0, \quad (2.28)$$

where $n = -u_a N^a$ is the particle number density and $n^a = h^a_b N^b$ is the particle flux vector relative to u^a . If the number of particles is conserved (i.e. $\nabla_a N^a = 0$) it is possible to define a unique 4-velocity $u^a_{(p)}$ by insisting that the particle flux vector vanishes in this frame

$$N^a = n_{(p)} u^a_{(p)}. \quad (2.29)$$

This is called the particle (or Eckart) frame. The existence of the fundamental 4-velocity is a crucial ingredient in cosmology. There is an inherent averaging scale associated with its definition. Obviously we could not use stars as the particles of the fluid because we know that the number of stars is not a conserved quantity. Actually it is very difficult to determine what the averaging scale is. Our current understanding of structure formation suggests that structures start forming from (initially small) density perturbations and then over time grow into the larger structures we observe today. With this view there must be an upper limit on the possible sizes of structures, anything larger simply would not have had enough time to form by gravitational collapse. This lends weight to the idea that there is a scale at which the number of particles remains constant. Note, however, that this scale is dynamical.

We should also mention that particle number is not the only conserved quantity that can be used to identify the fundamental 4-velocity. Another useful frame, called the energy frame, can be defined by insisting that the energy flux vanishes in that frame i.e.

$$q^a_{(e)} = T_{bc} h^ab_{(e)} u^c_{(e)} = 0. \quad (2.30)$$

For perfect fluids the energy and particle frames coincide so either could be used to define a unique hydrodynamic 4-velocity (see [25] for example).

Finally, we note that, once the fundamental frame u^a has been specified its covariant derivative can be decomposed into

$$\nabla_b u_a = -\dot{u}_a u_b + \frac{1}{3} \Theta h_{ab} + \sigma_{ab} + \omega_{ab}. \quad (2.31)$$

The above quantities, known as the kinematic quantities, characterise the kinematic features of the fluid flow and are defined as follows:

- $\dot{u}^a = u^b \nabla_b u^a =$ acceleration vector;
- $\Theta = \nabla_a u^a =$ expansion scalar;

- $\sigma_{ab} = h^c_{(a} h^d_{b)} \nabla_c u_d - \frac{1}{3} \Theta h_{ab} =$ shear tensor;
- $\omega_{ab} = \nabla_{[b} u_{a]} =$ vorticity tensor.

In cosmology we usually assume that $\omega_{ab} = 0$ i.e. the spacetime is irrotational. The shear tensor σ_{ab} describes the rate of deformation of the fluid. Note that, while $\sigma_{ab} = 0$ in homogeneous and isotropic models of the Universe, it is necessarily non-zero in models which are radially inhomogeneous (see the discussion in the last paragraph of § 3.3.1). As a result it can be used to construct a consistency relation which tests for the presence of inhomogeneities. The expansion scalar Θ is the only volume changing kinematic effect. It describes the rate of expansion of the fluid flow lines. The acceleration vector describes the motion of test particles in the presence of non-gravitational forces (e.g. pressure gradients). Note that it will vanish in homogeneous and isotropic models of the Universe regardless of the matter content. It also vanishes for the spherically symmetric models used in this work but only because the fluid we consider consists only of dust.

2.3 Comoving formalism

In this section we start by reviewing the Friedmann-Lemaître-Roberson-Walker (FLRW) cosmological model. We then investigate the more general spherically symmetric cosmological solutions that will be employed to formulate tests of the Copernican principle (CP) later on.

2.3.1 Maximally symmetric spacetimes

Here we derive the form of the metric in a spatially maximally symmetric universe in comoving coordinates. This is the most cosmologically relevant example. The symmetries stem from adopting the cosmological principle i.e. the assumption that, on large enough scales, the spatial sections of the Universe are statistically homogeneous (invariant under spatial translations) and isotropic (invariant under spatial rotations).

Definition - Spatial homogeneity A spacetime is considered to be spatially homogeneous [23] if there exists a one-parameter family of spacelike hypersurfaces Σ_t foliating the spacetime in such a way that for each value of t , and any two points $p, q \in \Sigma_t$, there exists an isometry of g_{ab} which maps p into q .

By itself homogeneity is already rather restrictive. When supplemented with the additional assumption of isotropy, however, it drastically restricts the form of the metric.

Definition - Spatial isotropy Consider an observer moving with 4-velocity $u^a = \frac{dx^a}{d\tau}$ along a timelike geodesic γ parametrised by proper time τ . If at any point p on γ we take two arbitrary linearly independent unit spatial tangent vectors in Σ_t , e_1^a and e_2^a say, then isotropy implies that there exists an isometry of the metric g_{ab} which rotates e_1^a into e_2^a while leaving p and u^a at p fixed.

Clearly this can only happen when e_1^a and e_2^a are both orthogonal to u^a . In particular, since the tangent vectors are chosen arbitrarily, we can always choose any two out of the three mutually orthogonal basis vectors in Σ_t . This implies that Σ_t itself is orthogonal to u^a which in turn establishes that the spacetime is irrotational about u^a i.e. $\omega_{ab} = 0$ (this is a consequence of Frobenius' theorem, see [23] for example). Furthermore, these hypersurfaces can be defined by relations of the form $\tau = \text{const.}$ so that $u_a = -\partial_a\tau$ is a normal to Σ_t .

Comoving coordinates $x^a = (t, y^i)$ can be constructed by labelling each timelike geodesic which intersect Σ_t by y^i and insisting that they keep their labels as Σ_t traverses the congruence. This induces a three dimensional metric h_{ij} on Σ_t . Using the coordinate t to measure proper time in the frame of an observer travelling with u^a , and noting that $u_a u^a = -1 \Rightarrow dt = -u_a dx^a$, we have

$$g_{ab} = -u_a u_b + h_{ab}, \quad \Rightarrow \quad ds^2 = -dt^2 + h_{ij} dy^i dy^j.$$

This is compatible with the interpretation (2.1) of the metric tensor i.e. when Σ_t is purely transverse to u^a the basis vectors $\frac{\partial}{\partial y^i}$ are everywhere orthogonal to $\frac{\partial}{\partial t}$. Note how the metric has split into distinct temporal $d\tau^2$ and spatial $h_{ij} dy^i dy^j$ separations. This is most reconcilable with our inevitable Newtonian interpretation of the world. Determining the actual form of h_{ij} is not difficult (at least for simply connected manifolds). It can be shown [23] that the Riemann tensor constructed from h_{ij} has to satisfy

$${}^{(3)}R_{ijkl} = K h_{k[i} h_{j]l}, \quad (2.32)$$

where K can be identified as the trace of the extrinsic curvature. Homogeneity (equivalently isotropy about each $p \in \Sigma_t$) then implies that K must be constant. To find the geometry of Σ_t , we simply need to enumerate all the possible spaces of constant curvature corresponding to $K > 0, K = 0$ and $K < 0$. All positive values of K (closed universes) correspond to 3-spheres, negative values to three dimensional hyperboloids (open universes) and $K = 0$ (flat universe) is just ordinary Euclidean space. The spatial part of the metric must be conformally related to one of these geometries. The homogeneity assumption further restricts the conformal factor to only depend on the one parameter which labels the hypersurfaces i.e. time t . Thus, in comoving

coordinates $x^a = (t, y^i)$, the metric must take the form

$$ds^2 = -dt^2 + a(t)^2 d\sigma^2, \quad d\sigma^2 = \gamma_{ij} dy^i dy^j = d\chi^2 + f^2(\chi) d\Omega^2, \quad (2.33)$$

where we have introduced ‘‘spherical’’ comoving coordinates $y^i = (\chi, \theta, \phi)$ in which $d\Omega^2 = d\theta^2 + \sin^2(\theta) d\phi^2$ is the usual solid angle on the sphere and χ is the comoving distance. The geodesic deviation equation applied to the connecting vector between a congruence of radial geodesics in Σ_t can be used to determine the form of the function $f(\chi)$. By imposing the limiting behaviour $f(\chi) \rightarrow 0$ as $\chi \rightarrow 0$, it can be shown that $f(\chi)$ takes one of three forms depending on the sign of K viz.

$$f(\chi) = \begin{cases} K^{-\frac{1}{2}} \sin(\sqrt{K}\chi), & \text{if } K > 0, \\ \chi, & \text{if } K = 0, \\ (-K)^{-\frac{1}{2}} \sinh(\sqrt{-K}\chi), & \text{if } K < 0. \end{cases} \quad (2.34)$$

The symmetries have therefore reduced the ten arbitrary functions in the metric to a single function $a(t)$ and a number K . This is the form of the FLRW metric. Using a radial coordinate in which $r = f(\chi)$, this metric can also be written as

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{1}{1 - Kr^2} dr^2 + r^2 d\Omega^2 \right). \quad (2.35)$$

Note that the quantity in front of $d\Omega^2$ is called the angular diameter or area distance. Denoting this quantity by D , we find that

$$D(t, r) = a(t)r, \quad (2.36)$$

is a separable function of t and r . Substituting (2.35) into the EFE’s (2.5), and using the simple linear equation of state (2.25), gives the familiar Friedmann and acceleration equations

$$H^2 = \left(\frac{\dot{a}}{a} \right)^2 = \frac{\kappa\rho}{3} - \frac{K}{a^2} + \frac{\Lambda}{3}, \quad (2.37)$$

$$\frac{\ddot{a}}{a} = -\frac{\kappa}{6}(\rho + 3w\rho) + \frac{\Lambda}{3}. \quad (2.38)$$

These equations are more frequently expressed in terms of the dimensionless density parameters defined by

$$\Omega = \frac{\kappa\rho}{3H^2}, \quad \Omega_\Lambda = \frac{\Lambda}{3H^2}, \quad \Omega_K = -\frac{K}{H^2 a^2}. \quad (2.39)$$

With these definitions the Friedmann equation (2.37) gives the constraint

$$\Omega + \Omega_\Lambda + \Omega_K = 1, \quad \Omega = \sum_x \Omega_x = \sum_x \frac{\kappa\rho_x}{3H^2}, \quad (2.40)$$

where we have decomposed the matter term to allow for multiple fluids and x labels components with different equations of state. In this model the MCE’s (2.27) are identically satisfied and

the ECE (2.26) gives

$$\dot{\rho}_x + 3H\rho_x(1 + w_x) = 0, \quad \Rightarrow \quad \rho_x = \rho_{x0} \left(\frac{a}{a_0} \right)^{-3(1+w_x)}, \quad (2.41)$$

where w_x specifies the equation of state of the fluid with density ρ_x . Equivalently we could give this solution in terms of the dimensionless density parameter Ω_x as

$$\Omega_x = \Omega_{x0} \left(\frac{a}{a_0} \right)^{-3(1+w_x)} \left(\frac{H_0}{H} \right)^2, \quad (2.42)$$

which also casts (2.40) into the form

$$H(a)^2 = H_0^2 \left(\sum_x \Omega_{x0} \left(\frac{a}{a_0} \right)^{-3(1+w_x)} + \Omega_{K0} \left(\frac{a}{a_0} \right)^{-2} + \Omega_{\Lambda 0} \right). \quad (2.43)$$

Given current values for the parameters a_0 , H_0 , Ω_{x0} and either $\Omega_{\Lambda 0}$ or Ω_{K0} , we can therefore integrate this expression backwards in time to get the age of the Universe as

$$t_0 = \int_0^{a_0} \frac{da}{aH(a)}. \quad (2.44)$$

Note that, in FLRW models, the scale factor is related to the redshift according to $\frac{a_0}{a} = 1+z$ (see §3.2) so that the $a(z)$ relation is known from the outset. Assuming that the only contribution to Ω_x comes from dust, and normalising $a_0 = 1$, casts (2.43) into the form

$$H(z) = H_0 \sqrt{\Omega_{m0}(1+z)^3 + \Omega_{K0}(1+z)^2 + \Omega_{\Lambda 0}}, \quad (2.45)$$

where Ω_{m0} is the current day density parameter of matter (CDM + baryons). Thus, given values for H_0 , Ω_{m0} and $\Omega_{\Lambda 0}$ (and hence Ω_{K0} from (2.40)), the $H(z)$ relation follows from (2.45) and the $\rho(z)$ relation can be found using (2.41) (note $\rho_0 = 3H_0^2\Omega_{m0}/\kappa$). Furthermore, the comoving distance between an observer located at $\chi = 0$ and an object at redshift z is found by integrating along a radial null geodesic $ds^2 = 0 = d\Omega^2$. According to (2.33) the comoving distance is therefore given by

$$d\chi = \frac{dt}{a} = \frac{da}{a^2 H(a)} = -\frac{dz}{H(z)}, \quad \Rightarrow \quad \chi(z^*) = \int_0^{z^*} \frac{dz}{H(z)}, \quad (2.46)$$

and all cosmological distances (see §3.4) can be related to $\chi(z)$. Note, in particular, that the comoving radial distance, r , is given by $r(z) = f(\chi(z))$ with $f(\chi)$ defined by (2.34). Using (2.36) to relate $r(z)$ to the angular diameter distance $D(z)$ then gives

$$r(z) = (1+z)D(z).$$

It should be understood that all the redshift relations given above hold on the current past lightcone (PLC0) once we have normalised $a_0 = 1$. However, since we can choose to set the

normalisation of the scale factor at an arbitrary point in time, these relations can also be extended off the PLC0 by a simple rescaling. This can be used to establish that the relation

$$r(z) = \frac{(1+z)D(z)}{a_c}, \quad (2.47)$$

holds on an arbitrary past lightcone (PLC). Here a_c is a constant which is equal to the value of the scale factor at the observer located, on an arbitrary PLC, at $\chi = 0$. As we will see it is not possible to establish such a simple relation between $r(z)$ and $D(z)$ in spherically symmetric models of the Universe (see §4.3).

It is therefore exceedingly simple to predict the forms of the observables in FLRW models. Since these are the models under scrutiny, we will, on multiple occasions, compare our results to the predictions of one such model viz. the Λ CDM model. In Table 2.1 we therefore show the parameters corresponding to the Λ CDM model which will serve as a reference model in the remainder of this thesis. The quantity ρ_c is the current critical density of the Universe and is defined by

$$\rho_c = \frac{3H_0^2}{8\pi G}, \quad (2.48)$$

where G is Newton's gravitational constant. Note that in FLRW models $\Omega_m = \frac{\rho}{\rho_c}$. The observables corresponding to our reference model will be shown in §3 where we also show the data we currently have available for some of them. Note that the parameter values in 2.1 correspond, approximately, to the current best fit Λ CDM model (see [4] for example). The form of the observables in an Λ CDM model corresponding to the parameters listed in 2.1 are shown in Figure 3.3.

Ω_{m0}	$\Omega_{\Lambda0}$	H_0	t_0	Λ	ρ_c
0.3	0.7	70 km s ⁻¹ Mpc ⁻¹	13.5 Gyr	0.114 Gpc ⁻²	9.2×10^{-30} g cm ⁻³

Table 2.1: Parameters in the reference Λ CDM model that will be used throughout

2.3.2 Spherically symmetric spacetimes

As we have just seen, the assumption of isotropy always implies that the foliating hypersurfaces Σ_t are orthogonal to the fundamental 4-velocity u^a . A spherically symmetric universe is obviously isotropic about its centre of symmetry. The lack of homogeneity however means that the curvature K is not necessarily constant on each Σ_t . Thus it is not possible to find the form of h_{ij} simply by enumerating all geometries corresponding to different values of K . It is possible

to show that the most general form of a spherically symmetric metric takes the canonical form (see §14 of [27] for example)

$$ds^2 = -N(x, y)^2 dx^2 + 2I(x, y) dx dy + X(x, y)^2 dy^2 + R(x, y)^2 d\Omega^2, \quad (2.49)$$

where we have introduced coordinates $x^a = [x, y, \theta, \phi]$ in which $0 < \theta < \pi$, $0 < \phi < 2\pi$ are coordinates that sweep out a 2-sphere and at this stage x and y are arbitrary. Choosing x to measure proper time in the frame of a geodesic observer, and noting that isotropy implies a hypersurface orthogonal foliation, we may set the coefficient $I(x, y)$ in front of the cross term $dx dy$ to zero. In this case the function $N(t, y)$ necessarily satisfies $N(t, 0) = 1$. It cannot necessarily be set to unity on the whole of Σ_t because radial inhomogeneity could, for example, result in non-vanishing pressure gradients at $y \neq 0$. The function N measures the time lapse between neighbouring geodesics. Using $y = r$, where r is a comoving coordinate which labels the spherical shells foliating Σ_t , we can write the metric in the more familiar diagonal form

$$ds^2 = -N(t, r)^2 dt^2 + X(t, r)^2 dr^2 + R(t, r)^2 d\Omega^2. \quad (2.50)$$

This metric is suitable for a comoving description of a universe whose energy and momentum content is spatially spherically symmetric, regardless of the number of fluids present. It should be clear that in spherical symmetry there are two distinct expansion rates in Σ_t . One of them, the radial or longitudinal expansion rate H_{\parallel} , describes how the volume element changes with dr . The other one, called the transverse or perpendicular (to the radial direction) expansion rate H_{\perp} , describes how the volume element changes with $d\theta$, $d\phi$. In terms of the metric components of (2.50) these are defined by

$$H_{\parallel} = \frac{\dot{X}}{X}, \quad \text{and} \quad H_{\perp} = \frac{\dot{R}}{R}. \quad (2.51)$$

Specialising to a single barotropic fluid, the EMT has contributions from the energy density $\rho = \rho(t, r)$ and isotropic pressure $P = P(t, r)$ only i.e.

$$T_{ab} = (\rho + P)u_a u_b + h_{ab}P, \quad (2.52)$$

where u^a is the 4-velocity of the fluid. Substituting (2.50) and (2.52) into the field equations (2.5), and manipulating the equations in exactly the same way as explained in §A.2, we find that the field equations reduce to

$$\frac{\dot{R}'}{R'} = \frac{N'\dot{R}}{NR'} + \frac{\dot{X}}{X} \quad (2.53)$$

$$\kappa R^2 R' \rho = M', \quad (2.54)$$

$$-\kappa R^2 \dot{R} w \rho = \dot{M}, \quad (2.55)$$

where we have used that $P = w\rho$ and defined the function $M(t, r)$ as

$$M(t, r) = -\frac{RR'^2}{X^2} + R + \frac{R\dot{R}^2}{N^2} - \frac{1}{3}\Lambda R^3. \quad (2.56)$$

Although it is possible to find first integrals of the ECE (2.26) and MCE (2.27) (see for example [26]) in the presence of radiation, the dependence of the function M on t makes it very difficult to write down a physically transparent analogue of the Friedmann equation (2.37). In particular, it is no longer to parametrise the free functions of the model purely in terms of radial coordinate on the current time slice Σ_{t_0} . This makes the model (known as the Lemaître model) very difficult to interpret. In fact an analytic solution in the presence of Λ is not known. We will not pursue the Lemaître model any further because we are interested in describing the dynamics of the Universe at late times i.e. during the matter dominated era. Moreover, as shown in § A, there are certain subtleties behind incorporating radiation and dust into inhomogeneous models of the Universe. In particular the assumption that the fluids are comoving is no longer allowed. This has non-trivial consequences for the interpretation of observables which rely on the CMB (see the discussion in § 6.4).

The model that results from assuming that the fluid is dust (i.e. $P = 0$) is often called the Λ -Lemaître-Tolman-Bondi (ALTB) cosmological model. It is a generalisation of the Lemaître-Tolman-Bondi (LTB) [58, 59, 60] model that includes a cosmological constant. For this model the exact solution is known (see [61] for example). The absence of pressure simplifies the field equations significantly. In particular, the MCE shows that the lapse satisfies $N' = 0 \Rightarrow N = N(t)$. However, since t measures proper time, we must have $N = 1$. Physically this means that all matter follows timelike geodesics; in the absence of pressure the only “force” is gravity. Also with $P = 0$ (2.55) shows that

$$\dot{M} = 0, \quad \Rightarrow \quad M = M(r), \quad (2.57)$$

another drastic simplification. Equation (2.53) immediately gives the first integral

$$X(t, r) = g(r)R'(t, r), \quad \text{where} \quad g(r) = \frac{1}{\sqrt{1 + E(r)}}. \quad (2.58)$$

Here $g(r)$ is a function of integration which we write in terms of the dimensionless function $E(r)$ for convenience. The metric can now be written as

$$ds^2 = -dt^2 + \frac{R'^2}{1 + E(r)}dr^2 + R^2d\Omega^2, \quad E(r) = K(r)r^2, \quad (2.59)$$

where we have written $E(r)$ in terms of the function $K(r)$ which now has the dimensions of curvature. Substituting (2.58) into the expression for M and dividing through by R^3 gives

$$\left(\frac{\dot{R}}{R}\right)^2 = H_{\perp}^2 = \frac{M}{R^3} + \frac{E}{R^2} + \frac{1}{3}\Lambda \quad (2.60)$$

A close analogy with FLRW can be obtained by making the following definitions:

$$\Omega_m = \frac{M}{H_{\perp}^2 R^3}, \quad (2.61)$$

$$\Omega_{\Lambda} = \frac{\Lambda}{3H_{\perp}^2}, \quad (2.62)$$

$$\Omega_K = \frac{E}{H_{\perp}^2 R^2} = 1 - \Omega_m - \Omega_{\Lambda}, \quad (2.63)$$

where the final constraint in (2.63) follows from substituting these definitions into (2.60). Evaluating the dimensionless densities on the current time slice Σ_{t_0} , we can also cast (2.60) into a Friedmann like equation for ALTB models

$$H_{\perp}^2 = H_{\perp 0}^2 \left(\Omega_{m0} \left(\frac{R}{R_0} \right)^{-3} + \Omega_{K0} \left(\frac{R}{R_0} \right)^{-2} + \Omega_{\Lambda 0} \right), \quad (2.64)$$

where it should be understood that all quantities with a subscript 0 are evaluated on Σ_{t_0} and are functions of r only. This equation can be integrated along worldlines with constant radial coordinate, $r = r^*$ say, to get the time coordinate as a function of the metric component R

$$t(\tilde{R}) - t_B(r^*) = \frac{1}{H_{\perp 0}} \int_0^{\tilde{R}} \frac{dR}{R \sqrt{\Omega_m \left(\frac{R_0}{R} \right)^3 + \Omega_K \left(\frac{R_0}{R} \right)^2 + \Omega_{\Lambda}}}, \quad (2.65)$$

where $t_B(r)$ is a function of integration which allows for a non-simultaneous big bang (therefore called the bang time function, see [26] for example). The age of the Universe along $r = 0$ will be given by $t_0 = t(R_0(0))$ if we scale the bang time such that $t_B(0) = 0$. The integrand is an elliptic function that can be cast into one of Carlson's symmetric forms viz.

$$R_J(x, y, z, p) = \frac{3}{2} \int_0^{\infty} \frac{dt}{(t+p) \sqrt{t+x} \sqrt{t+y} \sqrt{t+z}}. \quad (2.66)$$

Here x, y and z are the (possibly complex) roots of a cubic polynomial (see (A.34) of §A) the real parts of which must be positive. The elliptic function is only defined when not more than one of x, y and z are simultaneously zero. In §A it is shown that in this model the age of the Universe at $r = 0$ is given by³

$$t_0 = \frac{1}{H_0} \frac{2R_J(-x_1, -x_2, -x_3, 1)}{3\sqrt{\Omega_m(0)}}, \quad (2.67)$$

where the x_i are the roots of

$$(x+1)^3 + \frac{\Omega_K}{\Omega_m}(x+1)^2 + \frac{\Omega_{\Lambda}}{\Omega_m} = 0. \quad (2.68)$$

³Note that this quantity is not defined in terms of the elliptic integral (2.66) when Ω_K and Ω_{Λ} are simultaneously zero. However this is just Einstein-de Sitter in which case $t_0 = \frac{2}{3H_{\perp 0}}$.

The nature of the roots provides information about the expansion history along any given worldline. This is easiest to see if we go back to equation (2.60) written in the form

$$\dot{R} = \frac{1}{\sqrt{R}} \sqrt{M + ER + \frac{\Lambda R^3}{3}}. \quad (2.69)$$

Since R , the area distance, must be positive, the positive roots of the polynomial tell us how many times \dot{R} crosses zero. Even a single positive real root along any of the worldlines signifies that \dot{R} must have changed sign along that worldline. Physically this corresponds to a universe with both expanding and collapsing regions. See the discussion in [61] for further details.

It is worth noting some differences between equation (2.64) and its Friedmann analogue (2.37). Firstly, in Λ LTB, knowledge of the functions $H_{\perp 0}(r)$, $\Omega_{m0}(r)$ and either $\Omega_{\Lambda 0}(r)$ or $\Omega_{K0}(r)$ does not allow us to write H_{\perp} directly in terms of the redshift z . The reason for this is that we do not know the metric function R as a function of the redshift z from the outset. Instead we first have to solve the null geodesic equations which can be written in terms of the redshift as

$$\frac{dt}{dz} = -\frac{1}{(1+z)H_{\parallel}} \quad \text{and} \quad \frac{dr}{dz} = \frac{\sqrt{1-Kr^2}}{(1+z)H_{\parallel}R'}. \quad (2.70)$$

The solutions to these equations can then be used to express any function in terms of the redshift and hence confront the model with observations. Also note that the density ρ and its dimensionless counterpart are not simply related as in the FLRW case (2.39). Using (2.54) gives

$$\kappa\rho = \Omega'_m H_{\perp}^2 \frac{R}{R'} + 2\Omega_m H_{\perp} H'_{\perp} \frac{R}{R'} + 3\Omega_m H_{\perp}^2. \quad (2.71)$$

Thus the quantity $\frac{\rho}{\rho_c}$ is not the same as Ω_m in Λ LTB models. However, since $R(t, 0) = 0$ but $R'(t, 0) \neq 0$, we still have that

$$\Omega_{m0}(0) = \frac{\kappa\rho_0(0)}{3H_{\perp}(0)^2}, \quad (2.72)$$

which is the same as in FLRW. Given $\rho_0(0)$ and $H_{\perp 0}(0)$, (2.72) can be used to find $\Omega_{m0}(0)$. Given also the value of Λ , we can compute t_0 using (2.67). Note that t_0 is not required to get the Λ LTB solution in observational coordinates (see §4). However, it is required to compute the coordinate transformation and if we intend to use the age of the Universe as a data point.

2.4 Observational formalism

The ideas behind the formulation of observational coordinates (OC) developed alongside those used to study gravitational radiation [62, 63]. Both approaches resulted in coordinate systems based on null geodesics. OC are based on incoming radial null geodesics and were first used in [64, 65]. They were introduced in their modern form in [3] and are also discussed in depth in [25].

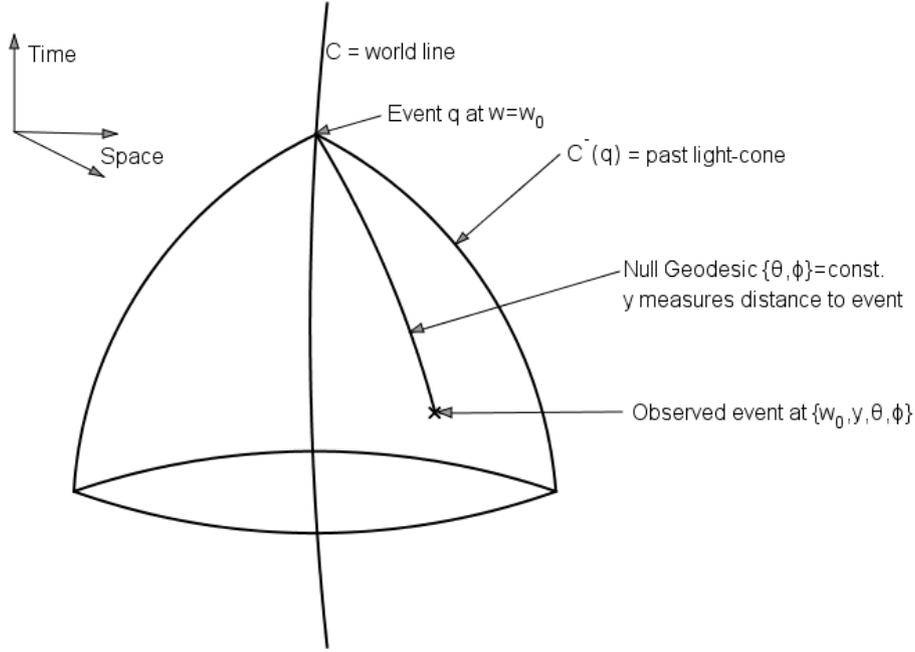


Figure 2.1: Observational coordinates: $x^a = \{w, y, \theta, \phi\}$

To define observational coordinates, we consider a congruence of null geodesics described by relations of the form $x^a(v, s)$, where v is an affine parameter for the geodesics and s labels geodesics in the congruence. Then $k^a = \frac{\partial x^a}{\partial v}$ is tangent to the geodesic, $\xi^a = \frac{\partial x^a}{\partial s}$ is a connecting vector and the following relations will hold in general (see §2.1)

$$k^a k_a = 0, \quad k^b \nabla_b k^a = 0, \quad \xi^b \nabla_b k^a = k^b \nabla_b \xi^a, \quad k^a \xi_a = 0. \quad (2.73)$$

We now embed a null hypersurface $\hat{\Sigma}$ into the spacetime with relations of the form $w(x^a) = \text{const.}$ where w is the defining scalar function of $\hat{\Sigma}$. Setting $k_a \propto \partial_a w$ ensures that the hypersurfaces are null since then $g^{ab} \partial_a w \partial_b w \propto g^{ab} k_a k_b = 0$. This reveals a strange property of null hypersurfaces viz. since k^a is both parallel and orthogonal to $\partial_a w$, the null geodesics are confined to lie within $\hat{\Sigma}$. We say the null geodesics fill or generate $\hat{\Sigma}$. Furthermore, as shown in [23] for example, it can be established that, when $k_a \propto \partial_a w$, the vector field $\mathbf{k} = \frac{\partial}{\partial v}$ must be hypersurface orthogonal. The vorticity of the congruence, defined as $\hat{\omega}_{ab} = k_{[a;b]}$, therefore vanishes. Observational coordinates are defined relative to a central observer moving with 4-velocity⁴ u^a . We assume that the observer is a fundamental observer whose worldline, denoted \mathcal{C} , is a timelike geodesic. An example might best serve to fix some of these ideas.

Consider an observer in Minkowski spacetime for which an appropriate 1+3 form of the metric

⁴Indeed the null hypersurfaces $\hat{\Sigma}$ cannot be uniquely defined by only specifying k^a , we need to specify the pair (k^a, u^a) . See [23] for example.

is

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2. \quad (2.74)$$

Information reaches the observer by means of photons which travel along a null geodesic congruence to the observer's location q , which is a caustic of the congruence. The only observable photons are those incoming in radial direction $d\Omega = 0$. As we have set $c = 1$ we know that the radial coordinate depends on the observer's time coordinate via $r = t_0 - t$, where t_0 is the proper time at the observer (recall that for an incoming radial geodesic t decreases with increasing r). This suggests using $w = t_0 = t + r$ as the defining scalar function of the null hypersurface. This also ensures that w increases towards the future. Substituting $dt = dw - dr$ into (2.74) yields

$$ds^2 = -dw^2 + 2dwdr + r^2 d\Omega^2. \quad (2.75)$$

Notice that in the observer's PLC $dw = 0$ automatically implies $ds^2 = 0$ radial null geodesics. Computing the tangent vectors we see that $k^a = \frac{dr}{dv} \delta^a_1$ since all of $dw, d\theta$ and $d\phi$ are zero along radial null geodesics in the observer's PLC. In Minkowski spacetime the radial coordinate is an affine parameter for null geodesics and we may use $v = r$. Indeed (2.75) is the form of the metric in observational coordinates for Minkowski spacetime.

Armed with a little bit of intuition we now define observational coordinates $x^a = \{w, y, \theta, \phi\}$ as follows:

w : Let w be a coordinate such that the surfaces $\{w = \text{const.}\}$ are the PLCs of events along \mathcal{C} .

In general there will still be some freedom in the definition of w which we fix by requiring that w measures proper time along the observer's worldline i.e. $w|_{\mathcal{C}} = \tau$, and that $w = w_0$ corresponds to the event q . Setting $k_a := \partial_a w$ means that the ruling geodesics of the PLC are generated by the null geodesic vector field $\mathbf{k} = \frac{\partial}{\partial w}$ where

$$k^a = \frac{dx^a}{dw}, \quad k^b \nabla_b k^a = 0, \quad k_{[a;b]} = 0 \quad \text{and} \quad k_a k^a = w_{,a} k^a = 0. \quad (2.76)$$

y : Let y be the coordinate measuring distance down the lightcone. We have complete freedom in how we choose y . Some sensible choices are the area distance D , the cosmological redshift z or an affine parameter v for the ruling null geodesics. It is also possible to choose y as one of these on the PLC0 and then insist that it is comoving with the fluid thereafter ($y_{,a} u^a = 0 \Leftrightarrow u^1 = 0$). Certain coordinate choices are more convenient for specific applications. For reasons explained in § 3.2.1, we will find it most convenient to choose $y = v$. This choice is used consistently throughout this thesis.

θ, ϕ : Let θ, ϕ be angular coordinates such that the geodesics generating the null cone are given by $\theta = \phi = \text{constant}$ in the surfaces on which w is constant ($\theta_{,a} k^a = 0 = \phi_{,a} k^a$). These are

normalized to be standard spherical coordinates at the observer. They are then paralelly propagated along \mathcal{C} .

The fundamental 4-velocity divides the set of vectors defined at any point in spacetime into two classes i.e. those that are future pointing and those that are past pointing. A vector X^a is said to be future or past pointing depending on whether $u_a X^a < 0$ or $u_a X^a > 0$, respectively. With the stated normalisation of w we have that

$$k_a u^a = \frac{\partial w}{\partial x^a} \frac{dx^a}{d\tau} \Rightarrow k_a u^a|_{\mathcal{C}} = 1, \quad (2.77)$$

which shows that the tangent vectors k^a are past pointing. Also note that the affine parameter is uniquely defined by specifying that $v = 0$ on \mathcal{C} . In observational coordinates the tangent vector satisfies

$$k_a \stackrel{w}{=} \delta^0_a, \quad k^a \stackrel{w}{=} \frac{1}{B} \delta^a_1 \quad \text{with} \quad \frac{1}{B} = \frac{dy}{dv}. \quad (2.78)$$

Thus if $y = v$ then $B = 1$ (but note in this case we cannot treat y as a comoving coordinate). The fact that k^a is hypersurface orthogonal determines some of the metric components directly

$$k_a \stackrel{w}{=} \delta^0_a = g_{ab} k^b = \frac{1}{B} g_{ab} \delta^b_1 \Rightarrow g_{1a} \stackrel{w}{=} B \delta^0_a. \quad (2.79)$$

Next introduce functions A, C_I and h_{IJ} (arbitrary for now) in the covariant form of the metric

$$g_{ab} = \begin{pmatrix} -A & B & C_2 & C_3 \\ B & 0 & 0 & 0 \\ C_2 & 0 & h_{22} & h_{23} \\ C_3 & 0 & h_{23} & h_{33} \end{pmatrix}. \quad (2.80)$$

Even though these functions were introduced somewhat arbitrarily they can be interpreted geometrically as follows. Lets start by noting that, out of the ten components of the Einstein field equations, four correspond to the complete freedom with which we can choose coordinates. Thus there remain six degrees of freedom in the metric. At first sight there seems to be seven free functions in (2.80). However, since B can be set to one with a coordinate choice, there are only six true degrees of freedom in the metric. Consider now that curves given by $v, \theta, \phi = \text{const.}$ must have tangent vector $\partial/\partial w$. Using the interpretation of the metric (2.1) we see that the direction and magnitude of these tangent vectors are given by the scalar products

$$-A = \frac{\partial}{\partial w} \cdot \frac{\partial}{\partial w}, \quad B = \frac{\partial}{\partial v} \cdot \frac{\partial}{\partial w}, \quad C_2 = \frac{\partial}{\partial \theta} \cdot \frac{\partial}{\partial w}, \quad C_3 = \frac{\partial}{\partial \phi} \cdot \frac{\partial}{\partial w}$$

The g_{0a} components of the metric therefore refer to characteristics transverse to the PLC. Since the intrinsic geometry of the PLC is described by h_{IJ} the null hypersurfaces are two dimensional.

The weak equivalence principle tells us that locally an observer will experience the Minkowski geometry. This implies the following central limit conditions [3] for the metric components of (2.80)

$$\lim_{y \rightarrow 0} A = 1, \quad \lim_{y \rightarrow 0} B = 1, \quad \lim_{y \rightarrow 0} (C_I/y^2) = 0 \quad \lim_{y \rightarrow 0} (h_{IJ}dx^I dx^J/y^2) = d\Omega^2 \quad (2.81)$$

Furthermore, the limiting behaviour of the metric components as we move away from \mathcal{C} can be established using the geodesic deviation equations (2.11) to describe the dynamics of the connecting vector ξ^a . The details can be found in [3], which uses an adaptation of the original technique used by [44] to obtain local expressions for the metric components and the matter 4-velocity in terms of the null geodesic affine parameter v . We quote the result here since it will be used in the numerical solution presented in §4. The limiting behaviour of the metric components in terms of observational coordinates with the choice $y = v$ (these are called optical coordinates in [3]) are

$$A(w, v, x^I) = 1 + A_2(w, x^I)v^2 + \mathcal{O}(v^3), \quad (2.82)$$

$$B(w, v, x^I) = 1, \quad (2.83)$$

$$C_i(w, v, x^I) = C_{I3}(w, x^I)v^3 + \mathcal{O}(v^4), \quad (2.84)$$

$$h_{IJ}(w, v, x^I) = \text{diag}(1, \sin^2(\theta))v^2 + h_{IJA}(w, x^I)v^4 + \mathcal{O}(v^5). \quad (2.85)$$

Note, in particular, that the first order term in (2.82) is zero so that $A_{,v}(0) = 0$.

Next we consider the implications of spherical symmetry. Firstly note that there is a difference between isotropy and spherical symmetry. While spherical symmetry implies isotropy, the converse is much harder to establish (see §8 of [25]). This highlights a fundamental difference between the top-down and bottom-up approaches to cosmology. The top-down approach is simpler because symmetry assumptions about the spacetime necessarily translate to symmetries in the observables. Working backwards, however, we would have to identify a maximal set of independent isotropic observables to reach the same conclusion. This highlights an unavoidable limitation of our formalism. In simple terms, we currently do not have access to enough model independent data to justify the assumption of spherical symmetry. To establish isotropy in a dust universe with Λ , an observer would have to measure isotropic area distances, number counts, bulk velocities and lensing (see §8.5 of [25]). Having admitted this limitation, we will proceed keeping in mind that the assumption of spherical symmetry is not completely justified. As we show in §3.3 the null affine parameter is indirectly observable in spherically symmetric dust universes. We therefore use $y = v$ as the coordinate measuring distance down the PLC. In this case the observational metric takes the simple form

$$ds^2 = -Adw^2 + 2dw dv + D^2 d\Omega^2. \quad (2.86)$$

Note also that $u^I = 0$ in spherical symmetry otherwise we could define preferred motions on the sky. The explicit form of u^a in observational coordinates is given by (3.16) in § 3.2 where we write it in terms of the cosmological redshift z . The field equations in observational coordinates for a spherically symmetric dust universe will be given in §4.1 where we also present a method to solve them. For now let us note, and this will be a reoccurring theme throughout, that spherically symmetric dust universes have two functional degrees of freedom, allowing for a cosmological constant introduces an additional free parameter viz. Λ .

3

Kinematics and observables in Spherical Symmetry

There is only a finite set of observables available in cosmology. Generating data for these observables relies on the ability to model astrophysical processes. In this chapter we aim to bridge the gap between these astrophysical models and the theoretical expressions for the observables. Understanding the data gathering process, which usually necessitates assuming a cosmological model, is crucial for the validity of our conclusions since it would be circular to presuppose a perturbed FLRW model. We therefore aim to be as explicit as possible about the assumptions that go into the data gathering process. Since we observe the Universe through electromagnetic radiation we start, borrowing extensively from the presentation given in [25] (see their §5.5 and §7.1 in particular), with the geometric optics approximation.

3.1 Geometric Optics

Here we apply the geometric optics limit to the electromagnetic field and assume that light propagation in a curved spacetime can be sufficiently well modelled using ray optics. The electromagnetic field is most conveniently described in terms of an antisymmetric type $(0, 2)$ tensor $F_{ab} = F_{[ab]}$ called the electromagnetic field tensor. We can achieve a 1+3 split of the field relative to u^a in terms of its electric E^a and magnetic B^a components

$$F_{ab} = 2u_{[a}E_{b]} + \eta_{abc}B^c \quad \text{where} \quad E_a = F_{ab}u^b, \quad B_a = \frac{1}{2}\eta_{abc}F^{cd}. \quad (3.1)$$

Charged particles contribute to the electromagnetic field giving rise to a four current J^a . An observer moving with four velocity u^a will measure this current as

$$J^a = \mu u^a + j^a, \quad j_a u^a = 0, \quad \mu = -u_a J^a, \quad j^a = h^a_b J^b, \quad (3.2)$$

where we interpret μ as the charge density and j^a as the three current associated with the field. Maxwell's equations can then be succinctly expressed as

$$\nabla_b F^{ab} = J^a, \quad \nabla_{[a} F_{bc]} = 0. \quad (3.3)$$

Defining $F_{ab} = \nabla_a A_b - \nabla_b A_a$ in terms of the four potential $A = [\phi, \mathbf{A}]$, where ϕ is the scalar potential sourcing the electric field and \mathbf{A} the the vector potential sourcing the magnetic field, ensures that F_{ab} is anti-symmetric. Plugging this definition into the second of (3.3) shows that it is automatically satisfied. Adding the gradient of an arbitrary scalar function of position to A_a i.e.

$$A_a \rightarrow A_a + \partial_a f(x^i), \quad (3.4)$$

has no effect of the form of F_{ab} since

$$F_{ab} = \nabla_a(A_b + \partial_b f) - \nabla_b(A_a + \partial_a f) = \nabla_a A_b - \nabla_b A_a + (\nabla_a \partial_b f - \nabla_b \partial_a f), \quad (3.5)$$

and the term in braces is zero. We now use this freedom to impose the Lorentz gauge $\nabla_a A^a = 0$. The gauge freedom (3.4) still exists provided that f is a harmonic function i.e. $\nabla_a \nabla^a f = 0$. The first of Maxwell's equations can now be written as

$$\nabla_b F^{ab} = \nabla_b \nabla^a A^b - \nabla_b \nabla^b A^a = J^a. \quad (3.6)$$

Upon using the Ricci identities (2.4), and noting that in a source free region $J^a = 0$, we find Maxwell's equations in vacuum as

$$\nabla_b \nabla^b A_a + R_{ab} A^b = 0, \quad \nabla_a A^a = 0. \quad (3.7)$$

A general solution to this equation is hard to find. However the solutions we are seeking should describe the paths of photons travelling from the source to the observer propagating as electromagnetic waves. Furthermore, since the interstellar medium is not conductive, these waves are not expected to suffer significant attenuation. As a result their phase will vary rapidly compared to their amplitude. Other than that the solutions we are seeking should be able to convey arbitrary information i.e. exhibit any mixture of polarisation states, amplitudes and frequencies. We therefore propose the following ansatz for the four potential

$$A_a = g(\psi) \alpha_a + \text{small terms}, \quad (3.8)$$

where $g(\psi)$ is an arbitrary function of the phase which varies rapidly in comparison with the amplitude α_a . When the spacetime curvature length is large compared to the wavelength of the

wave¹ the small terms do not effect photon propagation and can be ignored. We now define the propagation vector as

$$k_a := \psi_{,a} \quad \Rightarrow \quad k_{[a;b]} = 0, \quad (3.9)$$

and substitute our ansatz (3.8) into (3.7). Since $g(\psi)$ is arbitrary we can equate any of the coefficients of g, g' or g'' to zero. This can be used to establish that [25]

$$k^a k_a = 0 \quad \Rightarrow \quad k^a \nabla_a \psi = 0, \quad (3.10)$$

$$2k^b \nabla_b \alpha_a = -\alpha_a \nabla_b k^b. \quad (3.11)$$

Using equation (3.10) together with equation (3.9) shows that the light rays parallel to k^a are null geodesics, as expected. The second of (3.10) shows that

$$k^a \nabla_a \psi = \frac{d\psi}{dv} = 0, \quad \Rightarrow \quad \frac{dg}{dv} = \frac{dg}{d\psi} \frac{d\psi}{dv} = 0, \quad (3.12)$$

so that $\psi(x^a)$ is constant along the path of the ray. The surfaces $\psi = \text{const.}$ therefore define the future lightcones of the emitter. The constancy of the signal $g(\psi)$ along the rays also implies that the information it conveys remains unchanged. Finally, equation (3.11) establishes that (see the discussion in §7.1 of [25])

1. the polarisation of light is unaffected by the curvature of spacetime (at least when ignoring the small terms in (3.8)),
2. the magnitude of the amplitude propagates as

$$\frac{d}{d\nu} \alpha^2 = -\alpha^2 \nabla_a k^a. \quad (3.13)$$

The ansatz (3.8) seems to be sufficient to contain the solutions to (3.7) that we are seeking which therefore justifies the use of the geometric optics approximation to make astrophysical observations. In almost-FLRW expanding spacetime it is also possible to establish, at large distances from an isolated systems, that (3.8) is a necessary solution to (3.7). However, light propagation in general perturbed models of the Universe can have subtle consequences [66, 67] which merit careful scrutiny. Note that a solution to (3.7) is required if we are to infer the properties of astrophysical objects from observations made with earth based detectors. An observer moving with four velocity u^a will then be able to relate the measured F_{ab} (determined from the measured electric E_a and magnetic B_a components using (3.1)) to F_{ab} at the source.

¹Since the Ricci curvature $R \sim \frac{1}{l^2}$, where l is the curvature length scale, an equivalent statement is that the spacetime curvature is small compared to the frequency ν of the wave squared i.e. $R \ll \nu^2$.

3.2 Redshift

Observations are usually reported as functions of redshift, where the redshift z of a source is defined as

$$z := \frac{\lambda_o - \lambda_e}{\lambda_e} = \frac{\Delta\lambda}{\lambda_e}.$$

Here λ is the wavelength and the subscripts e and o label the emitter and observer respectively. As discussed in 2.1 the quantity $-u_a k^a$ is proportional to the photon energy $E = h\nu$ measured by and observer travelling with u^a . If we assume that both the emitter and observer are fundamental observers, then the cosmological redshift can be found from

$$1 + z = \frac{\lambda_o}{\lambda_e} = \frac{\nu_e}{\nu_o} = \frac{(u_a k^a)_e}{(u_b k^b)_o} = (u_a k^a)_e, \quad (3.14)$$

Since observational coordinates are defined in terms of past pointing k^a , the difference between future and past pointing vectors is a source of potential confusion in the above expression. The direction must be chosen consistently at both emitter and observer. However, since it is the observer that will see the redshift, we note that

$$k^a = -(1+z)(u^a + e^a), \quad \text{where} \quad e^a = \frac{h^a_b k^b}{(-u_c k^c)} \quad (3.15)$$

is always used for past pointing k^a . This establishes that

$$1 + z = k_a u^a = \partial_a w u^a \quad \Rightarrow \quad u^0 \stackrel{w}{=} 1 + z \quad \text{and} \quad u^1 \stackrel{w}{=} \frac{A(1+z)^2 - 1}{2(1+z)}, \quad (3.16)$$

where we have used the properties of the k^a (2.78) to find u^0 in observational coordinates and then used the normalisation $u_a u^a = 1$ with the form of the metric (2.86) to find u^1 . Also, in a universe with vanishing vorticity, we know that $u_a = -\partial_a \tau$. In comoving coordinates this implies²

$$1 + z = u_a k^a = -\partial_a \tau k^a, \quad \Rightarrow \quad k^0 \stackrel{\tau}{=} -(1+z) \quad \text{and} \quad k^1 \stackrel{\tau}{=} \frac{1+z}{X}, \quad (3.17)$$

where we have used the property of null vectors $k_a k^a = 0$ as well as the metric (2.59) to obtain the form of k^1 in comoving coordinates. Finally, we can use the above expressions to evaluate the form of the instantaneous direction of propagation of the ray e^a in both coordinate systems. For completeness we quote the result here

$$e^0 \stackrel{w}{=} -(1+z), \quad e^1 \stackrel{w}{=} -\frac{-A(1+z)^2 - 1}{2(1+z)^2}, \quad e^0 \stackrel{\tau}{=} 0 \quad e^1 \stackrel{\tau}{=} \frac{1}{X}. \quad (3.18)$$

From the expressions for u^a , k^a and e^a it is clear that the redshift already provides some kinematic information. However we should keep in mind that in reality there are a number

²Since $u_a k^a = \frac{-1}{a(t)}$ in FLRW this also justifies our assertion that $\frac{a_o}{a} = 1 + z$ in §2.3.1.

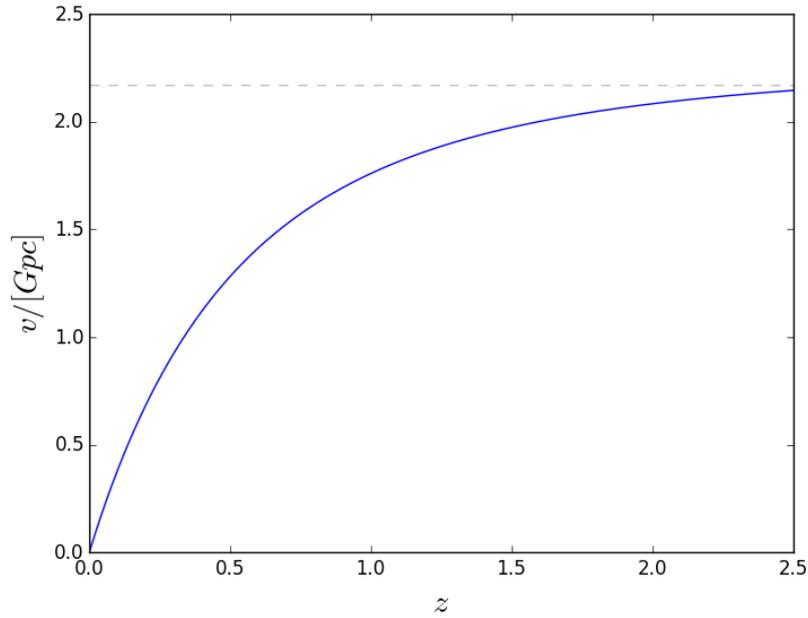


Figure 3.1: The $v(z)$ relation corresponding to our Λ CDM reference model defined by the parameters shown in Table 2.1. Note that $v(z)$ tends to an asymptotic maximum indicated by the dashed line in the figure.

of different contributions to the redshift. Examples include Doppler shifts induced by peculiar motions and a possible gravitational contribution to the redshift due to inhomogeneities in the gravitational potential. In what follows we assume that the reported redshift for any particular data set is the true cosmological redshift and neglect these additional contributions. For brevity we will sometimes use the shorthand notation $u = 1 + z$ where it should be understood that z is the true cosmological redshift.

3.2.1 The $v(z)$ relation

Next we obtain the null affine parameter as a function of redshift. Substituting (3.15) into the null geodesic and projecting along the direction of propagation of the ray gives

$$\begin{aligned} e^a k^b \nabla_b k_a &= -e^a \frac{dz}{dv} (u_a + e_a) + e^a (1+z)^2 (u^b + e^b) (\nabla_b u_a + \nabla_b e_a) = 0, \\ \frac{dz}{dv} &= u^2 e^a e^b \nabla_b u_a. \end{aligned} \quad (3.19)$$

If the vorticity and acceleration term in the decomposition of $\nabla_b u_a$ can be neglected, the last term on the right can be identified as the observed expansion rate H_{obs} defined in (3.25) below.

Thus the affine parameter as a function of redshift can be found by integrating (3.19)

$$\frac{dz}{dv} = (1+z)^2 H_{obs} \quad \Rightarrow \quad v(z) = \int_0^z \frac{dz^*}{(1+z^*)^2 H_{obs}(z^*)}. \quad (3.20)$$

Note that the $v(z)$ relation is not in general observable. In FLRW and LTB universe it follows from the existence of the fundamental four velocity and the vanishing of the acceleration and vorticity of (2.31). As opposed to the FLRW case where the acceleration vanishes by default (because of homogeneity), it would no longer be an observable relation in spherically symmetric universes with non-vanishing pressure. The $v(z)$ relation is central to our formalism because $v(z)$ is monotonically increasing as long as $H_{obs}(z)$ is everywhere positive (see Figure 3.1). This is a major advantage of using $y = v$ instead of $y = D$ since the $D(z)$ (see Figure 3.3) becomes multivalued beyond a certain limiting redshift which introduces additional boundary conditions when finding the solution numerically.

3.2.2 Redshift drift

In an expanding universe the redshift of an object does not remain constant over time. The evolution of redshift with proper time along our galaxy worldline is known as the redshift drift. It is possible to derive an expression for it in spherically symmetric observational coordinates by considering a 1st order Taylor expansion of $z(w_0 + \delta w, v_0 + \delta v)$ into

$$z(w_0 + \delta w, v_0 + \delta v) = z(w_0, v_0) + \left. \frac{\partial z}{\partial w} \right|_{v=\text{const.}} \delta w + \left. \frac{\partial z}{\partial v} \right|_{w=\text{const.}} \delta v. \quad (3.21)$$

Taking the limit of infinitesimal δw then gives³

$$\frac{dz}{dw} = \dot{z}|_{v=\text{const.}} + z'|_{w=\text{const.}} \frac{dv}{dw}. \quad (3.22)$$

When $ds^2 \neq 0$ we can use the observational metric to evaluate $\frac{dv}{dw}$ as

$$-d\tau^2 = -Adw^2 + 2dw dv = -dw^2 \left(A - \frac{dv}{dw} \right), \quad \Rightarrow \quad \frac{dv}{dw} = \frac{1}{2} \left(A - \frac{1}{u^2} \right), \quad (3.23)$$

where we have used $\frac{d\tau}{dw} = \frac{1}{u}$. Recognising that

$$\dot{z}|_{v=\text{const.}} = \dot{u} \quad \text{and} \quad z'|_{w=\text{const.}} = \frac{dz}{dv} = u^2 H_{\parallel}$$

then gives

$$\frac{dz}{dw} = \dot{u} + \frac{u^2 H_{\parallel}}{2} \left(A - \frac{1}{u^2} \right). \quad (3.24)$$

This describes the change in redshift as a function of proper time along \mathcal{C} . We will denote it as $\frac{\dot{z}}{\delta w}$ emphasising that it is not an on-the-lightcone observable. It is in principle completely model independent and, as we show in §6.3, will be very valuable to the observational cosmology programme (see also [68, 69, 70, 71, 72, 73]). There are a number of observational difficulties

³Note that $\delta v \rightarrow 0$ as $\delta w \rightarrow 0$

involved in actually measuring it though. In a Λ CDM universe the redshift drifts by approximately 10^{-8} per century at $z \approx 1$ (see Figure 3.3). Obtaining redshift drift data therefore requires the ability to measure redshifts to extreme precision. This is one of the design goals of the E-ELT's CODEX spectrograph [74] which faces a number of challenges [75]. Getting an accurate observation of the redshift drift will inevitably involve averaging the redshift drift of a large number of sources measured to fall within certain redshift bins (see the discussion in 6.3). The data will therefore have both vertical (i.e. uncertainty in $\frac{\delta z}{\delta w}$) and horizontal (i.e. uncertainty in z) error bars. This is an effect that we completely neglect in the simulations of §6.3. Properly accounting for the uncertainty in redshift requires knowing the uncertainty in the reported redshifts of the other observables. The possibility of incorporating uncertainty in redshift is briefly discussed in § 7.

3.3 Expansion

3.3.1 The longitudinal and transverse expansion rates

The existence (and differentiability) of the fundamental four velocity can be used to define a generalised Hubble relation in terms of the kinematic quantities of §2.2.2 as (see §4.6 of [25] for example)

$$H_{obs} := \frac{1}{3}\Theta + \sigma_{ab}e^ae^b + e^a\dot{u}_a, \quad (3.25)$$

Note that it does not necessarily refer to the same expansion rate as that in the Friedmann equation. Spherically symmetric universes can expand in two different directions. This is readily evident from the comoving form of the spherically symmetric metric (2.59), where the two different scale factors make it clear that the Universe can expand differently in directions longitudinal and transverse to our line of sight. Evaluating $\Theta = \nabla_a u^a$ with the metric (2.59) gives

$$\Theta = \frac{\partial_t X}{X} + 2\frac{\partial_t R}{R} = H_{\parallel} + 2H_{\perp}, \quad (3.26)$$

which we might have expected since there are two dimensional surfaces (the screenspace) transverse to each one dimensional line of sight. Direct evaluation of (3.25) using the metric (2.59) shows that $H_{obs} = H_{\parallel}$ in spherical symmetry. This is not surprising because we do, after all, observe along the radial direction. Evaluating (3.25) with the observational metric (2.86) gives

$$H_{\parallel} = \dot{u} + \frac{u'}{2u^2} + \frac{u'A + uA'}{2} = \frac{u'}{u^2}, \quad (3.27)$$

where the last step can be inferred directly from (3.20) or by using the MCE (4.15). The first expression will mainly be useful to derive the form of H_{\perp} below. The derivative $H_{\parallel,z}$ required

to compute the consistency relation (3.52) can therefore be written as

$$\frac{d}{dz}H_{\parallel} = \frac{dv}{dz} \frac{d}{dv}H_{\parallel} = \frac{1}{u^2 H_{\parallel}} \left(\frac{u''}{u^2} - 2 \frac{(u')^2}{u^3} \right) = \frac{u''}{u^2 u'} - 2 \frac{u'}{u^3}. \quad (3.28)$$

To derive an expression for H_{\perp} we note that, since Θ is a scalar, the RHS of (3.26) must hold in any coordinate system. Computing Θ with the observational metric (2.86) gives

$$\Theta = \dot{u} + \frac{uA'}{2} + \frac{u'A}{2} + \frac{u'}{2u^2} + \frac{2u\dot{D}}{D} - \frac{D'}{uD} + \frac{uD'A}{D}, \quad (3.29)$$

from which the transverse expansion rate can be found as

$$H_{\perp} = \frac{1}{2} (\Theta - H_{\parallel}) = \frac{1}{D} \left(u\dot{D} - \frac{D'}{2u} + \frac{uAD'}{2} \right). \quad (3.30)$$

Since $H_{\parallel} = H_{\perp}$ in any FLRW geometry these expressions can be used to test for radial inhomogeneity down the PLC. In particular we can form the dimensionless quantity

$$\begin{aligned} T_1 &= 1 - \frac{H_{\perp}}{H_{\parallel}}, \\ &= 1 - \frac{u^3}{u'D} \left(\dot{D} - \frac{D'}{2} \left(\frac{1}{u^2} - A \right) \right). \end{aligned} \quad (3.31)$$

which should be zero in the absence of radial inhomogeneity. This expression will be used as a consistency relation to test the Copernican principle in §6. There we will also use the numerical solution to the field equations to reconstruct T_1 on the inside of the PLC. Note that, in a real lumpy universe described by FLRW + perturbations, the quantity T_1 will only vanish in the background. Thus, even if the CP holds, T_1 will not turn out to be exactly zero. However, the covariance of the reconstructed distribution T_1 should lie within that predicted for a perturbed FLRW universe. In spherical symmetry we are only interested in its covariance down the PLC i.e. $\langle T_1(z), T_1(\tilde{z}) \rangle$ say. An estimate of the expected magnitude of this quantity is therefore required. Unfortunately this is beyond the scope of the current research (however see the discussion in §6.4).

Note that T_1 is related to the matter shear σ_{ab} . To see this, consider a spherical distribution of matter located somewhere along the PLC0 within a redshift bin, Δz_i say. If the expansion in the radial and transverse directions are not equal then this spherical distribution would become deformed, becoming oblate if $H_{\perp} > H_{\parallel}$ and prolate otherwise. Baryon acoustic oscillations (BAO) [76] should place very tight constraints on the expected magnitude of $\langle T_1(z), T_1(\tilde{z}) \rangle$ (see [77, 78]). However this will only be feasible when surveys become sensitive enough to separate the radial and transverse BAO scales, something that is difficult to do in a model independent way (however see [79]). Future surveys (e.g. intensity mapping surveys [80]) can in principle map the growth of structure with redshift and separate the transverse from the radial BAO scales. It should therefore be possible to constrain $\langle T_1(z), T_1(\tilde{z}) \rangle$ directly with projects such as the SKA, for example.

3.3.2 Cosmic chronometers

The cosmic chronometer approximation of passively evolving galaxies (PEGs) can be used to measure $H_{\parallel}(z)$ [81]. First note that applying the chain rule the first of (3.20) and using (3.17) gives

$$H_{\parallel}(z) = \frac{-1}{(1+z)} \frac{dz}{dt}, \quad (3.32)$$

Thus the longitudinal expansion rate follows directly from the knowledge of $\frac{dz}{dt}$. PEGs are galaxies that are not undergoing significant star formation. These are mainly red galaxies whose light is dominated by old stellar populations, many of which have been identified [82, 83]. We will not go into the details (which are well documented in [84, 85, 86, 87]) but simply outline the main idea. The essential feature of PEGs is that an observable feature of their spectra, called the 4000Å break and denoted $D4000$, is very nearly a linear function of their age t_a i.e.

$$D4000 = A(SFH, Z) \cdot t_a + B, \quad (3.33)$$

where $A(SFH, Z)$ is a factor that depends on the star formation history SFH and metallicity Z . This factor can be calibrated using stellar population synthesis (SPS) models (see [88] for example). Here B is another factor which is sensitive to a large number of statistical and systematic uncertainties. The differential dating technique avoids the majority of these uncertainties. This can be seen by using (3.33) to rewrite (3.32) as

$$H_{\parallel}(z) = -\frac{A(SFH, Z)}{1+z} \frac{dz}{dD4000}, \quad (3.34)$$

in which the B factor cancels out. Thus, by calibrating $A(SFH, Z)$ with SPS models, $H_{\parallel}(z)$ can be found by binning together PEGs at different redshifts and computing $\frac{dz}{dD4000}$. There are obviously a number of assumptions involved which can effect the systematics. The most relevant to the current discussion is the assumption that all PEGs formed at approximately the same time. This assumption is required to compare galaxies at different redshifts. As pointed out in [89], the assumption is not strictly valid in inhomogeneous models of the Universe because galaxy formation time can depend on position in ALTB models. Note that the authors of [84, 85, 86, 87] state quite explicitly that they analyse these data within the FLRW framework. Thus, in principle, it is possible to get $H_{\parallel}(z)$ in this way because the relation (3.32) holds true in both FLRW and ALTB models. For the time being we should make explicit an additional assumption which is necessitated by the availability, or lack thereof, of current model independent $H_{\parallel}(z)$ data viz. that galaxy formation starts, peaks and eventually falls off as predicted by the concordance model. Unfortunately this is a bit circular since the concordance model is the one under scrutiny. It might be possible to quantify the additional uncertainty introduced when this assumption

is not valid by properly analysing perturbations in radially inhomogeneous universes. Linear perturbations in LTB models are investigated in [90], for example.

The $H_{\parallel}(z)$ data used in this work were taken from [86] and [87] and are shown in Figure 3.3. Given the above assumption on the formation time of galaxies, we have to keep in mind that these data are inevitably biased towards models that were fairly homogeneous at early times (i.e. in models where the bang time $t_B(r)$ is constant). Allowing for the possibility of a varying formation time will increase the width of the error bars on the $H_{\parallel}(z)$ data. This would result in a widening of the reconstructed contours in the simulations performed in § 6.2 and will not alter the conclusions drawn from currently available data.

3.4 Distance

Distance, being the cornerstone of 20th century cosmology, is an essential astrophysical observable. Note that there are multiple notions of distance in cosmology some of which are not directly observable. Currently the most reliable way to measure cosmological distances is by treating supernovae of type Ia as standard candles [91]. This will probably remain true for some time [92]. The essential thing to realise though is that it is not actually the distance that is measured. The quantity that is measured is related to the flux passing through the surface of a detector. We therefore have to relate the observed flux to distances in cosmology.

3.4.1 Flux, luminosity and distances

The Maxwell energy momentum tensor is defined as

$$T^{ab} = -F^{ac}F_c{}^b - \frac{1}{4}g^{ab}F_{cd}F^{cd}. \quad (3.35)$$

In the geometric optics approximation (3.8) it therefore takes the form

$$T_{ab} = \alpha^2(g')^2 k_a k_b. \quad (3.36)$$

An observer with four velocity u^a finds the instantaneous flux \mathcal{F} across a surface perpendicular to k^a to be

$$\mathcal{F} = T_{ab}u^a u^b = \alpha^2(g')^2 (k_a u^a)^2. \quad (3.37)$$

However detectors do not measure \mathcal{F} . In practice the output from a detector is a time varying complex valued voltage for each frequency in the bandwidth it is sensitive to. Simplifying things a bit, the weighted time averages of these voltages can be turned into an observed flux F which can then be used to define the apparent (bolometric) magnitude m of the source as

$$m := -2.5 \log_{10} F + \text{const.}, \quad (3.38)$$

where the constant term on the right must be calibrated. This is usually done by insisting that the distance modulus μ vanishes at 10pc, where μ is defined as

$$\mu := m - M = -2.5 \log_{10} \frac{F}{F_{10}}. \quad (3.39)$$

Here F_{10} is the flux of a source observed at a distance of 10pc and M is its absolute magnitude. It does not really matter which distance is used because they are all approximately equal to the Hubble distance $D_{H_0} = \frac{c}{H_0}$ on scales below 10 Mpc. The Hubble constant is therefore the main cosmological parameter involved in calibrating the absolute magnitude. Strictly speaking we get a different $\mu(z)$ relation for each value of H_0 which should therefore be marginalised over in the analysis. This is an added level of detail that will not significantly alter our conclusions in §6. Opting for simplicity, we will therefore neglect the marginalisation in the current work and simply comment on how it could be expected to affect our results (see the discussion in §6.4). Next we relate μ to cosmological distances.

Consider the cross sectional area dS of a congruence of null geodesics diverging from a source. Then it can be established, for a source emitting isotropically, that

$$F dS = \frac{L}{4\pi} \frac{d\Omega}{(1+z)^2}. \quad (3.40)$$

Here L is the total luminosity of the source

$$L = \int_S (1+z)^2 F dS, \quad (3.41)$$

where the integral is taken over the surface of a sphere surrounding the source. Referring to figure 3.2, the source area distance r_s is defined by the relation

$$dS_o = r_s^2 d\Omega_s, \quad (3.42)$$

where $d\Omega_s$ is the solid angle at the source and dS_o is the cross sectional area at the observer. Since the choice of sphere to surround the source with is arbitrary, we could choose a sphere with radius r_s . However r_s is not directly observable. To turn (3.40) into an observable relation we further define the luminosity distance as

$$D_L := (1+z)r_s, \quad (3.43)$$

in which case the flux is related to D_L via

$$F = \frac{L}{4\pi D_L^2}. \quad (3.44)$$

Note that the cosmological model is not involved in the definition of D_L so that a model independent observation of the luminosity distance can in principle be determined from flux

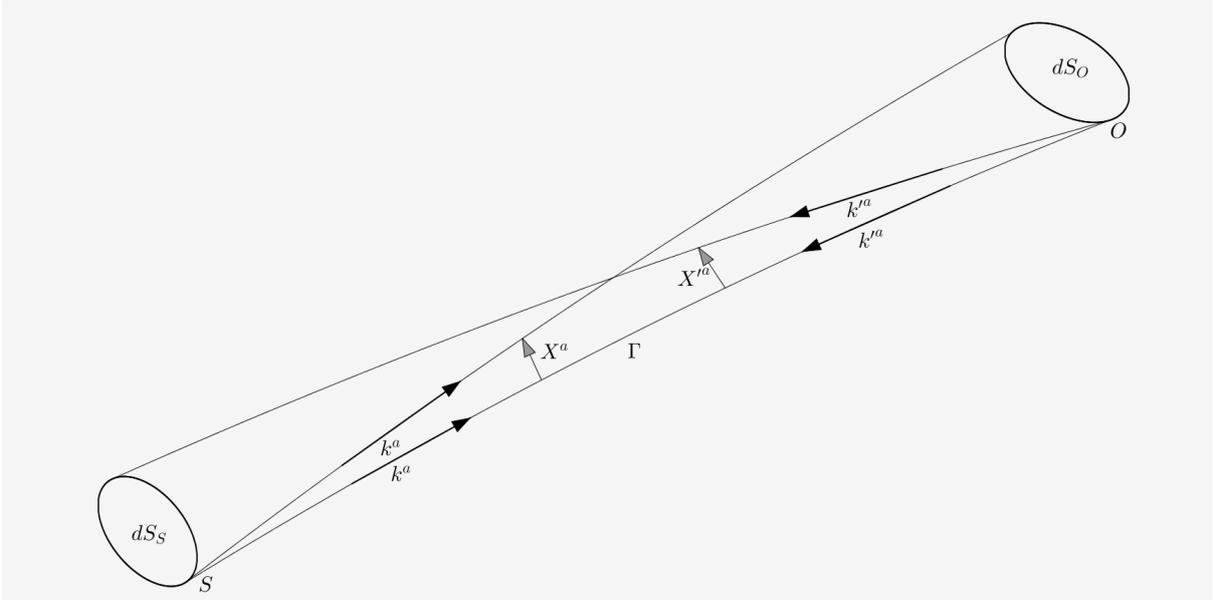


Figure 3.2: A demonstration of the areas and angles at the observer and source. This figure is inspired by Figure 7.1 of [25].

measurements. In practice the total luminosity of an arbitrary source is not usually known. However, note from (3.44) that the luminosity distances of two objects, one at a redshift of z^* and another at redshift z say, are related by

$$\frac{D_L(z)}{D_L(z^*)} = \sqrt{\frac{LF^*}{L^*F}}, \quad (3.45)$$

where L^* is the luminosity and F^* the flux of the source located at z^* . The idea behind standard candles is to identify and relate objects with known and effectively constant luminosities (more accurately reproducible light curves). In this case, if we know $D_L(z^*)$ and z^* say, the luminosity distance of the source located at a redshift z can be determined by measuring the ratio of the fluxes i.e.

$$D_L(z) = D_L(z^*) \sqrt{\frac{F^*}{F}}. \quad (3.46)$$

The luminosity distance therefore relates to the distance modulus via

$$\mu = 5 \log_{10} \frac{D_L}{10\text{pc}}. \quad (3.47)$$

Another common distance measure in cosmology, the area or angular diameter distance D , is defined by considering a bundle of rays diverging from the observer to the emitter. The ray bundle will subtend a solid angle $d\Omega_o$ with cross sectional area dS_s at the source i.e.

$$dS_s = D^2 d\Omega_o. \quad (3.48)$$

Given some prior knowledge about source, so that dS_s can be estimated, we can deduce D from the measured value of $d\Omega_o$. The BAO defined a standard ruler which can be used, at

least in perturbed FLRW models, to determine dS_s and hence the angular diameter distance D . When the size of an object (i.e. dS_s) is difficult to estimate it is useful to relate the angular diameter distance to the luminosity distance. This is possible as long as the number of photons is conserved and they travel along unique geodesics in which case the two distances are related by [93]

$$D_L = (1 + z)^2 D. \quad (3.49)$$

This result, known as the reciprocity or distance duality relation, will be used to convert $\mu(z)$ into $D(z)$ data. Next we derive another consistency relation which must hold in FLRW models and can hence be used to test the CP.

As proposed in [78] it is possible to formulate a test of the CP that is in principle independent of the matter content and particular theory of gravity employed. This can be done by investigating the consistency between distances and the expansion rate. The two main geometric effects on distance measurements are:

1. curvature bends null geodesics,
2. expansion changes radial distances.

In FLRW these two effects are coupled via the relation (assuming distance duality)

$$D = \frac{1}{uH_0\sqrt{-\Omega_{K0}}} \sin \left(\sqrt{-\Omega_{K0}} \int_0^z \frac{dz^*}{H(z^*)} \right). \quad (3.50)$$

This gives Ω_{K0} in terms of $H(z)$ and $D(z)$ as

$$\Omega_{K0} = \frac{[H(uD_{,z} + D)]^2 - 1}{[H_0uD]^2}. \quad (3.51)$$

Since Ω_{K0} is expected to be independent of z this expression can be derived to yield a quantity that should be zero in FLRW models viz.

$$T_2 = 1 + H^2 [u^2(DD_{,zz} - D_{,z}^2) - D^2] + uHH_{,z}D [uD_{,z} + D], \quad (3.52)$$

where we use $H = H_{\parallel}$ since it is the radial expansion that effects radial geodesics. Note that although this quantity is in principle independent of the matter content and theory of gravity employed, the quantity that we reconstruct is not. This is because we use one of the field equations to constrain $D(z)$ and its derivatives. Accurately correlating and reconstructing derivatives of $D(z)$ and $H_{\parallel}(z)$ in a non-parametric way is a very difficult task otherwise. Similar ideas as those regarding the expected value of T_1 in perturbed FLRW models (see the discussion in §3.3) apply to T_2 as well.

3.4.2 Standard candles

A standard candle is any astrophysical object whose luminosity is reproducible because of some characteristic shared by the entire class of such objects. Because of their characteristic light curves and high absolute luminosity, supernovae of type Ia (SNe Ia) are currently considered as the best standard candles for cosmology. Supernovae are amongst the best studied astrophysical objects with a long history devoted to understanding the physical mechanisms by which they occur (see for example [94, 95, 96, 97, 98, 99]). The focus of this section is on the assumptions that go into using SNe Ia to measure cosmological distances (see [91] and references therein for a detailed review).

All supernovae events are believed to be star explosions and are divided into two main classes viz. type I and type II. Type II supernovae (SNeII) exhibit hydrogen lines in their spectra and are typically fainter than type I supernovae (SNeI) which do not have hydrogen lines in their spectra. SNeI are further subdivided into three categories labelled by a, b and c. As well as the empirical relation between the peak and width of their luminosity curve, SNe Ia are characterised by the existence of a silicon absorption line in their spectra. Neither SNeIb or SNeIc exhibit silicon absorption lines, they are distinguished by the fact that SNeIb display a helium line, whereas SNeIc do not. All but SNe Ia are thought to be core-collapse explosions, the progenitors of which are massive stars. At the end of their lifetimes these stars start running out of fuel at which point, since the fusion chain is only exothermic up to iron nuclei, all further fusions are endothermic. These endothermic processes reduce the pressure generated from heat radiation which is subsequently overcome by the gravitational attraction and the star's core collapses. Most of the time these events result in the formation of a neutron star or black hole. Core-collapse supernovae can arise from progenitors with very different properties (e.g. wide range in mass, chemical abundance etc.). They therefore exhibit a broad variation of luminosities which makes it difficult to standardise their lightcurves.

SNe Ia on the other hand are thought to occur in binary systems when a carbon-oxygen white dwarf (WD) accretes the mass of its companion star up to a certain mass threshold called the Chandrasekhar mass $M_{Chan} \approx 1.4M_{\odot}$ [100]. Unlike the massive stars above, the gravitational attraction in WD stars is opposed by the Fermi pressure exerted by a degenerate electron gas. They are much colder than stars still undergoing nuclear fusion. When the mass of the WD reaches M_{Chan} the Fermi pressure is no longer able to balance the gravitational attraction. This causes the overall temperature to increase to a value high enough to trigger carbon fusions, initiating a fusion chain that can burn up all the material in the star within a short space of time [101], typically thought to be of the order of seconds. Although there are a number of

observational facts that support this paradigm, there is also some ambiguity about the path to the initial state [102, 103, 104]. Since the Chandrasekhar mass has to be reached from below, the WD has to gain mass by some mechanism. There are two obvious ways by which the WD can gain mass from its companion in a binary system. Firstly, it could slowly be accreting the surface elements, the most likely candidates being H and He, from a non-degenerate companion. In this case the absence of H and He in SNe Ia spectra is difficult to explain. Another possibility, which would explain the absence of H and He, is if the WD has another degenerate companion. The loss of orbital angular momentum to gravitational radiation will eventually result in a merger. However, while the expected merger rate from stellar population synthesis models is compatible with the frequency at which SNe Ia are observed, this scenario does not naturally result in an initial state with mass approximately fixed at M_{Chan} . Thus there remains considerable uncertainty in the exact mechanism by which SNe Ia occur. However the fact of the matter is that they exhibit highly reproducible light curves and therefore do make good standard candles. Of course there are outliers but these are excluded from the analysis when the aim is to determine cosmological distances.

Cosmological distances are determined from SNe Ia by fitting a model to the lightcurves obtained from some instrument (see [105] for example). A lightcurve is a measure of intensity as a function of time but it is often plotted as normalised flux vs. phase or as magnitude vs days of maximum brightness. All are equivalent. Many different instruments are used to measure lightcurves in different frequency bands. The most consistent band for SNe Ia is the B band, the U band for example is sensitive to the lightcurve model. The union 2.1 compilation [4] uses the SALT2 technique [106] which fits three parameters for each SNe Ia viz. an overall normalisation to the time dependent spectral energy distribution (SED) x_0 , the deviation from the average lightcurve shape x_1 and the deviation from the mean SNe Ia $B - V$ colour c . The parameters x_1 , c and the integrated B -band flux at maximum light m_B^{\max} are then combined with the host mass in a distance modulus model of the form

$$\mu_B = m_B^{\max} + \alpha x_1 - \beta c + \delta P(m_{\star}^{\text{true}} < m_{\star}^{\text{threshold}}) - M_B, \quad (3.53)$$

where M_B is the absolute B -band magnitude of the SN Ia with $x_1 = 0$, $c = 0$ and $P(m_{\star}^{\text{true}} < m_{\star}^{\text{threshold}}) = 0$. The parameter α therefore characterises the brighter-slower (lightcurve width) relationship and β the brighter-bluer (lightcurve color) relationship. The parameter δ is included to account for the fact that SN Ia are known to correlate with the mass of the host galaxy even after correcting for the lightcurve width and color [107] which could bias the cosmological results [108]. The essential point to notice here is that, once the data fixes m_B^{\max} , x_1 and $P(m_{\star}^{\text{true}} < m_{\star}^{\text{threshold}})$, the likelihood depends on the parameters α , β , δ and M_B . These details

will play no further role in the remainder of the thesis. Our reasons for mentioning them is simply to point out that there are multiple nuisance parameters, not excluding the absolute magnitude, involved when fitting a cosmology with SNe Ia data. The authors of [109, 105, 4] all advocate performing a blind analysis in which the nuisance parameters are inferred (and marginalised over) simultaneously with the cosmology. Such an analysis would be most in line with the bottom up approach we are pursuing in this work. However, note that an unbiased treatment of the systematics would have to adopt a blind analysis for all the data sets used i.e. simultaneously marginalising over the nuisance parameters involved in all data gathering processes. Unfortunately such an analysis is beyond the scope of the current research. In §5.3 we simply indicate how the algorithm presented in §5.3.3 could be modified to perform such a blind analysis.

The $D(z)$ data used in this work are shown in Figure 3.3. These data points were obtained from the $\mu(z)$ of the Union 2.1 supernovae data [4] by first using (3.47) to convert to $D_L(z)$ and then (3.49) to get $D(z)$. The fact that we do not explicitly marginalise over the nuisance parameters inevitably biases the analysis towards Λ CDM models. However, since Λ CDM is contained within ALTB, it is unlikely that the marginalisation will significantly affect the mean values obtained in our simulations in § 6. It will, however, most probably widen the reconstructed contours in all the figures. This does not alter the conclusions drawn from the simulations on real data (see the discussion in § 6.2).

3.5 Density

The total energy density of the Universe is very difficult to measure in a model independent way. The most direct way to do this, dubbed number counts by Ellis in 1971 [29], is to count the number of objects observed out to a certain volume, V say. This gives the observed number density n_{obs} of sources within V which, given the masses (or mass to luminosity ratio) of the sources, can then be converted into a density measurement. However, since the luminous matter only accounts for a small portion of the total matter in the Universe, there are already a number of assumptions involved in translating the observed number density n_{obs} into a proper number density, n say. Here we will attempt to highlight some of the difficulties involved in obtaining a model independent measurement of the density from galaxy redshift surveys and then present an alternative method to obtain the density in ALTB models. Note that it is in principle possible to count any class of objects to obtain number count data. Our discussion will be limited to galaxy number counts but it would also be possible to count galaxy clusters, for example. In this case the galaxy cluster mass function (see [110] for example) provides an alternative to using

the galaxy luminosity function discussed below. However, such an approach will share many of the same difficulties and, as with galaxy number counts, also requires assuming a cosmological model.

3.5.1 Number counts

Our discussion regarding relativistic number densities and the luminosity function will be somewhat superficial. The aim of this section is simply to point out where assumptions about the cosmological model enter the analysis. Further details can be found in [111, 112, 113] with an analysis devoted to number densities specifically in void LTB models presented in [114, 115]. We start by deriving an expression for the number density of sources in spherical symmetry.

Let dN denote the number of sources detected in a bundle of rays for a small affine parameter displacement v to $v + dv$. If k^a is the tangent vector to a past directed null geodesic, then this corresponds to a distance $dl = (k^a u_a)dv$. The cross sectional area of the bundle is given by $dS_0 = D(w_0, v)^2 d\Omega$ so that the volume increment is just $dV = dl dS_0 = (k^a u_a) D(w_0, v)^2 dv d\Omega$. Denoting the number density of sources at an affine parameter distance v by $n(v)$, the number of sources detected in the volume increment dV is given by $dN = n(v)dV$. Assuming that the true number density $n(v)$ can be related to the observed number density $n_{obs}(v)$ by a function $f(v)$ such that $n(v) = f(v)n_{obs}(v)$, the number of sources included up to some distance v^* is given by

$$N(v^*) = d\Omega \int_0^{v^*} f(v)n_{obs}(v)(k^a u_a)D(v)^2 dv. \quad (3.54)$$

The assumption of isotropy gives the total number of sources over the whole sky up to a distance v^* as

$$N(v^*) = 4\pi \int_0^{v^*} f(v)n_{obs}(v)(k^a u_a)D(v)^2 dv. \quad (3.55)$$

The $v(z)$ relation (3.20) then allows us to write this in terms of the redshift as (here z^* is the value of z at v^*)

$$N(z^*) = 4\pi \int_0^{z^*} \frac{f(z)n_{obs}(z)D(z)^2}{uH_{\parallel}(z)} dz, \quad (3.56)$$

where we have also used $u_a k^a = 1 + z$. Given number count data we can get $n(z)$ by direct differentiation

$$\frac{dN}{dz} = 4\pi \frac{f(z)n_{obs}(z)D(z)^2}{(1+z)H_{\parallel}(z)} \Rightarrow n(z) = \frac{(1+z)H_{\parallel}(z)}{4\pi D(z)^2} \frac{dN}{dz}. \quad (3.57)$$

Since luminous matter only accounts for a small portion of the total matter in the Universe, and because it is a biased tracer of the non-baryonic matter [116], a relation between the luminous and dark matter in the Universe needs to be specified. This is compounded by the fact that real detectors only detect sources above a certain flux cut. Deciding what fraction

of galaxies lie above this flux cut is very difficult to do in general and is usually where the assumption of spatial homogeneity is invoked (see the discussion in [115]). In this sense the function $f(z)$, which relates $n(z) = f(z)n_{obs}(z)$, is called the selection function and $n(z)$ is the number density of galaxies out to a redshift of z . However, because the resolution of any particular instrument is limited in practice, it is not possible to count this number directly, at least not for galaxies. Instead what is actually measured is the luminosity (more accurately the flux) behind a certain patch of sky and within a specific redshift bin. The link between observation and theory is provided by the galaxy luminosity function (LF), often denoted by ϕ , which expresses the number density of sources per unit of luminosity (see [117] for a review). Furthermore, this approach yields number densities within each frequency band used for the survey. These need to be combined in some way to obtain the total number density $n(z)$. To obtain the total matter density $\rho(z)$ from $n(z)$ requires a number of additional assumptions. Probably the most direct way to do this is to estimate the total mass (of both visible and dark matter) for each source in the survey. In a perturbed FLRW model the Press-Schechter formalism [118] (or something similar, see [119] for example) provides a possible way to do this. Ideally such an analysis should also incorporate the evolution of galaxy bias with redshift [120]. For inhomogeneous cosmologies it is often assumed that (see [114, 115] for example), on average, all galaxies contain approximately the same amount of matter. This is not really a satisfactory assumption within the bottom up approach that we are pursuing in this research.

Furthermore, it is practically impossible to determine the angular diameter distance to each galaxy in the survey. The distance to the galaxy therefore has to be deduced from its redshift, using SNe Ia as discussed above for example. Clearly there is some work to be done before we can extract reliable and model independent $\rho(z)$ data from galaxy redshift surveys. However, since spherically symmetric models of the Universe have only two free functions, it is always possible to deduce $\rho(z)$ once these two functions have been fixed. In observational coordinates there exists a relatively simple relation connecting ρ to D and H_{\parallel} which we will exploit later on.

3.5.2 Density from $D(z)$ and $H_{\parallel}(z)$

One of the field equations resulting from the metric (2.80) in a universe containing dust can be written as (see §4.1)

$$\rho = -\frac{2D''}{\kappa u_1^2 D}, \quad (3.58)$$

with a prime denoting, as usual, differentiation with respect to v . This provides a clear way to obtain the density from $H_{\parallel}(z)$ and $D(z)$ measurements viz. use $H_{\parallel}(z)$ in the expression

(3.20) to get $v(z)$ and then, since $v(z)$ is invertible, also $u_1 = 1 + z(v)$, $D(v)$ and (by numerical differentiation or otherwise) $D''(v)$. Once $\rho(v)$ has been obtained the $v(z)$ relation gives $\rho(z)$. This is remarkably simple only in principle, in practice it is very difficult to prescribe $H_{\parallel}(z)$ and $D(z)$ independently. To see this, apply the chain rule to D and use (3.20) to write the derivatives as

$$D' = u^2 H_{\parallel} D_{,z}, \quad D'' = 2u^3 H_{\parallel}^2 D_{,z} + u^4 H_{\parallel} H_{\parallel,z} D_{,z} + u^4 H_{\parallel}^2 D_{,zz}. \quad (3.59)$$

These expressions show that $H_{\parallel}(z)$ and $D(z)$ cannot be specified independently in Λ LTB models. In §5.3 we explain how to circumvent this by specifying a priors for $\rho(z)$ and $H_{\parallel}(z)$ and then inferring their posterior distributions from the available data. Once the posterior of $\rho(z)$ is known on the PLC0 it can in principle be used to work backwards to find $n(z)$. This is a possible application of the code that will be discussed further in §7. Note that the form of (3.58) does not seem to prohibit specifying $H_{\parallel}(z)$ and $\rho(z)$ independently, at least on the PLC0. In §4.5.4 we will verify that this is indeed possible.

The relation (3.58) actually plays a major role in this formalism. A number of interesting facts can be derived from it. In particular, note that observations of $D(z)$, $H_{\parallel}(z)$ and $\rho(z)$ can't constrain the value of Λ , no matter how precise they are. The absence of Λ from this equation also illustrates that LTB models can fit $D(z)$ data arbitrarily well by adjusting the two functions $\rho(z)$ and $H_{\parallel}(z)$ (which can be freely specified on any single PLC). As evidenced by the remaining field equations (see §4.1), the value of Λ does however affect the evolution of these function from one PLC to the next. Actually the evolution histories of LTB and Λ CDM models with nearly identical $D(z)$ and $H_{\parallel}(z)$ relations on the PLC0 are very different (see for example [121]). We therefore briefly turn to an additional piece of information that will allow us to also infer the value of Λ and hence the form of the cosmological metric on the inside of the PLC.

3.6 Age of the Universe

The age of the Universe t_0 along \mathcal{C} follows from integrating the $H_{\perp}(R)$ relation (2.64) in Λ LTB models. This value depends on the values of both free functions at the vertex as well as the value of Λ . Together with $D(z)$ and $H_{\parallel}(z)$ data, a model independent observation of t_0 therefore completely determines the Λ LTB model. While it is possible to put a lower bound on t_0 by age dating the oldest observed objects in the local vicinity of our galaxy, an upper bound is not as forthcoming, at least for the time being. Three commonly employed methods to determine the ages of stars are:

- nucleochronology,
- cooling of white dwarf stars,
- main sequence turn-off ages from stellar evolution models.

We will not delve into the details, which can be found in [122, 123, 124, 125, 126] for example, since, as we show in §6, redshift drift data are able to provide strong constraints on the value of Λ . Doing so will also not be worth the effort because, as we will see, the current bound on the minimum age of the Universe t_{min} does not place stringent constraints on the free functions of the model. Simply put, once the age of a star has been determined, it places a lower bound on the age of our galaxy, t_{gal} say. In order to derive a minimum bound on the age of the Universe, we also need to know how long after the big bang our galaxy formed. Galaxy formation is expected to have begun between $z = 20$ and $z = 5$ which, in a Λ CDM model, corresponds roughly to 0.1 - 2 Gyr respectively. A conservative (but not completely model independent) estimate of the age of the Universe is therefore $t_{min} \geq t_{gal} + 2$ Gyr. We will use the bound derived, using nucleochronology, in [127] viz. $t_{min} = 15.2 \pm 2.7$ Gyr. Since this is a lower bound we only allow it to contribute to the likelihood when $t_0 < t_{min}$, where t_0 is found by integrating the $H_{\perp}(R)$ relation. This is explained further when we discuss our likelihood model in §5.3.3.

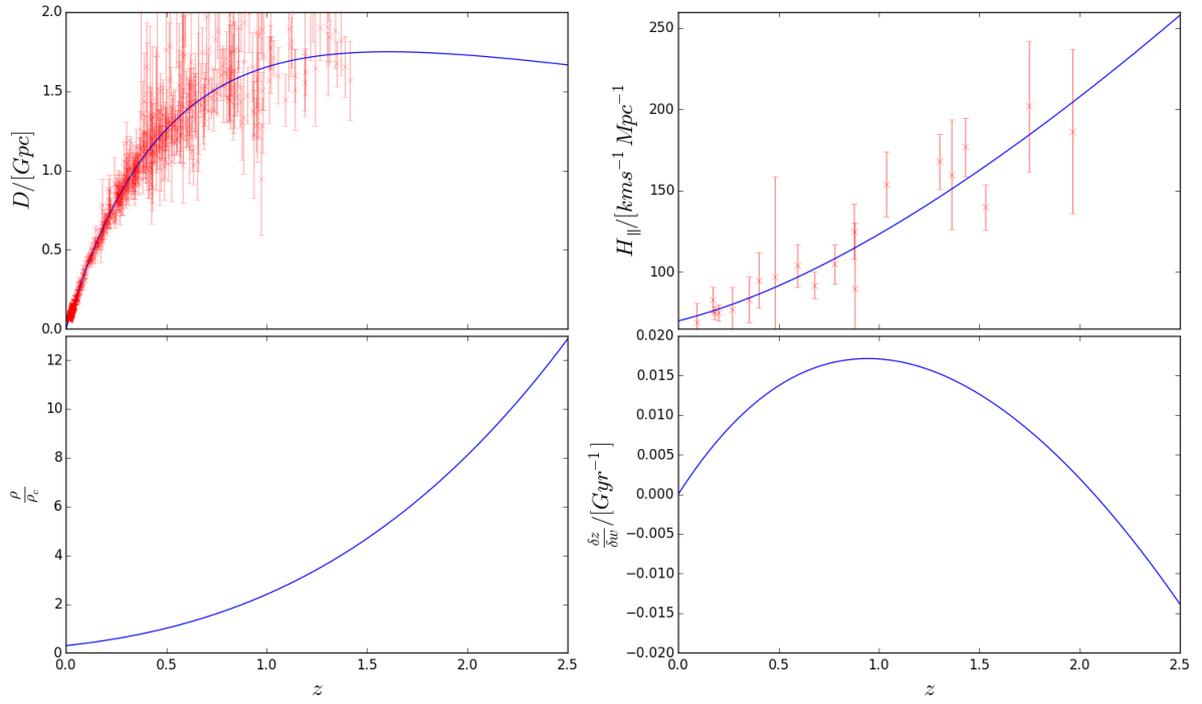


Figure 3.3: The observables as a function of redshift for corresponding to our Λ CDM reference model defined by the parameters shown in Table 2.1. Also shown are the $D(z)$ Union 2.1 supernovae data [4] and the $H_{||}(z)$ cosmic chronometer data [87].

4

Numerical Λ LTB solution

The Einstein field equations can be formulated as a characteristic initial¹ value problem in which the characteristics are radial null geodesics. The original numerical scheme used to integrate the field equations was developed, first for the case of a non-linear scalar wave [128], and then applied to the Einstein equations in [129]. Based on the formalism introduced by Bondi, Sachs and collaborators in the 1960's [62, 63], the codes of [128, 129] were used to study gravitational radiation. Since the Bondi-Sachs formalism uses a metric based on outgoing null geodesics it can, with some modifications, be adapted for the cosmological problem. In particular it was shown that the codes of [128] and [129] could be adapted for cosmology by changing the direction of integration in the “time” coordinate, effectively considering incoming instead of outgoing geodesics. These modifications were first implemented in [130]. Further developments followed in [121] where it was realised that the radial coordinate used in [130] (i.e. the angular diameter distance D) was not very well suited for going out to high redshifts. Because of the apparent refocusing of the angular diameter distance, $D(z)$ is not one-to-one over the whole domain of the problem. To go beyond this point the formalism was adapted to use the null geodesic affine parameter v instead. This led to the publication of the affine CIVP code which is described in depth in [131]. The affine CIVP code uses a metric based on outgoing instead of incoming null geodesics and simply reverses the direction of integration to get the past lightcone of the observer. In this chapter we adapt the affine CIVP code into a numerical scheme (henceforth referred to as the CIVP) more directly suitable for observational cosmology. We also test the convergence of the code when the initial data is specified in a non-parametric way.

¹For observational cosmology it is actually more appropriate to call it a final value problem because we start with data at the final state and evolve backwards in “time”.

4.1 Structure of the Equations

In their raw form the field equations (2.6) for the metric (2.86) in a dust universe are

$$\frac{1}{2}AA'' - 2\frac{\ddot{D}}{D} + \frac{A'\dot{D}}{D} - \frac{D'\dot{A}}{D} + \frac{AA'D'}{D} = \frac{1}{4}\kappa\left(\frac{\rho}{u^2} + u^2A^2\rho\right) - \Lambda A, \quad (4.1)$$

$$-\frac{1}{2}A'' - 2\frac{\dot{D}'}{D} - \frac{D'A'}{D} = -\frac{1}{2}\kappa u^2 A\rho + \Lambda, \quad (4.2)$$

$$-2\frac{D''}{D} = \kappa u^2 \rho, \quad (4.3)$$

$$-2D\dot{D}' - DD'A' - 2\dot{D}D' - A(D')^2 - ADD'' + 1 = \frac{1}{2}\kappa D^2\rho + \Lambda D^2. \quad (4.4)$$

This form of the equations is not very amenable to computation. Consider that our eventual aim is to provide initial data for this system in terms of the astrophysical observables discussed in §3. From our discussion at the end of §2.4 we should expect that fixing two functional degrees of freedom and the value of Λ provides sufficient information to solve the above system. We already know that $H_{\parallel}(z)$ provides the $v(z)$ relation required to convert observations reported as functions of z to functions of the coordinate v . Using v (as opposed to D) as the radial coordinate simplifies the treatment of the boundary regions² making $H_{\parallel}(z)$ a convenient choice for one of the free functions. Since the pair $[z, H_{\parallel}(z)]$ uniquely defines $v(z)$, and because this relation is always invertible, we will henceforth omit the functional dependence on z and v which are now completely interchangeable. As was shown in §3.5, ρ can in principle be obtained from H_{\parallel} and D by using (4.3). However, as evidenced by (3.59), the strong interdependence between D and H_{\parallel} makes it difficult to use D as the second free function. A more convenient choice is to use ρ as the second free function. With this choice we view (4.3) as a differential for D . The initial condition $D'(0) = 1$ then allows us to solve for D in a way that enforces its interdependence with H_{\parallel} by construction. With this consideration in mind we therefore cast (4.2)-(4.4) into a characteristic initial value problem of the form

$$D'' = -\frac{1}{2}\kappa D\rho u^2, \quad (4.5)$$

$$\dot{D}' = \frac{1}{2D} \left[1 - DD'A' - 2\dot{D}D' - A(D')^2 - ADD'' - \frac{1}{2}\kappa\rho D^2 - \Lambda D^2 \right], \quad (4.6)$$

$$A'' = \kappa Au^2\rho - 4\frac{\dot{D}'}{D} - 2\frac{A'D'}{D} - 2\Lambda, \quad (4.7)$$

$$\text{IC's} \quad D(0) = A'(0) = \dot{D}(0) = 0, \quad A(0) = D'(0) = 1, \quad (4.8)$$

where (4.8) are the initial conditions (IC's) which can be established from the central limit conditions (2.81) and by using the expressions obtained for the limiting behaviour of the metric components near \mathcal{C} , in particular (2.82). To obtain the solution numerically it is more convenient

²Note that with the choice $y = D$ there would be an additional boundary at the point where $D_{,z} = 0$.

to rewrite these equations as a system of first order equations. This is done by introducing new variables $S = D'$, $Q = \dot{D}$ and $Z = A'$, in terms of which the above system can be written as

$$D' = S, \quad (4.9)$$

$$S' = D'' = -\frac{1}{2}\kappa D\rho u^2, \quad (4.10)$$

$$Q' = \dot{D}' = \frac{1}{2D} \left[1 - DSZ - 2QS - AS^2 + \frac{1}{2}\kappa\rho D^2(Au^2 - 1) - \Lambda D^2 \right], \quad (4.11)$$

$$A' = Z, \quad (4.12)$$

$$Z' = A'' = \kappa\rho + 4\frac{QS}{D^2} + 2\frac{AS^2}{D^2} - \frac{2}{D^2}, \quad (4.13)$$

$$\text{IC's} \quad D(0) = Z(0) = Q(0) = 0, \quad A(0) = S(0) = 1. \quad (4.14)$$

These equations will be referred to as the hypersurface equations with the variables D , S , Q , A and Z being the corresponding hypersurface variables. Note that all derivatives of the hypersurface variables have been eliminated from the RHS. Once ρ , u and Λ are known they can be solved, using standard ODE techniques, as a single system of equations. As shown below the limiting behaviours of (4.11) and (4.13) are indeed regular despite the appearance of the factors of D in the denominator.

The input data needs to be evolved to the next PLC to find the solution in the interior of the PLC. This is achieved with the conservation equations $\nabla_a T^a_b = 0$. In observational coordinates the MCE (2.27) and the ECE (2.26) for the ALTB model are given, respectively, by

$$\dot{u} = \frac{1}{2} \left(\left(\frac{1}{u^2} - A \right) u' - Zu \right), \quad (4.15)$$

$$\dot{\rho} = \rho \left(-\frac{u'}{u^3} - 2\frac{Q}{D} + \frac{S}{D} \left(\frac{1}{u^2} - A \right) \right) + \frac{1}{2}\rho' \left(\frac{1}{u^2} - A \right). \quad (4.16)$$

These will be referred to as the evolution equations. They are used to evolve u and ρ to the next PLC where, along with the value of Λ , they can again be used to solve the hypersurface equations. Note that the evolution and hypersurface equations are solved separately as two systems of ordinary differential equations. This happens because the coordinate system is aligned with the characteristics of the solution.

The numerical scheme used to solve these equations is discussed in §4.4. Before it can be implemented though the regularity of the boundary regions first needs to be established.

4.2 Boundary regions

4.2.1 Inner boundary ($v \approx 0$)

The presence of the factors of D in the denominators of (4.11), (4.13) and (4.16) require special considerations close to the origin. The limiting behaviour as $v \rightarrow 0$ can be dealt with by

performing Taylor series approximations of the functions around $v = 0$. We start by writing the expansions of $u(w, v)$ and $D(w, v)$ on any single PLC as

$$\begin{aligned} u(w, v) &= u_0(w) + u'_0(w)v + \frac{1}{2}u''_0(w)v^2 + \mathcal{O}(v^3), \\ D(w, v) &= D_0(w) + D'_0(w)v + \frac{1}{2}D''_0(w)v^2 + \frac{1}{6}D'''_0(w)v^3 + \mathcal{O}(v^4), \\ A(w, v) &= A_0(w) + A'_0(w)v + \frac{1}{2}A''_0(w)v^2 + \frac{1}{6}A'''_0(w)v^3 + \mathcal{O}(v^4), \end{aligned}$$

where we show the coefficients up to third order since we will need the expansion for D' up to second order. The fact that $u(w, v) = 1 + z(w, v)$ can be used to establish

$$\begin{aligned} u(v) &= 1 + u'_0v + \frac{1}{2}u''_0v^2 + \mathcal{O}(v^3), \\ \dot{u}(v) &= \dot{u}'_0v + \frac{1}{2}\dot{u}''_0v^2 + \mathcal{O}(v^3), \end{aligned} \tag{4.17}$$

where it should be understood that the coefficients have an implicit dependence of w . Equation (4.17) establishes the boundary condition $\dot{u}(w, 0) = 0$. Next we note that, for ρ to be regular on \mathcal{C} , (4.5) implies $\lim_{v \rightarrow 0} D'' = 0$ on \mathcal{C} (i.e. for all w). Combining this with the fact that $D'_0 = 1$ along \mathcal{C} , gives

$$\begin{aligned} D(v) &= v + \frac{1}{6}D'''_0v^3 + \mathcal{O}(v^4), \\ D'(v) &= S(v) = 1 + \frac{D''_0}{2}v^2 + \mathcal{O}(v^3), \\ \dot{D}(v) &= Q(v) = \frac{\dot{D}'''_0}{6}v^3 + \mathcal{O}(v^4). \end{aligned}$$

This establishes that the limits

$$\begin{aligned} \lim_{v \rightarrow 0} \frac{Q}{D} &= 0, \\ \lim_{v \rightarrow 0} \frac{QS}{D^2} &= 0, \end{aligned}$$

are indeed regular. Furthermore, combining these expansions with the limiting behaviour of A along \mathcal{C} viz. (see (2.82))

$$A(w, v) = 1 + \frac{1}{2}A''_0(w)v^2 + \mathcal{O}(v^3),$$

shows that

$$\begin{aligned} \lim_{v \rightarrow 0} \frac{AS^2}{D} &= \frac{1}{v}, \\ \lim_{v \rightarrow 0} \frac{AS^2}{D^2} &= \frac{1}{v^2} + \frac{A''_0}{2} + D'''_0. \end{aligned} \tag{4.18}$$

The singular term in (4.11) is therefore exactly cancelled out and we have that

$$Q'(0) = 0. \tag{4.19}$$

This can be used to evaluate (4.11) at $v = 0$. Substituting the (4.18) into (4.13) also yields a regular expression for $Z'(0)$ which, upon using the limiting behaviour implied by (4.5), yields an identity. To find $A''(0) = Z'(0)$ we go back to the form (4.7) and use the expansions just derived to find

$$Z'(0) = \frac{\kappa\rho_0}{3} - \frac{2\Lambda}{3}. \quad (4.20)$$

This expression is used to evaluate (4.13) at $v = 0$. Finally, putting it all together, we also have that

$$\lim_{v \rightarrow 0} \frac{S}{D} \left(\frac{1}{u^2} - A \right) = -2u'_0. \quad (4.21)$$

Substituting this result into (4.16) gives

$$\dot{\rho}_0 = -3\rho_0 u'_0, \quad (4.22)$$

which can be used as boundary data for $\dot{\rho}$. This allows us to avoid the problematic factors of D in the denominators of (4.11), (4.13) and (4.16).

4.2.2 Characteristic cut-off

There is an additional subtlety that arises from the fact that our causal horizon evolves with time. In particular the observable part of an expanding universe was smaller in the past. As a result the maximum extent of v decreases as we move deeper into the PLC³. The maximum extent of v on each PLC can be computed by considering a radial null geodesic emanating from the maximum radial extent on the PLC0, v_{max} say, and intersecting \mathcal{C} (see figure 4.1). This geodesic, known as the characteristic cut-off line, represents the causal boundary of an observer that can only observe up to $v = v_{max}$ at $t = t_0$ and will be denoted by \mathcal{W} throughout. The change in the coordinate v as we move from one PLC to the next can be found directly from the form of the metric

$$ds^2 = 0 = -Adw^2 + 2dw dv, \quad (4.23)$$

$$dw \neq 0 \Rightarrow \Delta v = \frac{1}{2} \int_{w_i}^{w_f} Adw, \quad (4.24)$$

where Δv is the change in v as we move from a PLC defined by $w_i = \text{const.}$ to one defined by $w_f = w_i + \Delta w = \text{const.}$ The difficulty with performing the integral in (4.24) in practice is that both w and v change along the path of integration. Moreover, since we do not know the value of the coordinate v at the end of the path of integration, the integral cannot be performed. Instead an iterative procedure is used to determine \mathcal{W} . Denoting the maximum value of v on

³This is why we cannot treat v as a comoving coordinate.

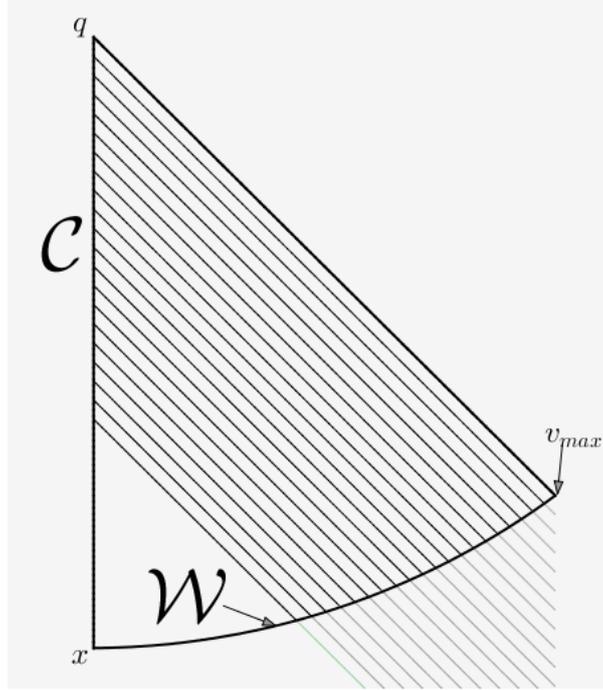


Figure 4.1: An illustration of the characteristic cut-off curve \mathcal{W} . This geodesic emanates from v_{max} and intersects \mathcal{C} at the point x . Nothing beyond \mathcal{W} can be causally connected to the observer at the point q .

the PLC labelled by i as v_i^* , we estimate the maximum value of v on the next PLC using

$$v_{i+1}^* = v_i^* + \Delta v = v_i^* + \frac{1}{2}A(w_i, v_i^*)\Delta w, \quad (4.25)$$

where we have used (4.23) to estimate Δv as we go from one PLC to the next. We then update the value of v_{i+1}^* iteratively according to

$$v_{i+1}^* = v_i^* + \frac{1}{2}A(w_{i+1}, v_{i+1}^*)\Delta w. \quad (4.26)$$

This is repeated on each PLC until the value of v^* converges. Since the value of v_{i+1}^* does not necessarily fall on a grid point, the value of $A(w_{i+1}, v_{i+1}^*)$ is calculated by linearly interpolating between the closest grid points on a specific w_i grid line. The integration is then performed only up to the closest spatial grid point which is less than v^* , anything beyond that is simply discarded. Next we compute the coordinate transformation which sends observational to comoving coordinates.

4.3 Coordinate transformation

A schematic of the transformation that relates observational to comoving coordinates is illustrated in Figure 4.2. If $\hat{t} = t(r)$ is the solution for t on the PLC0 then choosing $R(\hat{t}, r) = D(w_0, v)$

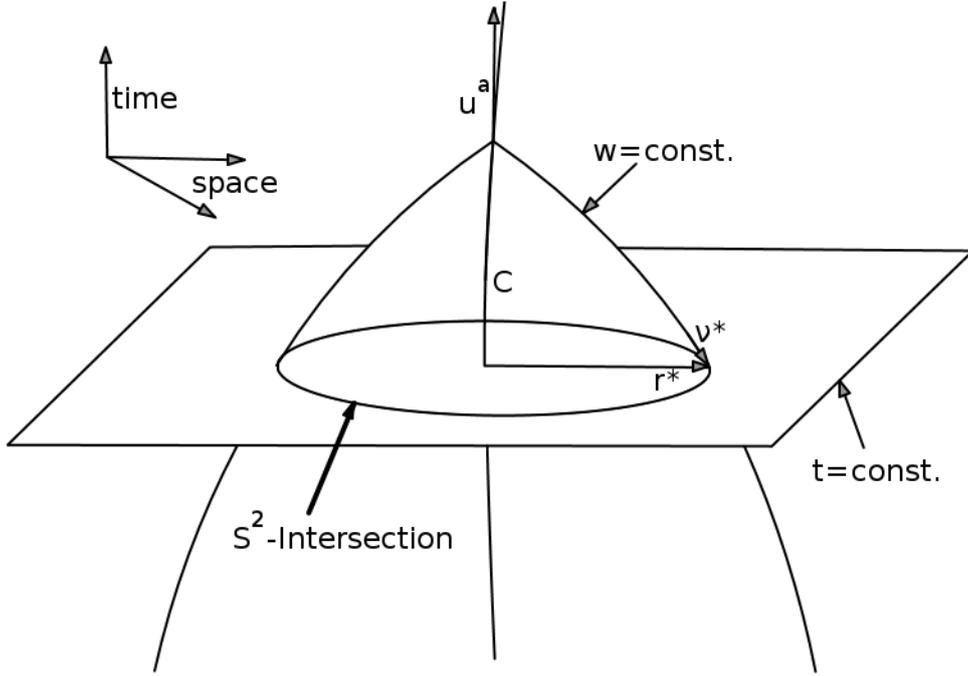


Figure 4.2: An illustration of the intersection of the two coordinate systems. It is only possible to reconstruct constant time surfaces that lie within the 2 sphere of intersection of the coordinate systems.

ensures that θ and ϕ have the same meaning in both coordinate systems. Thus we only need to consider the $(w, v) \leftrightarrow (t, r)$ transformation. The CIVP formalism solves for the background cosmological metric (2.86) in terms of observational coordinates x^a . We would like to compare these solutions to solutions in comoving coordinates $x^{\tilde{a}}$ with metric (2.59). Accordingly, we need to find both the metric components of (2.59) as functions of x^a and then explicitly solve for comoving coordinates in terms of x^a . The metric components follow from the transformation law

$$g_{\tilde{a}\tilde{b}} = \frac{\partial x^c}{\partial x^{\tilde{a}}} \frac{\partial x^d}{\partial x^{\tilde{b}}} g_{cd}. \quad (4.27)$$

Clearly we need expressions for these partial derivatives purely in terms of the observational metric and coordinates. Once the comoving metric components have been found in terms of observational coordinates, the geodesic equations can be solved to explicitly find comoving in terms of observational coordinates.

Gauge fixing the coordinate w to measure proper time along the central worldline makes the partial derivatives involving time straightforward to find. Note that

$$u_a u^a = -1 \Rightarrow dt = -u_a dx^a \quad \text{and} \quad k_a \stackrel{w}{=} \delta_a^0 \Rightarrow dw = k_a dx^a. \quad (4.28)$$

Interpreting $dt = -u_a dx^a$ as a coordinate transformation, we can use the form of u_a in observational coordinates to find $\frac{\partial t}{\partial w}$ and $\frac{\partial t}{\partial v}$. Using the observational form of the metric (2.86) to

obtain u_a from the expression (3.16) for u^a gives

$$\frac{\partial t}{\partial w} = \frac{Au^2 + 1}{2u}, \quad \frac{\partial t}{\partial v} = -u. \quad (4.29)$$

Next we can use the tensor transformation law to establish that

$$u^a = \frac{\partial x^a}{\partial x^{\bar{a}}} u^{\bar{a}} = \frac{\partial x^a}{\partial x^{\bar{a}}} \delta_0^{\bar{a}}, \quad \Rightarrow \quad u^1 \stackrel{w}{=} \frac{\partial v}{\partial t}.$$

Thus we can use the expression (3.16) for u^1 in observational coordinates to find $\frac{\partial v}{\partial t}$. Also noting, from (3.17), that $k_0 = u$, and interpreting $dw = k_a dx^a$ as a coordinate transformation, we have

$$\frac{\partial w}{\partial t} = u \quad \text{and} \quad \frac{\partial v}{\partial t} = \frac{Au^2 - 1}{2u}. \quad (4.30)$$

The partial derivatives involving r require a little more effort. The transformation (4.27), as well as its inverse, gives

$$d^2 X^2 = g^2 d^2 (a\dot{D} + bD')^2 = u^2, \quad (4.31)$$

$$bu - \frac{1}{2}auA - \frac{a}{2u} = 0, \quad (4.32)$$

$$cu + \frac{1}{2}duA - \frac{d}{2u} = 0, \quad (4.33)$$

$$ac + bd = 1, \quad (4.34)$$

where, for notational simplicity, we have abbreviated

$$a = \frac{\partial w}{\partial r}, \quad b = \frac{\partial v}{\partial r}, \quad c = \frac{\partial r}{\partial w}, \quad d = \frac{\partial r}{\partial v}.$$

and the first equality in (4.31) follows from (2.58) and the multivariate chain rule applied to $\partial_r R(t(w, v), r(w, v))$ i.e.

$$X = g(r)\partial_r R = g(r) \left(\frac{\partial w}{\partial r} \dot{D} + \frac{\partial v}{\partial r} D' \right).$$

Since there are five unknowns (viz. a, b, c, d and g) in four equations some additional information is required to solve this system. This is provided by the fact that the partial derivatives in these transformations commute. No new information can be derived from applying this criterion to a and b ; it simply recovers the momentum conservation equation (4.15). Applied to c and d however we get a partial differential equation (PDE) for d viz. $c' = \dot{d}$. This casts (4.33) into the following flux conservative form

$$\dot{d} = -\frac{\partial}{\partial v} \left(\frac{d}{2} \left(A - \frac{1}{u^2} \right) \right). \quad (4.35)$$

Given initial conditions for d this PDE can be solved alongside the CIVP. Initial conditions are provided by fixing the gauge freedom in r . It is clearest to proceed with an FLRW analogy. To

that end recall that in FLRW universes the comoving radial distance is simply $r = \frac{uD}{a_c}$, where a_c is a constant equal to the value of the scale factor on \mathcal{C} . Thus, in FLRW, the transformation is extremely simple, d is always given by

$$\frac{\partial r}{\partial v} = d = \frac{u'D + uD'}{a_c}. \quad (4.36)$$

Having fixed d we may use (4.31)-(4.34) to find the remaining unknowns viz. a, b, c and g . To find the transformation in general we should see (4.36) as a normalisation on the PLC0. The initial data for d then makes it possible to solve (4.35) and the remaining unknowns again follow from (4.31)-(4.34). Partial derivatives w.r.t. either r or t can then be expressed as derivatives w.r.t. w and v using straightforward tensor transformation laws.

One way to get $t(w, v)$ and $r(w, v)$ is to solve the following geodesic equations on each PLC

$$\frac{d^2 t}{dv^2} = -\frac{\partial_t X}{X} \left(\frac{dt}{dv} \right)^2 = -H_{\parallel} \left(\frac{dt}{dv} \right)^2, \quad (4.37)$$

$$\frac{d^2 r}{dv^2} = -\frac{\partial_r X}{X} \left(\frac{dr}{dv} \right)^2 - 2H_{\parallel} \frac{dr}{dv} \frac{dt}{dv}, \quad (4.38)$$

$$\text{IC's} \quad t(0) = t_0, \quad \frac{dt}{dv}(0) = -1, \quad r(0) = 0, \quad \frac{dr}{dv}(0) = \frac{1}{X}, \quad (4.39)$$

where the initial conditions follow the expressions (3.17) for k^a in comoving coordinates. These solutions allow us to associate the corresponding (t, r) pair to any (w, v) grid point. Note that the value of t_0 is required to obtain the coordinate transformation since it sets the initial conditions required to solve the system (4.37) and (4.38). The value of t_0 , and therefore the value of w on the PLC0, does not explicitly enter the CIVP equations and is therefore not required to obtain the ALTB solution in observational coordinates. Thus we mainly need the value of t_0 for algorithm verification purposes and to be able to use the age of the Universe as a data point.

It is difficult to compare the transformed CIVP solution to known ALTB solutions in comoving coordinates. This is because such solutions generally employ different gauge choices. In the comoving formalism gauge freedom in the coordinate r is usually fixed on Σ_{t_0} whereas we fix $r = uD$ on the PLC0. This choice ensures that we obtain the coordinates of the FLRW metric (2.35) when the universe is indeed homogeneous (see §2.3.1). It is therefore the most natural gauge choice to use when comparing the results of the CIVP to FLRW models. Next we describe the numerical procedure used to solve the CIVP and find the coordinate transformation.

4.4 Numerical Implementation

The numerical scheme described below is based on that developed in [131]. Besides change in notation and the reversal from outgoing to incoming geodesics, there are two main differences

between this scheme and the original viz.

- The hypersurface equations can be solved as a single system using Python’s built-in ode solvers. This allows the solution on the PLC0 (i.e. the one that will be confronted with data) to be found to greater accuracy using a high order scheme. Also, since the $v(z)$ relation has a horizontal asymptote (see Figure 3.1), derivatives in the CIVP equations can become very large towards the far end of the radial domain. Finding the solution out to high redshifts might therefore require adjusting the step size of the radial integrator. This is handled automatically by Python’s built-in solvers.
- The derivatives \dot{u} and $\dot{\rho}$ are evaluated at the inner boundary $v = 0$ using the boundary conditions derived in §4.2 without the need to resort to the interpolation scheme described in [131].

4.4.1 Initial data and radial integration

Since the hypersurface equations are solved using Python’s built-in ODE solvers there are a number of high order multi-step (explicit and implicit) methods available. However it should be kept in mind that using a higher order method than the one implemented for the evolution equations will not necessarily yield more accurate results in the interior of the PLC. Asking for higher order accuracy than that of the method used for the evolution equations can actually result in the integrator reporting non-convergence. As a result the integration domain has been split into two parts that are dealt with separately.

The first part consists of setting up the initial data and then finding the solution on the PLC0. The solution on the PLC0 is the only part of the solution that can be confronted with observations. However, before we can do that, it is also required to write the input data to a grid suitable for the numerical computation. This involves a number of interpolations and numerical integrations that can introduce additional numerical error into the scheme. These will be considered first. Note that, when referring to functions defined on a grid, we will use a superscript i to label the position on the temporal grid and a subscript j to label the position on the spatial grid. A superscript zero therefore refers to the data on the PLC0. We denote the number of spatial grid points by NJ , the number of temporal grid points by NI and the target error of the numerical scheme by ϵ_p .

In practice data are reported as functions of redshift at a finite number of points. This data has to be suitably smoothed before it can be used as input to the numerical integration scheme. The procedure used to smooth the data is described in §5.2.1, here we will focus on how to convert functions of the redshift to functions of the affine parameter. Thus it is assumed that

initial data are given as smooth functions $H_{\parallel}(z)$, $\rho(z)$ and the value of Λ . The $v(z)$ relation can then be obtained by interpolating the function

$$\frac{1}{(1+z)^2 H_{\parallel}(z)}, \quad (4.40)$$

and computing the cumulative integral (3.20) of the interpolated function. We will use cubic spline interpolation throughout. The error introduced by this operation depends on the number of redshift points used. We consistently choose enough evenly spaced redshift points, z_p say, so that the resulting error is at least two orders of magnitude below the target accuracy of the integration scheme. Since the error in cubic spline interpolation roughly scales like $\mathcal{O}(h^4)$, h being the step size, and the target error is $\epsilon_p \sim \mathcal{O}(h^2)$, this can usually be achieved by choosing the same number of points as is used to set up the spatial grid used in the integration scheme. The values $v_p = v(z_p)$ will not in general be evenly spaced apart. To write ρ and u to a regular spatial grid, v_j say, we perform another interpolation of $\rho(v_p)$ and $u(v_p)$. This gives the initial values $\rho_j = \rho(v_j)$ and $u_j = u(v_j)$ which can then be used, along with Λ , as input to solve the hypersurface equations. The procedure used to set up the initial data is summarised in Procedure 1. The notation $\mathcal{GP}_{H_{\parallel}}$ and \mathcal{GP}_{ρ} refers to the Gaussian processes for $H_{\parallel}(z)$ and $\rho(z)$ respectively and will be clarified in §5.2.1. For the time being, this notation simply means that we are able to generate random smooth realisations of $H_{\parallel}(z)$ and $\rho(z)$. Similarly \mathcal{N}_{Λ} is just a distribution from which to generate samples of Λ . The structures of Procedure 1 - Procedure 3 have been chosen to conform to their usage in Algorithm 1 by which time the notation will have been completely clarified.

Next we use this initial data, combined with the initial conditions (4.8), to find the solution on

Procedure 1

Input: ϵ_p , $\mathcal{GP}_{H_{\parallel}}$, \mathcal{GP}_{ρ} , \mathcal{N}_{Λ}

- 1 Set $\Delta v = \sqrt{\epsilon_p}$ and find NJ using (4.50);
- 2 Choose NJ values of z_p uniformly spaced in $[0, z_{max}]$;
- 3 Sample $H_{\parallel}(z_p) \sim \mathcal{GP}_{H_{\parallel}}$, $\rho(z_p) \sim \mathcal{GP}_{\rho}$ and $\Lambda \sim \mathcal{N}_{\Lambda}$;
- 4 Interpolate (4.40), compute $v_p = v(z_p)$ and set $v^* = \max(v_p)$;
- 5 Construct spatial grid v_j consisting of NJ points uniformly spaced in $[0, v^*]$;
- 6 Interpolate $\rho(v_p)$ and $u(v_p)$ and set $\rho_j^0 = \rho(v_j)$ and $u_j^0 = u(v_j)$;

Output: ρ_j^0 , u_j^0 , Λ

the PLC0. The radial derivatives u' , ρ' , d' and u'' , required to evaluate (4.15), (4.16), (4.35) and (3.52) respectively, are found numerically using fourth order accurate centred finite difference

stencils.⁴ The steps involved in finding the solution on any single PLC are summarised in Procedure 2.

The solution on the PLC0 can then be interpolated to reconstruct the observables at any value

Procedure 2

Input: $\rho_j^i, u_j^i, d_j^i, \Lambda$

- 1 Solve (4.9)-(4.13) for $D_j^i, S_j^i, Q_j^i, A_j^i$ and Z_j^i ;
if $i = 0$ **then** set d_j^0 with (4.36);
- 2 Find the coordinate transformation with (4.29)-(4.34);
- 3 Evaluate $(u_j^i)', (u_j^i)''$, $(\rho_j^i)'$ and $(d_j^i)'$ numerically using finite difference derivatives;
- 4 Solve (4.37) and (4.38) for t_j^i and r_j^i respectively;
- 5 Evaluate (4.15), (4.16) and (4.35) for \dot{u}_j^i , $\dot{\rho}_j^i$ and \dot{d}_j^i respectively;

Output: $D_j^i, S_j^i, Q_j^i, A_j^i, Z_j^i, \dot{u}_j^i, \dot{\rho}_j^i, \dot{d}_j^i, t_j^i, r_j^i$

of z . As explained in §5.1.3, this allows us to confront the solution corresponding to the initial samples of $H_{||}(z)$, $\rho(z)$ and Λ with data.

4.4.2 Temporal integration

The first step in finding the solution in the interior of the PLC is to construct the temporal grid. In order to compute the coordinate transformation, and use the t_{min} data point, we need to know the initial value of w viz. t_0 . After using $H_{||}(0), \rho(0)$ and Λ to compute the dimensionless density parameters in (2.61)-(2.63) at the vertex of the cone, the value of t_0 can be found with (2.67), which gives the current age of the universe along \mathcal{C} . We then construct a regular temporal grid w^i by choosing NI equally spaced points between $w^0 = t_0$ and $w^{NI-1} = t^*$, where t^* is the minimum time we wish to integrate up to. Thus t^* is the value of w on the final past lightcone (PLCF) considered.

Once the grid has been set up, the initial data is evolved from one PNC to the next using an explicit predictor-corrector method. The predictor step employs Euler's method to approximate an arbitrary function f on the next PLC as

$$\tilde{f}_j^{i+1} = f_j^i + \Delta w \dot{f}_j^i, \quad (4.41)$$

where a tilde is used to denote predicted values. We can therefore use (4.15), (4.16) and (4.35) to evaluate the temporal derivatives and predict \tilde{u}_j^{i+1} , $\tilde{\rho}_j^{i+1}$ and \tilde{d}_j^{i+1} respectively. These values are then used as input to Procedure 2 to get the corresponding predictions of the hypersurface

⁴One sided stencils are used at the edges of the computational domain.

variables on the next PLC. We then compute the new maximum extent v^* of the radial coordinate lying within the characteristic cut-off \mathcal{W} with the predicted value \tilde{A}_j^{i+1} as described in §4.2.2. Since v^* does not necessarily fall on a grid point, we let j_{max} denote the largest index of $v_j \leq v^*$ and discard the solution beyond j_{max} . The predicted hypersurface variables can then be used to evaluate the temporal derivatives \dot{u}_j^{i+1} , $\dot{\rho}_j^{i+1}$ and \dot{d}_j^{i+1} within the causal boundary of the next PLC. Then, by combining Euler's method (4.41) with the trapezoidal rule, the corrected values of u_j^{i+1} , ρ_j^{i+1} and d_j^{i+1} can be found using

$$f_j^{i+1} = f_j^i + \frac{\Delta w}{2} (f_j^i + f_j^{i+1}). \quad (4.42)$$

These corrected values are again used as input to Procedure 2 to find the hypersurface variables on time step $i + 1$. We then set $i \leftarrow i + 1$ and repeat the whole process until the domain of calculation has been exhausted. This is summarised in Procedure 3.

Procedure 3

Input: $D_j^0, S_j^0, Q_j^0, A_j^0, Z_j^0, \dot{u}_j^0, \dot{\rho}_j^0, \dot{d}_j^0$

- 1 Use Euler's method (4.41) to predict the values of \tilde{u}_j^{i+1} , $\tilde{\rho}_j^{i+1}$ and \tilde{d}_j^{i+1} ;
 - 2 Use \tilde{u}_j^{i+1} , $\tilde{\rho}_j^{i+1}$, \tilde{d}_j^{i+1} and Λ as input to Procedure 2;
 - 3 Find the characteristic cut-off as explained in §4.2.2;
 - 4 Use \dot{u}_j^{i+1} , $\dot{\rho}_j^{i+1}$ and \dot{d}_j^{i+1} from step 2 in (4.42) to find u_j^{i+1} , ρ_j^{i+1} and d_j^{i+1} ;
 - 5 Use u_j^{i+1} , ρ_j^{i+1} , d_j^{i+1} and Λ as input to Procedure 2;
- if** $i < (NI - 1)$ **then** set $i \leftarrow i + 1$ and go to 1;

Output: Solution

4.5 Diagnostics

The stability and convergence of the original code was investigated in detail in [131, 130, 121]. There it was shown that the code is stable against small errors on the input variables and second order convergent in both space and time. The order of convergence was established by comparing the numerical solution to a number of analytic solutions of the field equations viz. EdS, Λ CDM and a specific class of LTB models. However, in our application the initial data is specified in a non-parametric way. Since this introduces a number of additional interpolations and numerical integrations, we need to test the convergence of the scheme with these modifications. The fact that the solutions are not known analytically necessitates a three level convergence test. Such a test checks that successive refinements of the grid results in the expected improvement in

accuracy. Here we perform a three level convergence test for a number of samples drawn at random from Gaussian process priors over simulated initial data. We give a brief description of how data are simulated in §4.5.2. The process by which these data are smoothed is described in §5.2.1. Here we simply assume that a method for drawing smooth function realisations is available. The convergence and stability of the code will be confirmed for a large number of random realisations of $H_{||}(z)$, $\rho(z)$ and Λ . We start with a discussion about the possible sources of error in the integration scheme.

4.5.1 Stability and convergence

There are three sources of error involved in numerical solutions to initial value problems viz. truncation error, round off error and error due to incorrectly specifying initial or boundary data. Round off errors result from using floating point arithmetic. Truncation errors, on the other hand, result from discretising the computational domain and using truncated Taylor series expansions to evaluate derivatives. Errors from incorrect initial or boundary data can result in a number of different ways (e.g. user error or numerical reflections in boundary value problems). In our application these errors can result from errors introduced in setting up the initial data as described in Procedure 1. Although these errors are expected to be small, they can be amplified if the numerical scheme is not stable. A number of definitions are required before we can discuss the convergence of the CIVP.

The convergence of initial value numerical integration schemes can be analysed in terms of the well-posedness of the model as well as the consistency and stability of the finite difference approximation [132].

Well-posedness: An initial value problem is said to be well-posed if there exists a unique solution that is stable when subjected to small perturbations.

Consistency: A finite differencing representation is said to be consistent if the difference between the analytic solution to a PDE and its finite difference representation can be shown, in theory, to vanish as the grid is refined. Thus the truncation error of a consistent finite difference representation vanishes as the step size tends to zero.

Stability: A numerical scheme is said to be stable if none of the possible sources of error grow exponentially in the sequence of numerical calculations from one time step to the next.

Convergence: A numerical scheme is said to be convergent if the finite difference solution approaches the analytic solution of the PDE as the grid is refined.

The difference between consistency and convergence stem from the fact that the finite difference representation employed in a numerical integration scheme can be consistent even if the integration scheme itself is not stable. Stability considerations are more subtle. In fact it is very difficult to provide a general rule that ensures stability, especially for non-linear PDEs. For the linear case we have the following result due to Lax (see §3 of [132] for example):

Lax’s equivalence theorem: Given a well-posed initial value problem and a consistent finite difference approximation to it, stability is the necessary and sufficient condition for convergence.

There is no general analogue of this result for non-linear PDEs. Strictly speaking the stability analysis of each such PDE should be performed using, for example, Fourier or von Neumann analysis on a linearised version of the PDE. However, even then stability of the non-linear PDE is not guaranteed. In practical applications, a useful guideline is the Courant-Friedrichs-Lewy (CFL) condition. The CFL condition is the necessary (note not sufficient) condition for certain explicit first order initial value problems to converge. The CFL puts an upper limit on the maximum allowed time step Δw in the numerical integration scheme for a given spatial step Δv via

$$\left| \frac{\Delta w}{\Delta v} s_{max} \right| \leq 1, \quad (4.43)$$

where s_{max} is the maximum advection or wave speed in the problem, in our case the speed of light $s_{max} = c = 1$. Specific stability conditions for the CIVP were not derived explicitly. However, stability and convergence can be tested for numerically and it was found in [131, 130, 121] that the condition

$$\left| \frac{\Delta w}{\Delta v} \right| \leq 0.5, \quad (4.44)$$

usually results in a stable, second order convergent scheme. Note that the CFL condition imposes a limit on the computable domain inside the PLC. Since the number of grid points decrease as we move deeper into the PLC, and at least two grid points are required to compute numerical derivatives, there is always a tiny portion of the PLC that is beyond our reach viz. the intersection of \mathcal{W} with \mathcal{C} . However, for the grid spacing used in this work (typically $\Delta v \approx 1$ Mpc), this region lies well within the expected averaging scale. As already mentioned the stability and convergence of the original CIVP code was established by comparing the numerical solution to a number of exact solutions. Our goal now is to verify that second order

convergence is obtained for arbitrary initial data. First however, we will digress slightly to explain the method by which we simulate these data.

4.5.2 Simulating “realistic” test data

Simulated data will be used to perform the three level convergence test below and to verify the algorithm presented in §5.3.3. There are two main considerations behind the method we describe here viz. realistic data are not distributed as a perfect Gaussian about the background model and the uncertainty in cosmological data sets generally increases as a function of redshift. To account for this, we simulate data for an arbitrary function $F(z)$ according to the following procedure:

1. Draw n redshift values in the relevant range from a uniform distribution $z_n \sim \mathcal{U}(0, z_{max})$.
2. Assume that the relative error of the function F scales as a power law $\frac{\sigma_F}{\bar{F}} = \sigma_{F_0}(1+z)^\alpha$ where \bar{F} is its mean and σ_{F_0} its uncertainty at $z = 0$. We then draw N samples of F using

$$F(z_n) = \bar{F}(z_n) (1 + \sigma_{F_0}(1 + z_n)^\alpha \mathcal{N}(0, I_N)), \quad (4.45)$$

where $\mathcal{N}(0, I_N)$ is uncorrelated Gaussian noise (i.e. I_N is an $N \times N$ identity matrix).

3. Sort the samples at each redshift and call the value at $N/2$ the mean, the value at $0.16N$ the lower $1 - \sigma$ value and use this to set the error bars on the data.

Note that the value of N controls the degree of Gaussianity in the data, as N is finite the distribution of error is not perfectly Gaussian. We will use a value of $N = 21$ throughout. This way of simulating the data is not supposed to be realistic, it is simply a way to impose imperfections on the data. It is however more realistic than assuming we have data with perfectly symmetric error bars with means centred exactly on the underlying model. It still results in rather idealised data because we usually have fewer observations at high redshifts making it unlikely that the redshift points of a realistic survey will be uniformly distributed in the interval $[0, z_{max}]$. Note that, with this method, there are also no outliers.

For the test we perform below, we simulate 50 data points with $z_{max} = 2$, $\alpha = 0.5$ and $\delta_{H_{||0}} = 10\%$ and $\delta_{\rho_0} = 25\%$ around our reference Λ CDM model defined by the parameters in Table 2.1. Note that such small errors on $H_{||0}$ and ρ_0 are optimistic and have been chosen for reasons explained in § 6.1. We then smooth these data using GPR as described in §5.2.1 and draw function realisations of $H_{||}(z)$ and $\rho(z)$ from their posterior distributions. We additionally set a Gaussian prior over the cosmological constant

$$\Lambda \sim \mathcal{N}(\bar{\Lambda}, \sigma_\Lambda), \quad \sigma_\Lambda = 0.05\bar{\Lambda},$$

where $\bar{\Lambda}$ is the value corresponding to the reference Λ CDM model of Table 2.1. The standard deviation on Λ was chosen arbitrarily and is not realistic. In § 6.1 we choose a much larger value (viz. $\sigma_\Lambda = 0.05\bar{\Lambda}$) to ensure that we have an over-dispersed prior. Note that, even though the background model is Λ CDM, the model reconstructed from random realisations of $\rho(z)$, $H_{\parallel}(z)$ and Λ is not necessarily a Λ CDM model. It will however satisfy all the constraints of a Λ TB model (see §4.5.4).

4.5.3 Three level self convergence test

Since the initial data is specified non-parametrically we do not have access to exact solutions. To test the order of convergence without knowledge of the exact solution we successively refine the grid in three steps. The CIVP is first solved using initial grid spacings of Δw_1 and Δv_1 in time and space respectively. Denoting this solution by S_1 we should expect, for a p -th order scheme, to find that (note that the CFL condition implies $\Delta v > \Delta w$)

$$S_1 - S = \alpha \Delta v_1^p, \quad (4.46)$$

where S is the unknown exact solution and $\alpha \sim \mathcal{O}(1)$ is a real coefficient. The grid spacing is then halved $\Delta w_2 = \Delta w_1/2$ and $\Delta v_2 = \Delta v_1/2$ to get the solution S_2 , for which we expect to find

$$S_2 - S \propto (\Delta v_2)^p \propto \left(\frac{\Delta v_1}{2}\right)^p. \quad (4.47)$$

Finally we generate a reference solution S_r by halving the grid spacings again. For this solution we should have

$$S_r - S \propto (\Delta v_r)^p \propto \left(\frac{\Delta v_2}{2}\right)^p \propto \left(\frac{\Delta v_1}{4}\right)^p. \quad (4.48)$$

To test the overall convergence order of the scheme we then form the error ratio

$$R = \frac{\|S_1 - S_r\|_2}{\|S_2 - S_r\|_2} = 2^p - 1 \quad \Rightarrow \quad p = \log_2 |R + 1|, \quad (4.49)$$

where $\|\cdot\|_2$ denotes the standard Euclidean 2-norm and we compare the solutions on the coarsest grid (i.e. at points separated by Δv_1). We can in principle perform this test on any of the reconstructed functions on any earlier PLC. We have chosen to perform the convergence test on T_1 and T_2 since these are the quantities that our proposed test of the CP relies on. The test is performed on both the initial PLC and one approximately 3 Gyr in the past. This tests both the spatial and the temporal order of convergence.

The target redshift values z_p at which we reconstruct these functions consist of NJ values uniformly spaced between $z = 0$ and $z = 2$, where NJ is the number of points chosen for the reference solution S_r . For each set of initial data, $H_{\parallel}(z)$, $\rho(z)$ and Λ , we perform the integration

	T_{1i}	T_{1f}	T_{2i}	T_{2f}
\bar{p}	2.009	2.0018	2.01	1.60
σ_p	0.010	0.0056	0.13	0.26

Table 4.1: Mean \bar{p} convergence factors and their standard deviations σ_p . The subscripts i and f denote initial and final PLC's respectively.

with three different grid spacings corresponding to $NJ, NJ/2$ and $NJ/4$ grid points. Since $v(0) = 0$ the spatial grid spacing on the finest grid is

$$\Delta v = \frac{v_{max}}{NJ - 1}. \quad (4.50)$$

The number of time grid points NI should be chosen so that (4.44) is satisfied i.e.

$$\frac{\Delta w}{\Delta v} \leq 0.5, \quad \Rightarrow \quad \Delta w \leq \frac{0.5v_{max}}{NJ - 1}. \quad (4.51)$$

Erring on the side of caution we therefore choose

$$\Delta w = \frac{v_{max}}{3(NJ - 1)} \quad \Rightarrow \quad NI = \text{ceil} \left(\frac{3(NJ - 1)(t_0 - t^*)}{v_{max}} + 1 \right), \quad (4.52)$$

where $\text{ceil}(\cdot)$ is the ceiling function which rounds up to the nearest integer. We repeat this procedure 1000 times and compute the convergence factor p using (4.49). We report the average convergence factor and its standard deviation in table 4.1. We see that the reconstructed value of $T_1 \propto \Delta v^2$ on both the PLC0 and the PLCF. The degradation in the time convergence factor of T_2 could be a due to the large number of terms involved in its computation. However it is second order on the PLC0 and, at the very worst, first order on the PLCF.

4.5.4 ALTB consistency relation

Without analytic solutions we cannot test the consistency of the CIVP since we can't verify that the numerical solution approaches the exact solution as the grid is refined⁵. However, there is an alternative way to test the consistency and stability, and hence verify convergence, of the CIVP. This is possible because we haven't used one of the field equations (i.e. (4.1)) in solving the CIVP. Since the solution can be found without using (4.1) there should be a way to relate it to the others. However this relation is not immediately obvious (note the factors of \ddot{D} and \dot{A}) so we will turn it into a consistency relation by writing it in the form

$$E = \frac{1}{2}AA''D - 2\ddot{D} + A'\dot{D} - D'\dot{A} + AA'D' - \frac{1}{4}\kappa\rho D \left(\frac{1}{u^2} + u^2A^2 \right) + \Lambda AD. \quad (4.53)$$

⁵Note that the solution from the new version of the CIVP always agrees, within the expected tolerance, with that developed in [131].

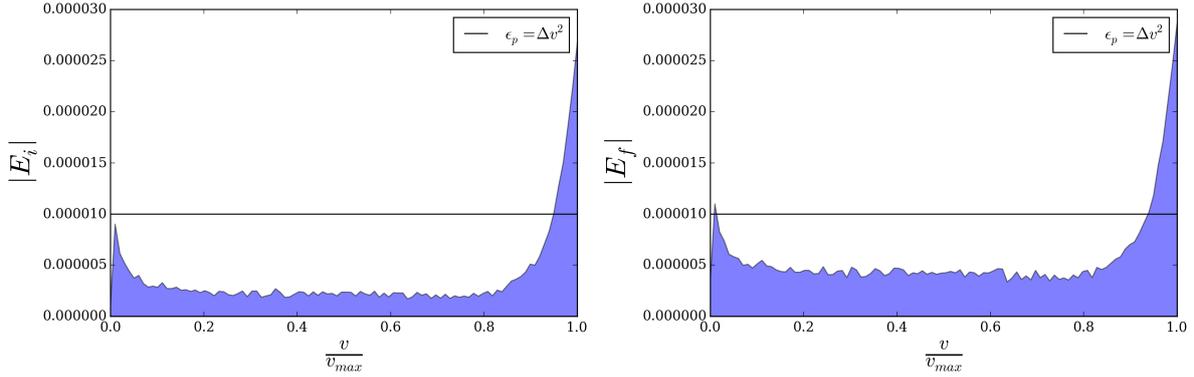


Figure 4.3: The upper bound on the consistency relation (4.53) for 1000 integrations of random realisations of $H_{\parallel}(z)$, $\rho(z)$ and Λ . Left) The the upper bound on $|E|$ on the PLC0. Right) The the upper bound on $|E|$ the PLCF. The horizontal line is the expected error of the numerical scheme viz ϵ_p .

Any instabilities and inconsistencies should then result in $E \neq 0$. Note that this constraint can be violated if either, there are errors in the initial data or, the numerical scheme is not stable and consistent.

Errors in the initial data could result in one of two ways. The first is unavoidable and results from the interpolations and numerical integrations involved when converting functions of z into functions of v . These errors are expected to be small however, if the numerical scheme is not stable, they will grow exponentially and degrade the solution in the interior of the PLC. The second relates to our discussion in §3.5. There we suggested that $H_{\parallel}(z)$ and $\rho(z)$ can, to some extent, be specified freely on the PLC0. If this is not the case we should find, already on the PLC0, that $E \neq 0$. Of course in practice we should expect to find $E \approx \mathcal{O}(\delta v^2)$ for a second order scheme. In Figure 4.3 we show the upper bound on $|E|$ found for the 1000 random realisations of $H_{\parallel}(z)$, $\rho(z)$ and Λ of the above convergence test. Note that the maximum error sometimes peaks slightly above ϵ_p towards the edges of the computational domain. This is most likely because both \ddot{D} and \dot{A} are found numerically using finite differencing. Note that we are already using one sided finite differences towards the inner $v = 0$ and outer $v = v^*$ radial boundaries of the computational domain. However $|E|$ is at least of the same order as ϵ_p on both the PLC0 and the PLCF. Since $E_i \sim \mathcal{O}(\epsilon_p)$ on the PLC0 the initial data must be valid. This verifies that $H_{\parallel}(z)$ and $\rho(z)$ can be specified independently on the PLC0, at least for the samples we considered here. Also, since $E_f \sim \mathcal{O}(\epsilon_p)$, the numerical scheme seems to be stable. Of course these results only hold for the 1000 realisations that were actually tested and within the domain on which the problem was solved. Any numerical scheme will have a stability limit and the

CIVP is no different. We found that trying to push the temporal integration too far into the past can result in numerical instabilities. There is no clear limit on how far back into the past we can go. However we found that the CIVP is always stable for an integration domain which satisfies $w^0 - w^{NI-1} \leq 5$ Gyr. This is discussed further in §5.3.3. In §6 we will monitor the consistency relation (4.53).

5

Statistical methodology

The aim of this chapter is to provide an overview of our statistical analysis. Since we work exclusively in a Bayesian framework we start by developing notation and introducing some concepts in Bayesian statistical modelling with particular emphasis on applications of the Gaussian distribution. Bayesian modelling is a very flexible framework with diverse applications in a variety of fields. The particular way in which it is utilised depends very strongly on the context. As a result our interpretation of the definitions made below is inevitably biased to our intended application and should not be taken as generic.

5.1 Bayesian statistics

5.1.1 Probability

The joint probability of a finite dimensional vector $\mathbf{y} = [y_1, y_2, \dots, y_D]$ of dimension D is denoted $p(\mathbf{y})$. The set \mathbf{y} can be partitioned into a number of disjoint subsets, each playing a different part in the model. This simple operation is at the heart of Bayesian modelling. To illustrate the general idea we partition \mathbf{y} into two disjoint subsets \mathbf{y}_A and \mathbf{y}_B and write their joint probability distribution as $p(\mathbf{y}_A, \mathbf{y}_B)$. The marginal probability of \mathbf{y}_A is defined as

$$p(\mathbf{y}_A) := \int p(\mathbf{y}_A, \mathbf{y}_B) d\mathbf{y}_B. \quad (5.1)$$

The marginal probability expresses the joint probability of \mathbf{y}_A regardless of the possible values that \mathbf{y}_B can take. If the parameters are independent the act of marginalisation has no effect. In that case the joint probability is simply the product of the marginals i.e. $p(\mathbf{y}_A, \mathbf{y}_B) = p(\mathbf{y}_A)p(\mathbf{y}_B)$. Another concept that plays an important role in Bayesian modelling is the conditional probability. The joint probability distribution of \mathbf{y}_A conditioned on the values of \mathbf{y}_B is

defined as

$$p(\mathbf{y}_A|\mathbf{y}_B) := \frac{p(\mathbf{y}_A, \mathbf{y}_B)}{p(\mathbf{y}_B)} \quad (5.2)$$

Whereas (5.1) allows us to describe the joint probability distribution of \mathbf{y}_A independently of \mathbf{y}_B , (5.2) describes the joint probability of \mathbf{y}_A given the values (or distributions) of $\mathbf{y}_B = Y_B$. If \mathbf{y}_A and \mathbf{y}_B are independent then these two operations are equivalent and the conditional and marginal distributions will be identical. Note that we can combine the two definitions (5.1) and (5.2) to rewrite the marginalised probability as

$$p(\mathbf{y}_A) = \int p(\mathbf{y}_A|\mathbf{y}_B)p(\mathbf{y}_B)d\mathbf{y}_B. \quad (5.3)$$

Bayes' theorem, which can be established from the fact that $p(\mathbf{y}_A, \mathbf{y}_B) = p(\mathbf{y}_B, \mathbf{y}_A)$ and the definition of conditional probability, states that

$$p(\mathbf{y}_A|\mathbf{y}_B) = \frac{p(\mathbf{y}_B|\mathbf{y}_A)p(\mathbf{y}_A)}{p(\mathbf{y}_B)} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}. \quad (5.4)$$

In the Bayesian interpretation we identify $p(\mathbf{y}_B|\mathbf{y}_A)$ as the likelihood, $p(\mathbf{y}_A)$ as the prior, $p(\mathbf{y}_B)$ as the evidence or marginal likelihood and $p(\mathbf{y}_A|\mathbf{y}_B)$ as the posterior probability. Note that the sets \mathbf{y}_A and \mathbf{y}_B can be further partitioned to describe nested hierarchical models. This is, at least in part, the power behind the Bayesian approach as it makes it possible to split a problem into a number of simpler sub-problems. However, in practice integrals such as (5.3) can quickly become analytically intractable, sometimes necessitating certain simplifying assumptions. Even then the complexity of the resulting models might require approximate numerical techniques to interpret. Markov-Chain-Monte-Carlo (MCMC) sampling algorithms are very powerful and flexible tools that can be used in a wide variety of problems. Before we discuss MCMC however it will be useful to review the most frequently employed simplifying assumption of them all i.e. that of jointly Gaussian random variables.

5.1.2 The multivariate Gaussian distribution

Recall that a jointly Gaussian random vector $\mathbf{y} = [y_1, y_2, \dots, y_n]$ has probability density given by

$$p(\mathbf{y}|\bar{\mathbf{y}}, \Sigma_y) = (2\pi)^{-n/2} |\Sigma_y|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \bar{\mathbf{y}})^T \Sigma_y^{-1} (\mathbf{y} - \bar{\mathbf{y}})\right), \quad (5.5)$$

where $\bar{\mathbf{y}}$ is its mean vector (of length n) and Σ its (symmetric, positive definite) covariance matrix (of size $n \times n$). This is abbreviated using the notation $\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{y}}, \Sigma_y)$. If we partition \mathbf{y} into two disjoint subsets as $\mathbf{y} = [\mathbf{y}_A, \mathbf{y}_B]$, then

$$\begin{bmatrix} \mathbf{y}_A \\ \mathbf{y}_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{\mathbf{y}}_A \\ \bar{\mathbf{y}}_B \end{bmatrix}, \begin{bmatrix} \Sigma_A & C \\ C^T & \Sigma_B \end{bmatrix}\right), \quad (5.6)$$

where Σ_A is the covariance matrix of \mathbf{y}_A , Σ_B is the covariance matrix of \mathbf{y}_B and C expresses the covariance between them. The marginal (5.1) and conditional (5.2) distributions are given by

$$\mathbf{y}_A \sim \mathcal{N}(\bar{\mathbf{y}}_A, \Sigma_A), \quad (5.7)$$

$$\mathbf{y}_A | \mathbf{y}_B \sim \mathcal{N}(\bar{\mathbf{y}}_A + C\Sigma_B^{-1}(\mathbf{y}_B - \bar{\mathbf{y}}_B), \Sigma_A - C\Sigma_B^{-1}C^T). \quad (5.8)$$

The product of two Gaussians gives another (un-normalized) Gaussian

$$\mathcal{N}(\mathbf{x}|\mathbf{a}, A)\mathcal{N}(\mathbf{x}|\mathbf{b}, B) = Z^{-1}\mathcal{N}(\mathbf{x}|\mathbf{c}, C), \quad (5.9)$$

$$\text{where } \mathbf{c} = C(A^{-1}\mathbf{a} + B^{-1}\mathbf{b}) \text{ and } C = (A^{-1} + B^{-1})^{-1}, \quad (5.10)$$

and the normalization constant also looks like a Gaussian

$$Z^{-1} = (2\pi)^{D/2} |A + B|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{a} - \mathbf{b})^T (A + B)^{-1} (\mathbf{a} - \mathbf{b})\right). \quad (5.11)$$

5.1.3 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methods [133, 134] have become increasingly more sophisticated since their introduction in [135, 136]. There exists many different MCMC algorithms, some more suitable in certain applications than others. Our exposition will be limited to some of the practical considerations involved when implementing MCMC samplers. We refer, in particular, to [133] and [137] for derivations of the theoretical results upon which these algorithms are based.

At its core there are really only two essential ingredients in a Bayesian model. The first is a model which allows us to evaluate the probability of a known quantity (frequently the observed data \mathcal{D}), conditioned on an unknown quantity (e.g. a vector of parameters θ) i.e. $p(\mathcal{D}|\theta)$. Such a model is called a likelihood model. When the data are fixed it allows us to assign a likelihood $L(\theta) \propto p(\mathcal{D}|\theta)$ to any sample of θ ¹. As an example, suppose we have measured data $\mathcal{D}_y = [x_i, y_i, \Sigma_y]$ for a function which we believe can be described by a parametric model of the form

$$y_i = f(x_i, \theta) + \epsilon \quad \text{where } \epsilon \sim \mathcal{N}(0, \Sigma_y). \quad (5.12)$$

¹Of course this is a statistical tool and should not be taken too literally. In reality the data \mathcal{D} are the true stochastic variables. As a result there is always an intrinsic uncertainty involved in observing the data. The parameters, on the other hand, are not really stochastic variables. The intrinsic uncertainty in the data does however translate to uncertainties in the true values of the parameters. The likelihood should therefore always be regarded as the likelihood of observing data \mathcal{D} given parameters θ , even when it is written as a function of the parameters. The notation and terminology used below should cause no confusion as it is simply a statistical tool with which to propagate uncertainties in the data onto uncertainties about the true values of the parameters.

Here θ is a D dimensional vector of parameters, $f(x, \theta)$ is the underlying model (or mean function) and ϵ is Gaussian noise. Since the data are fixed, the assumption that the y_i are jointly normally distributed (i.e. has mean $f(x_i, \theta)$ and covariance matrix Σ_y) allows us to assign a likelihood to a specific realisation of the parameters, $\theta = \theta^*$ say, using (5.5) i.e.

$$L(\theta^*) \propto \exp\left(-\frac{1}{2}(Y - f(X, \theta^*))^T \Sigma_y^{-1}(Y - f(X, \theta^*))\right). \quad (5.13)$$

Here we have introduced the notation $Y = [y_1, y_2, \dots, y_N]$ and $X = [x_1, x_2, \dots, x_N]$ (i.e. Y and X are numerical vectors formed by concatenating the observations y_i and the corresponding points in the domain x_i respectively) where N is the number of data points. We will often work with the negative of the log-likelihood called the potential function which, in the case of normally distributed residual vectors ϵ , is also called the χ^2 . Note that, although the data are normally distributed, the parameters are not necessarily normally distributed because the model $y(x, \theta)$ can be highly non-linear. Actually, as is often the case, we will not even have direct access to observations of the function itself but rather some means of deriving the observables from a realisation of the function. In this case we call f a latent or hidden variable.

The second ingredient is a marginal distribution $p(\theta)$ for the unknowns θ . This we call the prior. As the name suggests $p(\theta)$ expresses prior information about the parameters. Note that the “absence” of a prior really expresses our initial belief that $p(\theta)$ is a uniform distribution in the parameters (called a flat prior). Furthermore, if we are only interested in real valued parameters (as we will be), then θ can take any value in \mathbb{R}^D . Since it is impossible to sample the whole of \mathbb{R}^D in practice, some care should be taken when specifying “flat” priors.

These two ingredients allow us to infer $p(\theta|\mathcal{D})$ using Bayes’ theorem

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}. \quad (5.14)$$

The evidence $p(\mathcal{D})$ is given by the marginal distribution

$$p(\mathcal{D}) = \int p(\mathcal{D}|\theta)p(\theta)d\theta. \quad (5.15)$$

When there are multiple models to compare to the data, $p(\mathcal{D})$ can be used to perform Bayesian model selection, a topic we will only briefly touch upon. However, for fixed \mathcal{D} , the integral (5.15) is simply a normalisation constant. We may therefore write

$$p(\theta|\mathcal{D}) \propto L(\theta)p(\theta). \quad (5.16)$$

This fact can often be used to circumvent computing the integral in (5.15) if our main goal is simply to infer the overall shape, or sample from, the distribution $p(\theta|\mathcal{D})$. We now illustrate how to do this using the Metropolis-Hastings (MH) sampler [135, 136].

To implement the MH sampler we need a way to generate random realisations of θ . This is achieved by specifying a proposal distribution denoted by $q(\cdot)$. The MH sampler is a procedure which, given an initial sample $\theta_{(0)}$ and corresponding likelihood $L(\theta_{(0)})$, iterates over the following steps:

1. Propose $\theta^* = q(\theta|\theta_{(k)})$, where an index appearing as a subscript in round braces labels steps in the Markov chain.
2. Compute the acceptance probability

$$\alpha(\theta_{(k)}, \theta^*) = \min \left(1, \frac{f(\theta^*)q(\theta_{(k)}|\theta^*)}{f(\theta_{(k)})q(\theta^*|\theta_{(k)})} \right). \quad (5.17)$$
3. Accept the proposed value of θ by setting $\theta_{(k+1)} = \theta^*$ with probability $\alpha(\theta_{(k)}, \theta^*)$. Otherwise reject it by setting $\theta_{(k+1)} = \theta_{(k)}$.
4. Set $k \leftarrow k + 1$ and repeat.

This remarkably simple algorithm will converge to the target distribution $p(\theta|\mathcal{D})$ as $k \rightarrow \infty$ as long as the Markov chain is reversible. Note that the Markovian property of the chain

$$p(\theta_{(k+1)}|\theta_{(k)}, \theta_{(k-1)}, \dots, \theta_{(0)}) = p(\theta_{(k+1)}|\theta_{(k)}),$$

is ensured by choosing $q = q(\theta|\theta_{(k)})$ in the proposal step. The form of the acceptance probability ensures that the chain is reversible as it enforces the detailed balancing condition

$$f(\theta^*)q(\theta_{(k)}|\theta^*) = f(\theta_{(k)})q(\theta^*|\theta_{(k)}). \quad (5.18)$$

Furthermore, when the proposal distribution is symmetric about the current state (i.e. $q(\theta^*|\theta_{(k)}) = q(\theta_{(k)}|\theta^*)$), the acceptance probability simplifies to

$$\alpha(\theta_{(k)}, \theta^*) = \min \left(1, \frac{f(\theta^*)}{f(\theta_{(k)})} \right), \quad (5.19)$$

and we refer to the sampling scheme as the Metropolis algorithm.

In theory this procedure will converge with probability one as $k \rightarrow \infty$. This is obviously not possible in practice and we have to decide when it is safe to terminate the chain. This is the main consideration in designing MCMC algorithms and there is no definitive guideline that applies generically to all algorithms. Convergence may be delayed if:

1. The chain is initialised in a region far away from the maximum and it ends up spending a significant fraction of the total number of iterations in this region;

2. There are islands of high probability separated by very low probability regions causing the chain to get stuck in a local probability maximum for a significant fraction of the total number of iterations.

The first of these can be addressed by allowing for a so called burn-in period in which we simply discard the first portion of the chain. However there is no definitive guideline for deciding how long the burn-in period should be. The second issue is more difficult to overcome. Naively we may try making the variance of the proposal distribution large enough so that the proposed samples can jump from island to island. In practice this does not always work because proposal distributions with variance much larger than the target distribution will result in too few proposals being accepted. Note that the opposite is also true, proposals that underestimate the variance of the target will result in too many samples being accepted so that the chain will not mix well. The acceptance rate of an MCMC algorithm is defined as the fraction of accepted samples out of the total. Choosing an optimal acceptance rate (i.e. the one that leads to quickest convergence) is not always straightforward [138]. For one dimensional algorithms the optimal acceptance rate is around 44%. As the dimensionality of the algorithm increases the optimal acceptance rate drops. It can be shown that, under certain idealised circumstances, the optimal acceptance rate for an infinite number of parameters is about 23% [138]. However, even when an algorithm has a good acceptance rate, convergence after a finite number of samples is still not guaranteed. This is especially true when case 2 above applies. In such cases simple MH MCMC samplers might not be appropriate and problem specific sampling algorithms have to be considered. We will look at an alternative in §5.3.2. For now however will simply note that the convergence of any specific MCMC algorithm can usually be diagnosed. One common way to do this, and the one that we will employ, is to use the Gelman-Rubin diagnostic [139] (see [140] for some criticisms of the Gelman-Rubin diagnostic).

The idea behind the Gelman-Rubin diagnostic is to initialise multiple MCMC chains with overdispersed starting values (i.e. drawn from a distribution that overestimates the variance of the posterior) and then to check that the variance computed for each chain separately (called the within chain variance and denoted W) is the same as the total variance computed by combining all the chains (called the between chain variance and denoted B). Intuitively we should expect to find $W \approx B$ if the MCMC has converged. This can be tested for formally by constructing the potential scale reduction factor (PSRF) as follows.

Suppose we start M chains

$$\theta_{(k)}^i \quad \text{where} \quad k = 0, 1, \dots, T \quad i = 1, 2, \dots, M \quad M \geq 2, \quad (5.20)$$

where each $\theta_{(k)}^i$ is a vector containing values of the parameters for the chain i at step k . We then discard the first half of the samples and, for each scalar parameter $\theta_{j(k)}^i$, compute the means

$$\bar{\theta}_j^i = \frac{1}{n} \sum_{k=\frac{T}{2}}^T \theta_{j(k)}^i, \quad \text{and} \quad \bar{\theta}_j = \frac{1}{M} \sum_{i=1}^M \bar{\theta}_j^i, \quad (5.21)$$

so that $\bar{\theta}_j^i$ is the mean of the j -th parameter for chain i and $\bar{\theta}_j$ is the mean of the j -th parameter for all the chains. Note that $n = \frac{T}{2}$ since the first half of the chain has been discarded. We then estimate W and B using

$$W = \frac{1}{M} \sum_{i=1}^M \frac{1}{n-1} \sum_{k=\frac{T}{2}}^T \left(\theta_{j(k)}^i - \bar{\theta}_j^i \right)^2, \quad (5.22)$$

$$B = \frac{n}{M-1} \sum_{i=1}^M \left(\bar{\theta}_j^i - \bar{\theta}_j \right)^2. \quad (5.23)$$

The PSRF, which is defined as

$$\sqrt{R} = \sqrt{\frac{\frac{n-1}{n}W + \frac{1}{n}B}{W}}, \quad (5.24)$$

should then be computed for each parameter. Note that, if overdispersed starting values were used, B overestimates the variance of the chains and, while the chains have not converged, W underestimates the variance of each chain. Thus the PSRF should therefore approach 1 from above and it is considered safe to use the samples from $k = \frac{T}{2}$ onwards if $\sqrt{R} \approx 1$ for each parameter. In practice, since \sqrt{R} will never be exactly equal to 1, we terminate the chains when \sqrt{R} is close enough to 1. In our implementation we have used the threshold $\sqrt{R} < 1.1$. However, before we can even perform the Gelman-Rubin diagnostic, the MCMC has to be initialised with reasonable starting values. To find such reasonable starting values we directly supervise the MCMC during a number of trial runs. The length of each trial run will depend on the application, in most cases $T = 2500$ samples turns out to be sufficient. Note that we use Gaussian proposal distributions in all the applications so that $q(x|y) \sim \mathcal{N}(y, \Sigma_q)$ where Σ_q is the variance of the proposal distribution. It can become cumbersome to initialise Σ_q with a reasonable value when we only have vague prior information at our disposal. In general, the higher the dimensionality of the problem the more difficult it is to set Σ_q in a way that leads to reasonable acceptance rates. Gaussian process regression (see §5.2.1 below) provides a convenient way to do this when the target, denoted x (i.e. x is the quantity we want to infer from the data), is a smooth function. Once Σ_q has been specified, we also need an initial sample, \hat{x} say, which satisfies all the physical constraints of the model. When possible we will use Python's basin hopping optimisation algorithm [141] to set \hat{x} as the value of x which maximises (minimises) the likelihood (potential) function. We then initialise a trial period with starting

value $x_{(0)} \sim q(x|\hat{x})$ and inspect the acceptance rate, α_a say, after the initial $T \simeq 2500$ samples. If the acceptance rate is $\alpha_a < 0.2$ or $\alpha_a > 0.5$ we adjust Σ_q and repeat the procedure until we find $0.2 < \alpha_a < 0.5$. At this we can initialise multiple MCMC chains and perform the Gelman-Rubin diagnostic to determine when it is safe to stop sampling. This is done by initialising 10 MCMC chains and computing the PSRF (5.24) of each parameter of the chain after every $T = 2500$ samples. To ensure that the chains are initialised with overdispersed starting values, we draw $x_{(0)}$ from a distribution which slightly overestimates the expected variance of the posterior, using a variance of $2\Sigma_q$ say. Once a PSRF of $\sqrt{R} < 1.1$ is found for each parameter of the chain, we discard the first half of the samples and continue sampling until we have at least 10^4 samples for each chain. The posterior distributions we report are therefore computed with at least 10^5 samples. We will use this strategy to implement all our MCMC algorithms. We will now illustrate this with an example, the aim of which is to highlight some of differences between the observational and comoving formalisms.

Example 5.1.1. Fitting LTB to data:

Dropping the parameter Λ in the model outlined in §2.3.2 gives the LTB model [26]. Initial data for the model can be specified in terms of $H_{\perp 0}(r)$ and $\Omega_{m0}(r)$. We follow the simple void parametrisation used in [142] viz.

$$H_{\perp 0}(r) = H_o + (H_i - H_o) \left(\frac{1 - \tanh\left(\frac{r-r_0}{2\Delta r}\right)}{1 + \tanh\left(\frac{r_0}{2\Delta r}\right)} \right), \quad (5.25)$$

$$\Omega_{m0}(r) = \Omega_o + (\Omega_i - \Omega_o) \left(\frac{1 - \tanh\left(\frac{r-r_0}{2\Delta r}\right)}{1 + \tanh\left(\frac{r_0}{2\Delta r}\right)} \right). \quad (5.26)$$

The six parameters in the model, which must all be strictly positive, are:

Ω_i - The value of Ω_{m0} at the origin.

Ω_o - The value of Ω_{m0} asymptotically far away from the origin. For simplicity we demand asymptotic flatness and set $\Omega_o = 1$ throughout.

H_i - The value of $H_{\perp 0}$ at the origin.

H_o - The value of $H_{\perp 0}$ asymptotically far away from the origin.

r_0 - The width of the void.

Δr - The peakedness of the void.

Note that this is a subclass of the ALTB solution. Given values for the parameters above the LTB solution is obtained as outlined in § A. Observables are found as a function of redshift by solving (2.70) for the $t(z)$ and $r(z)$ relations. We will use the data discussed in §3 viz. Union 2.1 distance modulus $\mu(z)$ [4], cosmic chronometer expansion rate $H_{\parallel}(z)$ [87] and the minimum bound t_{min} [127] on the current age of the universe. The $\mu(z)$ and $H_{\parallel}(z)$ data are shown in

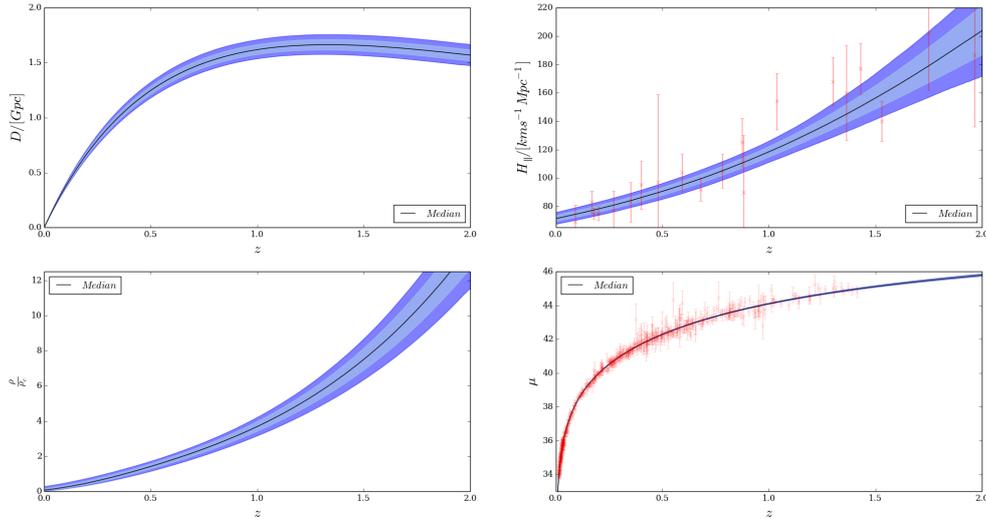


Figure 5.1: Observables for the parametric LTB model.

Figure 5.1 and we use the value $t_{min} = 15.2 \pm 2.7 \text{Gyr}$. The potential (or likelihood) is computed as²

$$\chi^2 = \sum_y \chi_y^2, \quad (5.27)$$

where y labels the data set and we compute the χ_y^2 for each of $\mu(z)$ or $H_{\parallel}(z)$ using an equation such as (5.13). Since the techniques described in §3.6 only places a lower bound on the current age of the universe, it should only contribute to the likelihood when $t_0 \leq t_{min}$. To implement this we use a one-sided version of (5.13) viz.

$$L(t_0) = \Pi(\bar{t}_{min} - t_0) \exp\left(-\frac{(t_0 - \bar{t}_{min})^2}{2\sigma_{t_{min}}^2}\right), \quad (5.28)$$

where $\Pi(\bar{t}_{min} - t_0)$ is the step function (i.e. $\Pi(\bar{t}_{min} - t_0) = 1$ if $t_0 \leq \bar{t}_{min}$ and zero otherwise). The MCMC is performed using the strategy described above with a standard multivariate Gaussian as the proposal distribution for the parameters. We show the posterior distributions of $D(z)$, $H_{\parallel}(z)$, $\rho(z)$ and $\mu(z)$ generated by the MCMC in figure 5.1. The blue shaded contours are found by constructing the empirical distribution function at each point in the domain. Note that the input functions $H_{\perp 0}(r)$ and $\Omega_{m0}(r)$, defined by (5.25) and (5.26) respectively, are not directly observable i.e. this is a latent variable model. This is one of the main drawbacks of the comoving formalism in which priors have to be specified on some constant time slice Σ_t and are therefore necessarily not directly observable. In principle these priors could also be specified non-parametrically. Such an approach has been undertaken for a Λ LTB model in [143]. While

²Samples in which any of the parameters turn out to be negative are rejected i.e. $\chi^2 = \infty$.

this approach has intuitive appeal, it requires fixing the priors from a priori knowledge and is therefore not compatible with the bottom up approach to cosmology. As we will see in §6.2 the current lack of model independent density data also necessitates that we fix the prior over $\rho(z)$ in a rather ad-hoc and cumbersome manner. However, the algorithm we describe will be straightforward to adapt when such data become available.

5.2 Gaussian processes

Gaussian processes have recently become quite popular as a means to perform non-parametric regression on cosmological data sets (see for example [144, 145, 146]). In its most basic form a Gaussian process (GP) is a collection of random variables, any finite subset of which have a joint Gaussian distribution [147]. A GP can be completely characterised by specifying its mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \tilde{\mathbf{x}})$ functions. The mean and covariance function of a real process $f(\mathbf{x})$ are defined by

$$m(\mathbf{x}) = \mathbb{E}_{\mathcal{N}}[f(\mathbf{x})], \quad (5.29)$$

$$k(\mathbf{x}, \tilde{\mathbf{x}}) = \mathbb{E}_{\mathcal{N}}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\tilde{\mathbf{x}}) - m(\tilde{\mathbf{x}}))]. \quad (5.30)$$

This is conveniently abbreviated using the notation $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \tilde{\mathbf{x}}))$. Note that the notation $\mathbb{E}_{\mathcal{N}}[\cdot]$ is used to denote the expectation value with respect to a Gaussian distribution. GP's are employed in a wide variety of applications (see [148, 149, 150, 151, 152, 153, 154] for example). Here we use them to perform non-parametric regression on discrete data sets of the form

$$\mathcal{D}_y = [z_i, y_i, \Sigma_y], \quad i = [1, 2, \dots, n], \quad (5.31)$$

where the z_i are the values in the domain of the function at which we have data (i.e. the redshift), the y_i are the observed values of the function and Σ_y is the covariance matrix of the data.

5.2.1 Gaussian process regression

The Gaussian process regression (GPR) problem aims to reconstruct the posterior distribution of a function, f say, from a number of discrete observations as in (5.31). Under the assumption of spherical symmetry, astrophysical observables are reported as functions of the redshift. Thus we will only be concerned with one dimensional regression problems and we will consistently use z to denote the variable in the domain of the function. As a simple illustration, suppose we seek the underlying function $f(z)$ in

$$y = f(z) + \epsilon, \quad \text{where } f(z) \sim \mathcal{GP}(m(z), k(z, \tilde{z})) \quad \text{and} \quad \epsilon \sim \mathcal{N}(0, \Sigma_y). \quad (5.32)$$

Denoting the points at which the function is to be reconstructed by z_p , means that the finite collection of random variables whose joint Gaussian distribution we want to infer from the data is $f_p = f(z_p)$. This is achieved by specifying the joint distribution between the observations and the random variables f_p . Suppose we select a finite subset of function values and partition them as $[Y, f_p]$, where $Y = [y_1, y_2, \dots, y_n]$ is a numerical vector formed by concatenating all the observations y_i of the function. Under a GP prior, the joint distribution of the data and the f_p is a Gaussian distribution of the form

$$\begin{bmatrix} Y \\ f_p \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(X) \\ m(X_p) \end{bmatrix}, \begin{bmatrix} k(X, X) + \Sigma_y & k(X, X_p) \\ k(X_p, X) & k(X_p, X_p) \end{bmatrix} \right), \quad (5.33)$$

where $X = [z_1, z_2, \dots, z_n]$ is the vector formed by concatenating all the inputs and X_p is the vector formed by concatenating all the points at which we wish to reconstruct the function z_p . For brevity we sometimes omit the arguments of the mean and covariance functions and refer, respectively, to $M = m(X)$ and $K = k(X, X)$ as the vector and matrix formed by evaluating the mean and covariance functions at the points X , and similarly $M_p = m(X_p)$, $K_p = k(X, X_p)$, $K_p^T = k(X_p, X)$ and $K_{pp} = k(X_p, X_p)$.

The ‘‘trick’’ employed in GPR is to specify a sufficiently generous prior distribution and then restrict the f_p to be compatible with the observations by using (5.7) and (5.8) to form the conditional distribution i.e.

$$f_p \mid X, Y, X_p \sim \mathcal{N}(\bar{f}_p, \text{cov}(f_p)), \quad \text{where} \quad (5.34)$$

$$\bar{f}_p = M_p + K_p^T K_y^{-1} (Y - M), \quad (5.35)$$

$$\text{cov}(f_p) = K_{pp} - K_p^T K_y^{-1} K_p. \quad (5.36)$$

Here $\bar{f}_p = \mathbb{E}_{\mathcal{N}}[f_p | X, y, X_p]$ is the posterior mean, $\text{cov}(f_p)$ the posterior covariance matrix and $K_y = K + \Sigma_y$. Eqns (5.34) - (5.36) are the key predictive equations for GPR. This remarkably simple procedure can accommodate a very large class of functions. The properties of these functions are determined by the mean and covariance functions.

When performing GPR everything hinges on the form of the prior distribution and the quality of the data. If we specify an incorrect prior from the outset then GPR will perform poorly regardless of the quality of the data. Note that, as evidenced by equation (5.36), the form of the prior mean function doesn’t enter into the computation of the posterior covariance matrix at all. Its role is basically to zero the data. It can significantly alter the results in regions where the data are sparse. In particular, in absence of a prior mean function (i.e. $m(\cdot) \equiv 0$), it is not recommended to try and predict values of the function beyond, or near, the edges of the data set. This is because in such regions the contribution of the prior to the likelihood will dominate

and therefore tend to bias the predictions towards zero. These “edge effects” are unavoidable without using a mean function. Note that the prior will also dominate the likelihood in regions where the data are sparse compared to the length scale of the posterior covariance function. It is easiest to illustrate this with an example. One popular covariance function is the squared exponential, which takes the form

$$k(z, \tilde{z}) = \sigma_f^2 \exp\left(-\frac{(z - \tilde{z})^2}{2l^2}\right). \quad (5.37)$$

This is an example of a stationary covariance function i.e. it only depends on the difference $z - \tilde{z}$. Note that we do not explicitly indicate the functional dependence on the parameters σ_f and l , which allow for more flexibility in the prior. Parameters that are required in the specification of the prior (here σ_f and l) are called hyperparameters and we will denote them collectively by θ . With the form (5.37), the roles of l and σ_f can be interpreted as follows:

- The parameter l is called the length scale. It sets a characteristic length which determines the strength of correlation between function values for a given separation between two points in their domain.
- The parameter σ_f is called the signal variance. It constrains the amount by which a function can change for a given l and separation between points in the domain.

We now come back to the statement made above regarding the role of the mean function when the data are sparse compared to the length scale of the posterior covariance function. In light of (5.37), or in fact any stationary covariance function, we now see that input points separated by distances $\gg l$ will convey negligible information about the covariance between the values of the function at these points. An implication for cosmology is, for example, that we could not meaningfully incorporate the angular diameter distance at decoupling $D(z_{dec})$ without specifying a prior mean function. However, when sufficient data are available, the form of the mean function is not very important and it is often safe to assume a zero mean function. The form of the covariance function however can be crucial, and so are the means by which we adapt it to the data.

5.2.2 Marginalisation and the choice of covariance function

The hyperparameters are not arbitrary and must be learnt from the data. This is where Bayes’ theorem is employed. In the GPR context it takes the following form

$$p(f|Y, X) = \frac{p(Y|f, X)p(f|X)}{p(Y|X)}. \quad (5.38)$$

Here the likelihood $p(Y|f, X)$ is given by

$$p(Y|f, X) = (2\pi)^{-\frac{n}{2}} \Sigma_y^{-\frac{1}{2}} \exp\left(-\frac{(Y-f)^T \Sigma_y^{-1} (Y-f)}{2}\right). \quad (5.39)$$

The prior is also a Gaussian and, with the understanding that there is an implicit dependence on the choice of mean and covariance function, can be written as

$$p(f|X) = (2\pi)^{-\frac{n}{2}} K^{-\frac{1}{2}} \exp\left(-\frac{(f-M)^T K^{-1} (f-M)}{2}\right). \quad (5.40)$$

Thus the marginal likelihood associated with GPR is given by the integral

$$p(Y|X) = \int p(Y|f, X) p(f|X) df. \quad (5.41)$$

Using (5.9)-(5.11), or simply noting that $Y \sim \mathcal{N}(0, K + \Sigma_y)$, the marginal log-likelihood is found to be

$$\log(p(Y|X, \theta)) = -\frac{1}{2} Y^T K_y^{-1} y - \frac{1}{2} \log |K_y| - \frac{n}{2} \log(2\pi). \quad (5.42)$$

Here we show the explicit dependence on θ because (5.42) can be used, either to infer the values of θ that maximises the likelihood (i.e. the maximum a posteriori (MAP) estimate), or to marginalise over the hyperparameters. For a truly Bayesian reconstruction of a function we ideally want to marginalise over θ . This can have an especially significant effect on the shape (and therefore the derivatives) of the reconstructed function. However, when the marginal likelihood is strongly peaked, the marginalisation will not have a very significant effect. This is more so for certain choices of covariance function than for others.

The covariance function encodes assumptions about the function we wish to infer from the data. Valid covariance functions are necessarily positive semi-definite and symmetric. In our application of GPR we further restrict to stationary and isotropic (i.e depends only on $r = |z - \tilde{z}|$) covariance functions. This expresses our belief that proximity of the inputs should be the dominant feature in determining the covariance between function values. Furthermore, as we intend to use them as input to the numerical integration scheme, we want the reconstructed functions to be smooth. Smoothness relates to the differentiability of the covariance function. For example using the squared exponential covariance function (5.37) results in functions that are truly smooth i.e. infinitely differentiable. However we found that this choice was particularly sensitive to the marginalisation over θ and actually obtained better results using the Matern class of covariance functions with $\nu = 5/2$ (see [147] for more on Matern covariance functions)

$$k(x, \tilde{x}) = \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \sigma_f^2 \exp\left(-\frac{\sqrt{5}r}{l}\right). \quad (5.43)$$

Gaussian processes described by this covariance function are continuous stochastic processes, the sample paths of which are twice differentiable. This property, as well as ensuring a certain

degree of smoothness for the samples drawn from (5.34), allows us to incorporate up to second derivative observations into the smoothing process. Our reasons for choosing it are motivated more by trial and error than theoretical considerations. Model selection could in principle be used to determine optimal covariance functions for the data sets employed (see §5 of [147]).

5.2.3 Derivative observations and sampling

Sometimes, especially when smoothing multiple dependent data sets, it is necessary to incorporate certain physical constraints into the smoothing process. In general, this is not an easy task when using non-parametric smoothing algorithms. The ability to incorporate derivative observations during GPR allows at least some of these constraints to be enforced.

Recall that under the GP model the covariance between function values at two different points in the domain is determined by the covariance function

$$\text{cov}(f(z), f(\tilde{z})) = k(z, \tilde{z}). \quad (5.44)$$

Using the linearity of the differential operator, the covariance between a function and its derivative is given by

$$\text{cov}\left(f(z), \frac{\partial f(\tilde{z})}{\partial \tilde{z}}\right) = \frac{\partial}{\partial \tilde{z}} \text{cov}(f(z), f(\tilde{z})) = \frac{\partial}{\partial \tilde{z}} k(z, \tilde{z}), \quad (5.45)$$

$$\text{cov}\left(\frac{\partial f(z)}{\partial z}, \frac{\partial f(\tilde{z})}{\partial \tilde{z}}\right) = \frac{\partial^2}{\partial z \partial \tilde{z}} \text{cov}(f(z), f(\tilde{z})) = \frac{\partial^2}{\partial z \partial \tilde{z}} k(z, \tilde{z}). \quad (5.46)$$

Thus, as long as the covariance function is sufficiently differentiable, the covariance between a function and its derivatives, as well as the covariance between derivatives of the function, can be specified up to arbitrary order. Suppose we are given a data set of a function and its derivatives $\mathcal{D}_y = [y_i, dy_i, \dots, z_i, dz_i, \dots, \Sigma_y, \Sigma_{dy}, \Sigma_{ydy} \dots]$ where the notation $d^n y$ is used to denote an observation of the n^{th} derivative of the function at the input point $d^n z$. In the GPR context we can therefore specify the following prior over the observations

$$\begin{bmatrix} Y \\ dY \\ \cdot \\ \cdot \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X, X) + \Sigma_y & \partial_z K(X, dX) + \Sigma_{ydy} & \cdot & \cdot \\ \partial_z K(dX, X) + \Sigma_{ydy}^T & \partial_z \partial_z K(dX, dX) + \Sigma_{dy} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}\right). \quad (5.47)$$

If the data cross-covariance matrices (i.e. the off diagonal block Σ 's) are not available we simply assume they are zero. Actually, in this work, we assume that the only relevant contributions in the data covariance matrices are the diagonal terms. As such we assume that the error terms ϵ in (5.32) are independent but, as we have shown above, this assumption is not necessary. It

is useful to denote the full prior over the data using the notation K_y i.e. with the implicit understanding that when there are derivative observations we really mean (5.47). With this understanding the predictive formulas (5.34)-(5.36) remain unchanged.

Finally, note that the linearity of the differentiation operator can also be used to construct posterior predictive distributions for the derivatives of a function. However there is an important caveat to doing so. The textbook advice (see [147] for example) for drawing function realisations (i.e. a sample of $f(z_p)$ from (5.34)) is to perform Cholesky decomposition on the posterior covariance matrix (5.36) and then form

$$f_p = \bar{f}_p + Lu, \quad \text{where } u \sim \mathcal{N}(0, I) \quad \text{and} \quad LL^T = \text{cov}(f_p), \quad (5.48)$$

and similarly for the derivatives of the function. However there are two important things to take note of here:

- It is true that the posterior mean of the derivative of the function \bar{f}'_p is the same as the quantity found by taking the derivative of \bar{f}_p numerically (to within the expected accuracy of the numerical method of course). However, this is not the case for random realisations of the function i.e. $f'_p \sim \mathcal{GP}(\bar{f}'_p, \text{cov}(f'_p))$ is not in general the same as the quantity found by taking the derivative of a realisation $f_p \sim \mathcal{GP}(\bar{f}_p, \text{cov}(f_p))$ numerically, even when the same random seed u is used. Thus, when it is important that the function and it's derivatives satisfy certain physical constraints, we cannot sample them from their respective GPR posteriors separately. Instead the derivatives of a specific realisation f_p have to be found numerically.
- The Cholesky decomposition $LL^T = \text{cov}(f_p)$ is unstable when performed numerically. This is because $\text{cov}(f_p)$ can be very nearly singular, i.e. $\det |\text{cov}(f_p)| \approx 0$. In [147] it is recommended to add a small multiple of the identity matrix (called jitter) to $\text{cov}(f_p)$ to cure this instability. Although this works and is nearly imperceptible in most applications, it does not work when we wish to find derivatives numerically. Jitter effectively adds a small independent and identically distributed component to $\text{cov}(f_p)$ and this shows up as jagged lines in the numerical derivatives.

To draw samples from (5.34) we instead perform the eigendecomposition

$$\text{cov}(f_p) = U\Lambda U^T = (U\Lambda^{\frac{1}{2}})(U\Lambda^{\frac{1}{2}})^T,$$

where U is the matrix of eigenvectors and Λ the diagonal matrix of eigenvalues. Although the eigendecomposition is stable, it can result in eigenvalues which are smaller than machine

tolerance. In practical applications these small eigenvalues therefore have to be set to zero. This has a negligible effect on the resulting samples, which can be generated using

$$f_p = \bar{f}_p + (U\Lambda^{\frac{1}{2}})u \quad \text{where} \quad u \sim \mathcal{N}(0, I). \quad (5.49)$$

Direct substitution confirms that $\bar{f}_p = \mathbb{E}_{\mathcal{N}}[f_p]$ and $\text{cov}(f_p) = \mathbb{E}_{\mathcal{N}}[(f_p - \bar{f}_p)(f_p - \bar{f}_p)^T]$ as required (the latter follows because $\mathbb{E}_{\mathcal{N}}[u] = 0$). Note that the samples resulting from (5.49) can be differentiated numerically.

Example 5.2.1. Smoothing $H_{\parallel}(z)$ data:

Here we perform GPR on the cosmic chronometer $H_{\parallel}(z)$ data of [87]. For illustration we show the result with and without the marginalisation in figure 5.2. The optimised hyperparameters are found using Python’s basin hopping optimiser [141] subject to the constraint $\theta > 0$. For the marginalisation the MCMC is implemented using the strategy outlined in §5.1.3. Interestingly, the error contours are larger when using the optimised hyperparameters. The reason for this is that, with the limited number of data points available for $H_{\parallel}(z)$, the marginal likelihood does not contain much information. We found that the MCMC spends a significant amount of time in regions of high l and σ_f . This results in functions that seem nearly linear in the redshift range $z \in [0, 2.0]$. For this reason we have opted not to include the marginalisation over the hyperparameters in our main application presented in §6. However, from a Bayesian perspective, our conclusions would be more robust if the hyperparameters were marginalised over. With this in mind we will now construct a Bayesian model which, among other things, allows for this possibility in the future.

5.3 Inference in a non-parameteric Λ LTB universe

5.3.1 Bayesian model

In order to perform inference using the numerical Λ LTB solution discussed in §4 we need to construct a Bayesian model which makes explicit the prior and the likelihood as well as the parameters that enter the model. The scientific parameters that we wish to infer are $H_{\parallel}(z_p)$, $\rho(z_p)$ and Λ where, in accordance with the Gaussian process interpretation, we view the values of the functions H_{\parallel} and ρ at each redshift z_p as parameters. However, there are also nuisance parameters (e.g. calibration parameters that enter the likelihood during the data gathering process) and hyperparameters (i.e. those involved in the specification of the prior) that need to be marginalised over. For simplicity we will not attempt a rigorous treatment of either of these and simply indicate where they should be incorporated. We will use the following notation:

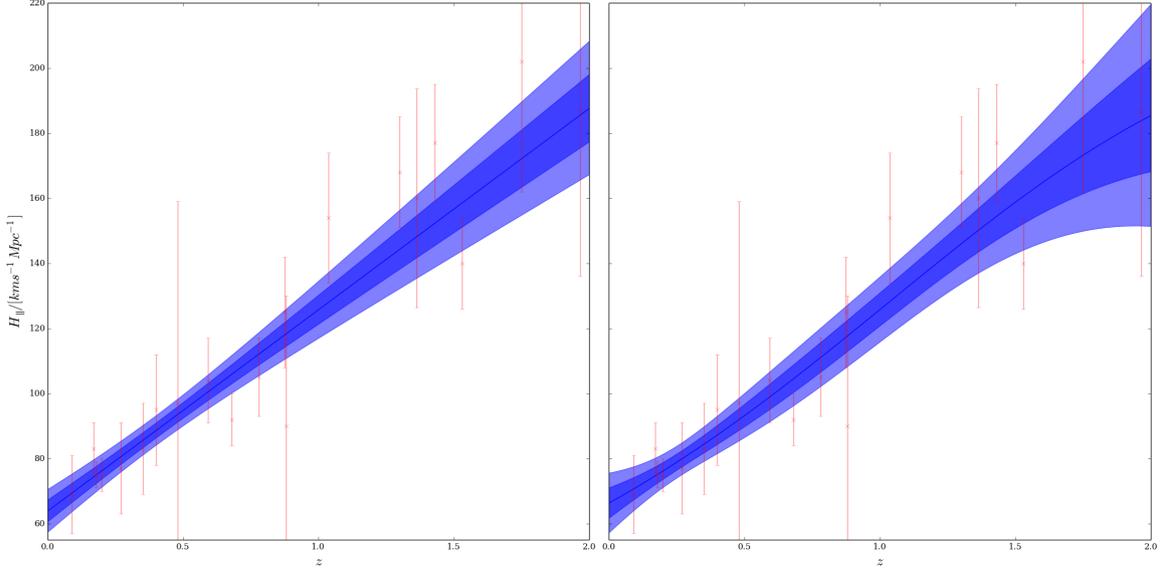


Figure 5.2: The smoothed cosmic chronometer $H_{\parallel}(z)$ data. The figure on the left shows the result of GPR when marginalising over the hyperparameters. On the right we show the result using optimised hyperparameters. This example illustrates a subtlety of using GPR viz. when the marginal posterior (5.42) does not contain much information (i.e. is not significantly peaked) then the marginalisation does not always improve the reconstruction.

- Our hypothesis \mathcal{H} is that the universe on large scales can be described by a Λ LTB model. We can therefore identify the hypothesis with a system of partial differential equations which act on the initial data, x say, to generate the observables y_i i.e.

$$y = \mathcal{H}(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma_y). \quad (5.50)$$

Here x could be a latent variable as in example 5.1.1. However, the advantage of specifying the model directly from observations on the PLC0 is that x could also be partially or fully observed. In an ideal scenario we would have observations for all of $H_{\parallel}(z)$, $\rho(z)$ and Λ . Using GPR to smooth these data would give a very efficient and robust prior over x which requires as little a priori assumptions as possible. In our case, as we only have “model independent” $H_{\parallel}(z)$ data available, it will be partially observed. Note that this idea, in particular models of the form (5.50), could be extended to more general models of the universe.

- The Λ LTB solution can be found from initial samples of $H_{\parallel}(z)$, $\rho(z)$ and Λ . Thus we form the target vector x (i.e. the random vector whose joint probability distribution we want

to infer from the data) as

$$x = \begin{bmatrix} H_{\parallel} - \bar{H}_{\parallel} \\ \rho - \bar{\rho} \\ \Lambda - \bar{\Lambda} \end{bmatrix}, \quad (5.51)$$

where, as usual, barred quantities denote the mean. We will discuss prior specification in more depth in §6.

- For purposes of illustration the full set of hyperparameters associated with the model is denoted as θ . These will typically consist of the hyperparameters of the GP priors.
- Similarly we use ψ to denote the full set of nuisance parameters associated with the model.
- Let \mathcal{D} denote all the astrophysical data at our disposal. In general \mathcal{D} will consist of multiple data sets and it is important that we have sufficient data to fully determine the model. We refer to a specific data set by using a subscript of the function e.g. \mathcal{D}_{μ} and $\mathcal{D}_{H_{\parallel}}$ refer to data sets for distance modulus and expansion rate respectively.

The goal is to infer the posterior of x under the hypothesis (5.50) with the given data i.e. $p(x|\mathcal{D}, \mathcal{H})$. Here we explicitly indicate the dependence on our assumed hypothesis since, in general, x (i.e. the initial data required to fix the model of the universe) will be different under different hypotheses. In order not to over-determine the system it is important to identify a minimal set of initial data that is sufficient for this purpose. That said, and since we only work with one single hypothesis, we will henceforth omit the explicit dependence on \mathcal{H} for notational simplicity. We now use Bayes' theorem to express $p(x|\mathcal{D})$ as

$$p(x|\mathcal{D}) = \frac{p(\mathcal{D}|x)p(x)}{\int p(\mathcal{D}|x)p(x)dx}. \quad (5.52)$$

In general the prior $p(x)$ will depend on a number of hyperparameters so that we have to marginalise $p(x|\theta)$ to find $p(x)$. Fortunately, for the modified random walk algorithm discussed in §5.3.2 below, we only need to be able to sample from $p(x)$ and won't need to explicitly compute $p(x = X)$. The marginalisation could therefore be performed by sampling the GPR marginal posterior as discussed in Example 5.2.1.

The likelihood, on the other hand, additionally depends on a number of nuisance parameters. Thus we actually have access to $L(x|\psi) = p(\mathcal{D}|x, \psi)$ and we need to be able to compute the marginal

$$L(x) = \int L(x|\psi)p(\psi)d\psi. \quad (5.53)$$

The importance of this was discussed in §3 where we pointed out that ideally the astrophysical parameters should be inferred simultaneously with the cosmology. This will become more

important for the robustness of our conclusions as the quality and availability of data improve. With the direct sampling based approach that we propose here, the likelihood could be modified, either by performing analytic marginalisation (as discussed in appendix C of [105] for example) or by a direct Bayesian marginalisation. Since the number of points z_p at which we wish to reconstruct H_{\parallel} and ρ is arbitrary, the inference framework must be able to accommodate arbitrarily high dimensional problems. This is next to impossible to do in general. However, when the target consists of a finite number of smooth functions, and possibly a handful of parameters, inference can be achieved with the following modified random walk algorithm.

5.3.2 MCMC for functional spaces

In this section we follow the presentation given in [137] to construct an efficient MCMC algorithm for our problem. We will not attempt a detailed derivation of the theoretical results. The reader can consult [137] for further details regarding the notation and terminology employed in this section.

We use $\pi(x)$ to refer to the measure on x while $\pi_0(x)$ is used to refer to the measure on the prior over x . Choosing $\pi_0(x)$ to be the dominating reference Gaussian measure of the Markov chain, Bayes' theorem is expressed with the corresponding Radon-Nikodym derivative as

$$\frac{d\pi}{d\pi_0}(x) \propto L(x), \quad \text{where} \quad L(x) = \exp(-\Phi(x)). \quad (5.54)$$

Here L is the likelihood with corresponding real valued potential Φ . The preconditioned Crank-Nicolson (pCN) proposal of [137] takes the form

$$y_{(k)} = \sqrt{(1 - \beta^2)}x_{(k)} + \beta\delta, \quad \text{with} \quad \delta \sim \mu_0(x), \quad (5.55)$$

where $0 \leq \beta \leq 1$ is a constant that can be adjusted to control the acceptance rate and, as before, we indicate the step in the chain by a subscript in braces. This proposal is a specific discretization of a class of stochastic partial differential equations which are invariant for either the reference or target measure (see [137] for the explicit construction of this proposal). Defining the acceptance probability as

$$a(x, y) = \min(1, \exp(\Phi(x) - \Phi(y))), \quad (5.56)$$

ensures that the Gaussian reference measure is exactly preserved as the likelihood drops to zero i.e. it does not reject the proposal in the case where $\Phi = 0$ but rather accepts the move with probability one. This method differs from the standard random walk method since the proposal is not centred but rather autoregressive with order one (AR(1) type). Note that the proposal is reversible with respect to π_0 so that π is indeed the stationary distribution of a Markov chain

with transition kernel defined by (5.56). Note that, although the theory required to justify this statement is non-trivial, it is remarkably simple to implement this algorithm in practice. The main differences between this and the MH sampler of §5.1.3 are that the proposal (5.55) uses samples drawn from the prior and that the acceptance rate is modified to (5.56). We can therefore use the same MCMC strategy as outlined in §5.1.3 using GPR to set Σ_q and \hat{x} as (5.36) and (5.35) respectively. We now put everything together and present an algorithm that can be used to infer the geometry of the universe directly from data.

5.3.3 The Copernicus algorithm

Let us again emphasise that spherically symmetric dust universes have two free functions. In our algorithm we select $\rho(z)$ and $H_{\parallel}(z)$ as the two free functions. Also setting the value of Λ completely specifies the model and allows the solution to be found numerically as discussed in §3. Since the entire solution is known on the grid and, using the $v(z)$ relation, can be written as a function z , we are able to assign a likelihood to the current sample of $H_{\parallel}(z)$, $\rho(z)$ and Λ by confronting this solution with the available data. Thus, given priors from which to sample $H_{\parallel}(z)$, $\rho(z)$ and Λ , we can use the above MCMC to perform inference. Prior specification is an important aspect of the algorithm and should be kept as flexible as possible. Although we will be using simple Gaussian priors in this work, the inference framework only requires that the prior is absolutely differentiable with respect to an appropriately chosen Gaussian distribution [137].

Next we need to specify the likelihood function $L(x)$. In accordance with our model hypothesis (5.50), we use a potential function Φ of the form

$$\Phi(x) = \frac{1}{2}(y - \mathcal{H}(x))^T \Sigma_y^{-1} (y - \mathcal{H}(x)), \quad L(x) = \exp(-\Phi(x)), \quad (5.57)$$

where y consists of the functions for which we have data. This is the form of a χ^2 likelihood without the degrees of freedom correction. This likelihood has been adopted for simplicity, it is not a requirement of the algorithm. The only restriction on the allowed form of the likelihood function is that it satisfies (5.54) i.e. it can be written as the Radon-Nikodym derivative of the target measure $\pi(x)$ with respect to the prior $\pi_0(x)$ where either $\pi(x)$ or $\pi_0(x)$ is absolutely continuous with respect to the dominating reference Gaussian measure. The inference framework is therefore quite general and could be adapted to perform a blind analysis in which the nuisance and hyperparameters are inferred simultaneously with the cosmology.

With the prior and likelihood specified, our final consideration is to make sure that all the physical constraints are satisfied. This is achieved simply by rejecting a sample (equivalently

setting $\Phi = \infty$) which violates any of the physical constraints. We explicitly check the following constraints on the PLC0:

- Positive energy density $\rho > 0$;
- No shell crossings $\partial_r R \neq 0$ (equivalently that $\partial_r R$ doesn't change sign)³;
- The ALTB consistency relation (4.53).

The field equations ensure that these constraints remain satisfied if they hold on the PLC0. They could however be violated in the interior if there are instabilities or convergence issues in the numerical integration scheme. As was shown in §4 the CIVP is stable and second order convergent for a large class of initial data and within the domain of integration investigated there. However, since these tests were performed on a finite number of initial samples, we also monitor the ALTB consistency relation while running the algorithm. In all our simulations we found results on the PLC0 which look virtually identical to those shown in Figure 4.3. However, we found that the stability limit of the integration scheme can sometimes be exceeded in the interior of the PLC. This can happen when the initial samples of H_{\parallel} , ρ and Λ yield a very high value of t_0 and we try to integrate very deep into the PLC, typically when the temporal domain exceeds $w^0 - w^{NI-1} \approx 7$ Gyr. However it does not happen for every integration in which the $w^0 - w^{NI-1} > 7$ Gyr and seems to be correlated with the maximum value of the radial coordinate v^* on the PLC0. Samples with a lower value of v^* on the PLC0 tend to exceed the stability limits of the code more often. A possible explanation for this is that, for a fixed redshift interval, lower values of v^* on the PLC0 results in larger derivatives of the hypersurface variables at the far end of the radial domain. This might necessitate adapting the step size of the radial integrator. Samples which violate (4.53) in the interior of the PLC by more than an order of magnitude are discarded (in the interior of the PLC only). This should be kept in mind when interpreting the posterior distributions on the inside of the PLC. Note however that this stability limit does not affect the posterior distributions of reconstructed quantities on the PLC0.

With these considerations in mind the geometry of the universe can be reconstructed from data as summarised in Algorithm 1⁴. It should be understood that each accepted sample of x corresponds to a full ALTB solution viz. $\mathcal{H}(x)$. This makes it possible to investigate any aspect of the model within the domain in which the solution has been calculated. However,

³This does not exclude any potentially relevant scenarios that are not already excluded by the fluid assumption.

⁴The code implementing the Copernicus algorithm is available for download <https://github.com/landmanbestor/Copernicus>. Documentation and examples will be made available in due course.

Algorithm 1

- 1 Set $k = 0$, choose $x_{(k)}$ and compute $\Phi(x_{(k)})$;
 - 2 Implement Procedure 1 to get $\delta \sim \pi_0(x)$;
 - 3 **if** $\text{any}(\rho < 0)$ **then** set $x_{(k+1)} = x_{(k)}$ and go to 6;
else Propose $\tilde{x}_{(k)} = \sqrt{1 - \beta^2}x_{(k)} + \beta\delta$ and implement Procedure 2;
 - 4 **if** $\partial_r R = 0$ **or** $Q > \mathcal{O}(\Delta v^2)$ **then** set $x_{(k+1)} = x_{(k)}$ and go to 6;
else Compute $a(\tilde{x}_{(k)}, x_{(k)})$ with (5.56) and sample $u \sim \mathcal{U}(0, 1)$;
 - 5 **if** $u \leq a(\tilde{x}_{(k)}, x_{(k)})$ **then** set $x_{(k+1)} = \tilde{x}_{(k)}$ and implement Procedure 3;
else set $x_{(k+1)} = x_{(k)}$ and go to 6;
 - 6 Set $k \leftarrow k + 1$ and go to 2 ;
-

to be concise, we will summarise the results for each simulation in a series of four figures (see Figures 6.1-6.4 for example). The first figure shows the posterior distribution of the observables on the PLC0. The second shows the posterior distributions of the quantities T_1 and T_2 on the PLC0 and on a PLC defined by $w \approx 10\text{Gyr}$ which we abbreviate as PLCF. We then show the posterior distributions of the metric components X and R of (2.59) and of ρ and H_\perp on a constant time slice defined by $t \approx 10\text{Gyr}$ (Σ_{t^*}). The final of the four figure series shows the joint constraints on the parameters Ω_{m0} vs. $\Omega_{\Lambda 0}$ and on t_0 vs. Λ . Note that contours on the reconstructed functions are found by constructing the empirical distribution function (*EDF*) at each point in the domain. The black line in all these figures is the median ($EDF = 0.5$) of the posterior distribution, the dark blue region is the 1- σ quantile ($0.16 < EDF < 0.84$) and the light blue region is the 2- σ quantile ($0.025 < EDF < 0.975$). Contours on the parameters are found using a kernel density estimate on the values of the parameters at the vertex q of the PLC. We use a Gaussian kernel for the kernel density estimate and set the bandwidth using Scott's rule (see [155] for example). The dark and light blue regions again correspond to the 1- and 2- σ contours.

6

The metric from observations

In this section we employ the methodology developed thus far to reconstruct the geometry of the universe directly from data. We then use the result to test the Copernican principle (CP) by reconstructing the quantities T_1 and T_2 in the consistency relations (3.31) and (3.52) respectively. We will use the solution to the CIVP to show the forms of T_1 and T_2 on the PLC0 and on the PLCF i.e. the PLC located at $w \approx 10 \text{ Gyr}^1$. It is shown that current data, although perfectly compatible with ΛCDM , are not sufficient to confirm or refute the validity of the CP. This is followed by an investigation using forecast data from future surveys. Our goal is to determine which data will best be able to constrain violations of the CP. We start this section by testing the Copernicus algorithm on simulated data.

6.1 Simulations and verification

To test our numerical implementation of Algorithm 1 we simulate data around our reference ΛCDM model defined in Table 2.1. The data are simulated according to the procedure described in §4.5.2. In what follows the simulated data could fall into one of two different categories viz. the data used to set priors and the data used to perform inference. These do not have to be mutually exclusive. Any of the available data could be used for inference as long as the data gathering process is model independent and we can correctly incorporate correlations between data sets. For simplicity we will assume that the data are independent in all the applications that follow. This would not be a satisfactory assumption if the data were precise enough to place tight constraints on allowed violations (or lack thereof) of the CP. Robust tests of the CP obviously have to account for correlations in the data.

In verifying that our algorithm performs as expected, we will be optimistic and assume that

¹Note it is not exactly 10 Gyr because we use the solution at the w grid point closest to 10 Gyr.

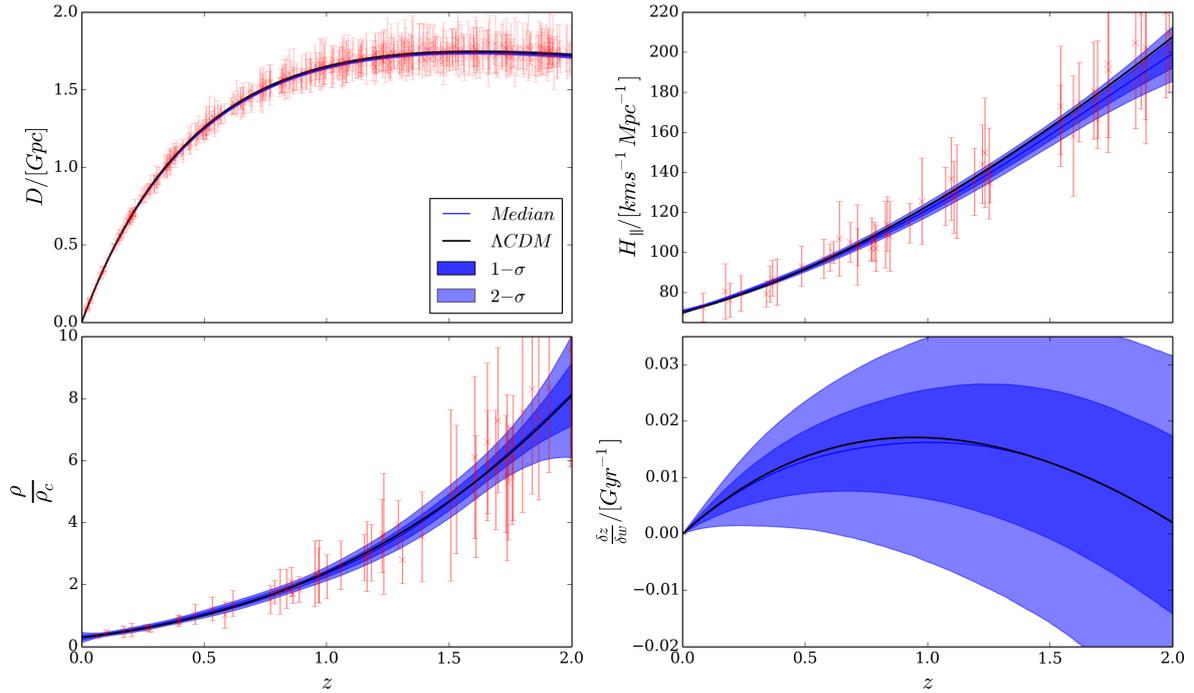


Figure 6.1: Observables on the PLC0 for simulated test data. Top Left: Posterior distribution of $\mu(z)$ on the PLC0. Top Right: Posterior distribution of $H(z)$ on the PLC0. Bottom Left: The posterior distribution of $\rho(z)$ on the PLC0. Bottom Right: The posterior distribution of the redshift drift $\frac{\delta z}{\delta w}$.

we have model independent data available for $D(z)$, $H_{\parallel}(z)$, $\rho(z)$ and t_0 . We will be even more optimistic in the way we choose the uncertainty on the simulated data points and prescribe relatively small (compared to current data) error bars. We use exactly the same simulated $H_{\parallel}(z)$ and $\rho(z)$ data as was used for the convergence test of 4.5.3. The prior over $H_{\parallel}(z)$ and $\rho(z)$ is set using GPR with optimised hyperparameters². The prior over Λ is Gaussian with mean corresponding to the reference model of Table 2.1 and an uncertainty of 60%. The large uncertainty on the prior over Λ ensures that we maintain an over-dispersed prior. Treating H_{\parallel} , ρ and Λ independently in the prior, we draw samples from $\pi_0(x)$ by sampling each of these separately and constructing the vector x as in (5.51). The constraints $\rho > 0$ and $\Lambda \geq 0$ are enforced simply by rejecting negative samples during the MCMC.

We further simulate 500 data points for $D(z)$ in the same interval i.e. $z \in [0, 2]$ with relative error governed by the power law $\frac{\sigma_D}{D} = 0.05\sqrt{1+z}$ i.e. 5% uncertainty at the origin. We also assume that t_0 can be measured with high precision and create a data point centred on the background Λ CDM value with 1- σ error bars at 2%. Algorithm 1 is run using all these data for

²Note that, since data are available over the whole input domain, there is no need to specify prior mean functions for either of these GPs.

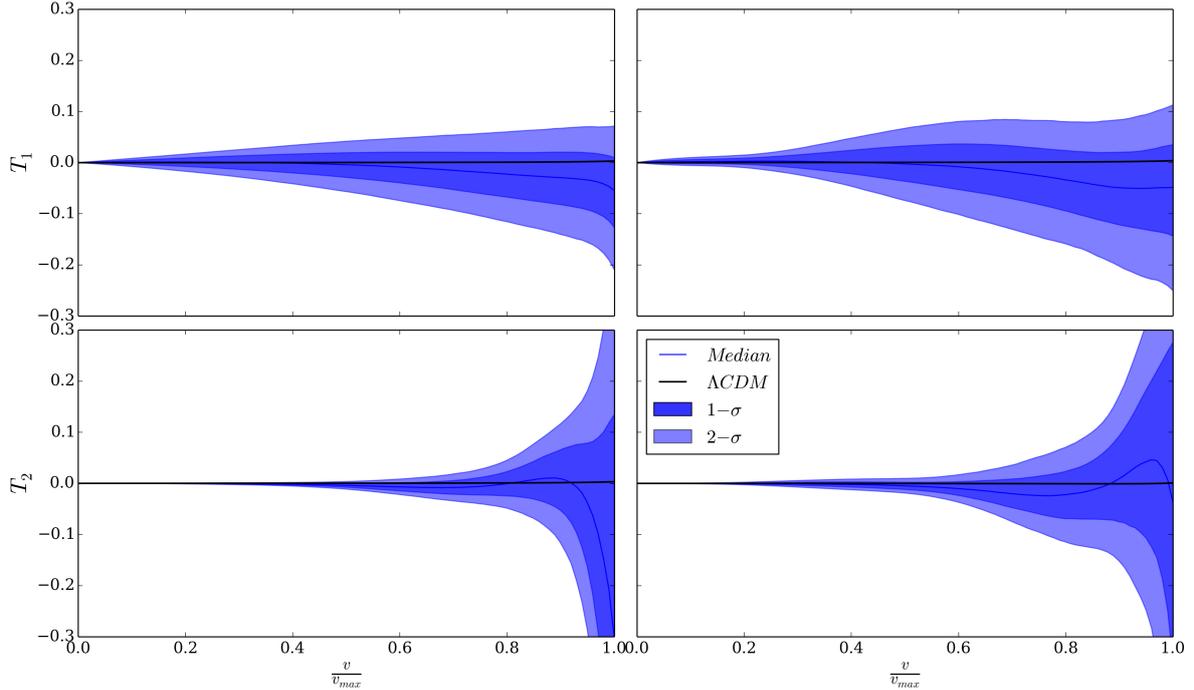


Figure 6.2: The quantities T_1 and T_2 from simulated test data plotted as functions of the normalised affine parameter $\frac{v}{v_{max}}$. The figures on the left (right) correspond to the PLC0 (PLCF).

inference and using the MCMC strategy described in §5.1.3. We found that a value of $\beta = 0.25$ in the proposal (5.55) lead to typical acceptance rates of between 25-35%. For this simulation the the PSRF (5.24) of the Gelman-Rubin diagnostic reaches a value of 1.1 after approximately 10^4 samples. For each of the 10 MCMC chains we therefore set a burn-in period of 5000 samples and then draw an additional 10^4 samples for each chain. Each simulation therefore consists of 10^5 samples in total.

The results for simulated test data are summarised in Figures 6.1-6.4. As can clearly be seen from the figures, the input model always falls within the $2\text{-}\sigma$ contours but are more often confined to the $1\text{-}\sigma$ contours. Note that, in Figure 6.1, the contours for the observables $D(z)$, $H_{||}(z)$ and $\rho(z)$ all lie well within the error bars of the data but that there is still a surprising amount of uncertainty in the redshift drift $\frac{\delta z}{\delta w}(z)$. We will investigate the possibility of including redshift drift data in 6.3. Figure 6.2 shows the posterior distributions of T_1 and T_2 on the PLC0 (left) and the PLCF (right). Note that there can still be a surprising amount of uncertainty in these quantities (especially at high redshift) even with reasonably precise observations of $D(z)$, $H_{||}(z)$, $\rho(z)$ and t_0 . Since the maximum extent of the affine parameter v_{max} varies from sample to sample, we plot T_1 and T_2 against the normalised quantity $\frac{v}{v_{max}}$. In the reference model of

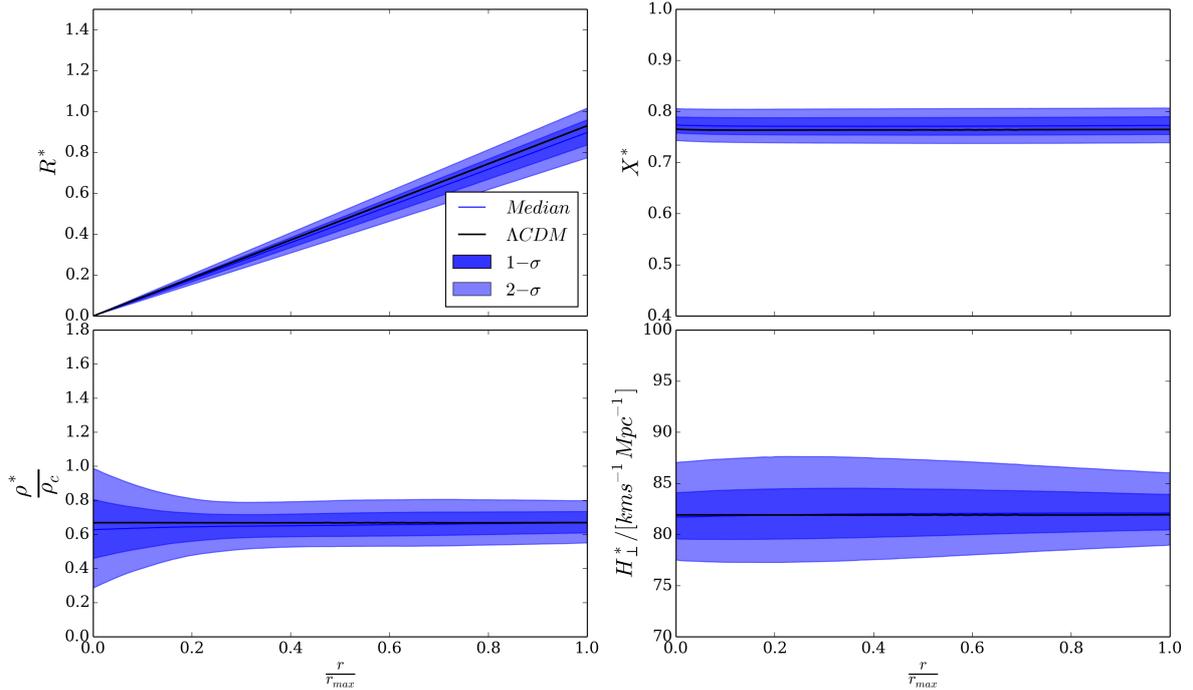


Figure 6.3: Posterior distributions on a constant time slice Σ_t defined by $t^* = 10$ Gyr for simulated data. We plot X , R , ρ and H_\perp as functions of the normalised comoving radial distance $\frac{r}{r_{max}}$. Top Left: The metric function R . Bottom Left: The density on Σ_t plotted in units of the critical density ρ_c today (see Table 2.1). Top Right: The metric function X . Bottom Right: The transverse Hubble rate H_\perp . Clearly the algorithm correctly recovers the expected forms of these functions.

Table 2.1 a value of $\frac{v}{v_{max}} = 0.5$ corresponds to $z \approx 0.4$ (see Figure 3.1). Thus there seem to be fairly tight constraints on T_2 out to $z \approx 0.4$. However these constraints are meaningless by themselves. To draw meaningful conclusions from them we would need to know the expected variance of T_2 within the concordance model. This will be discussed further in §6.4. For a comparison with the expected values of T_1 and T_2 in the best fit LTB model of Example 5.1.1 see Figure 2 in [156]. In 6.3 we will also investigate how data expected from future surveys can further tighten these constraints.

We should note an important feature that will be prevalent in all the simulations that follow. As can be seen from Figure 6.2 there is a tendency of both T_1 and T_2 towards negative values at the far end of the radial domain. At this stage it is not completely clear why this happens. An explanation that seems convincing is the presence of “edge effects” as discussed in §5.2.1. Close inspection of Figure 6.1 shows that the most likely realisations of $H_\parallel(z)$ fall slightly below the what is expected from the background model. If this tendency is due to edge effects then it

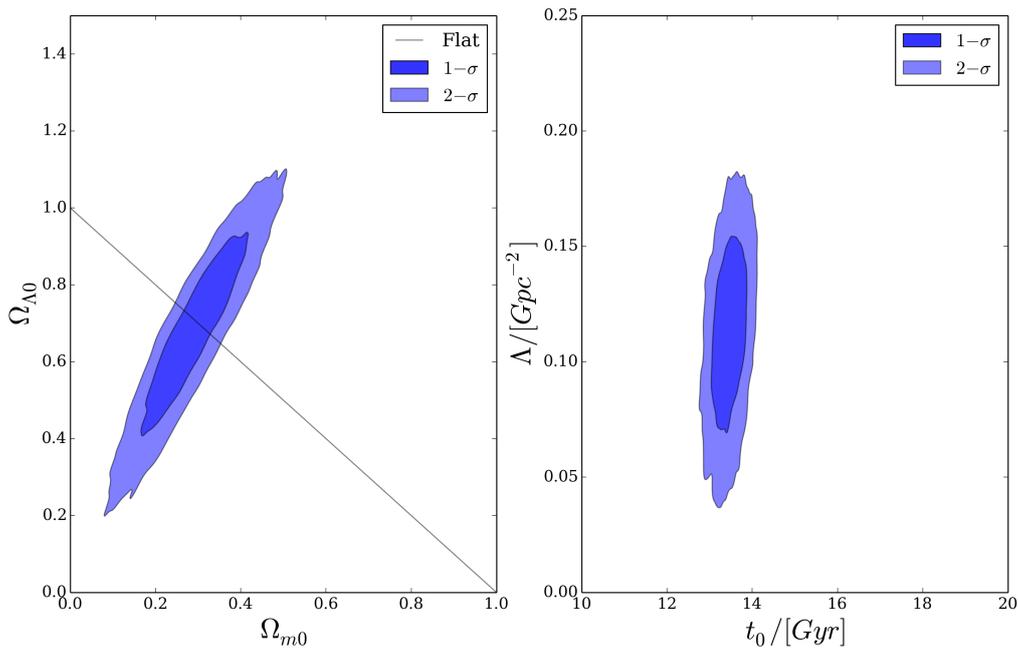


Figure 6.4: Contour plot of Ω_{m0} and $\Omega_{\Lambda0}$ at the vertex for simulated test data. Also shown is a contour plot of Λ and the age of the Universe t_0 .

would be difficult to avoid without specifying a prior mean function.

In Figure 6.3 we use the coordinate transformation to reconstruct X , R , ρ and H_{\perp} on Σ_{t^*} . Clearly the most probable profiles agree with what is expected from the background Λ CDM model. Finally in Figure 6.4 we show the joint constraints on the parameters Ω_{m0} vs. $\Omega_{\Lambda0}$ and t_0 vs. Λ at the vertex of the PLC0. These again contain the background model but they also illustrate an interesting feature of our formalism. The additional degree of freedom introduced by dropping the CP significantly relaxes the constraints on these parameters. In particular, as evidenced by the slanted Ω_{m0} vs. $\Omega_{\Lambda0}$ contours, models which have a non-negligible contribution from curvature can also fit the data with high confidence. Since we used fairly optimistic error bars on the simulated data sets, we must conclude that models with non-negligible curvature would be hard to rule out by using only $D(z)$, $H_{\parallel}(z)$, $\rho(z)$ and t_0 data. Our investigation in 6.3 aims to determine which data would be able to remove this ambiguity. However first we will see what constraints can be derived from currently available data.

6.2 Status of the Copernican principle

To run the current data through the algorithm we first need to specify the prior $\pi_0(x)$. The priors over $H_{\parallel}(z)$ and Λ are specified in the same way as for the test simulation above except

that we allow for 100% uncertainty in Λ . Unfortunately the lack of density data means we have to specify a fairly ad-hoc prior for $\rho(z)$. We will create a Gaussian prior by using mock density data. To do so we first need to specify a background mean function $\bar{\rho}(z)$. As an initial guess we choose a $\bar{\rho}(z)$ corresponding to our reference Λ CDM model defined by the parameters listed in Table 2.1. We choose 50 uniformly spaced redshift values between $z = 0$ and the maximum redshift up to which we have $H_{\parallel}(z)$ data i.e. $z_{max} = 1.965$. For an overdispersed prior the uncertainty around these data points should be chosen to overestimate that of the final reconstructed function. We assume that the relative uncertainty vs. redshift relation is governed by the power law

$$\frac{\sigma_{\rho}}{\rho} = \sigma_f(1+z)^{\alpha}. \quad (6.1)$$

Choosing values for σ_f and α would then allow us to perform GPR on the mock data and draw smooth function realisations as usual. Substituting (6.1) and the form of the covariance function (5.43) into the formula for the predictive posterior covariance (5.36), we find that

$$\sigma_f = \sqrt{2}\sigma_{\rho_0}, \quad (6.2)$$

where σ_{ρ_0} is the desired uncertainty of the prior over ρ at the origin. An estimate of this error can be found from

$$\rho(0) = -\frac{2D'''(0)}{\kappa},$$

where we have substituted the limiting expressions of D and its derivatives derived in § 4.2 into (4.5). An expression for D''' , in terms of u , H_{\parallel} and D , as well as their derivatives w.r.t. z , can be derived by applying the chain rule to the relation (3.59) for D'' and using the expression (3.20) for $\frac{dz}{dv}$. Taking the limit as $v \rightarrow 0$ gives

$$D'''_{\mathcal{C}} = \left[H_{\parallel}^3 (6D_{,z} + 6D_{,zz} + D_{,zzz}) + H_{\parallel}^2 H_{\parallel,z} (8D_{,z} + 3D_{,zz}) + D_{,z} (H_{\parallel} H_{\parallel,z}^2 + H_{\parallel}^2 H_{\parallel,zz}) \right]_{\mathcal{C}},$$

where the subscript \mathcal{C} denotes the quantities evaluated on the central worldline of the observer. To get estimates of the uncertainties in all the quantities on the RHS, we perform GPR on the data for $H_{\parallel}(z)$ and $D(z)$ separately. This allows us to reconstruct the posterior distributions of the derivatives as explained in § 5.2.3. An estimate of the error on $\rho(0)$ can then be found simply by propagating the uncertainties in their posterior distributions at $z = 0$. Such a procedure will overestimate the error on $\rho(0)$ because it assumes that the given data sets for $H_{\parallel}(z)$ and $D(z)$ are independent. With the Union 2.1 data from [4] and the cosmic chronometer data from [86] and [87] it yields a value of $\frac{\sigma_{\rho_0}}{\rho_0} \approx 0.4$. This sets the initial guess of σ_{ρ_0} and hence also σ_f^3 .

³Note that this procedure yields a wildly inaccurate value of ρ_0 (often very large and even negative). This is because we have neglected the interdependence between $H_{\parallel}(z)$, $D(z)$ and their derivatives. We therefore use the value of $\rho_0 = \bar{\rho}(0)$ i.e. the density of our reference Λ CDM model 2.1 at the vertex of the PLC0.

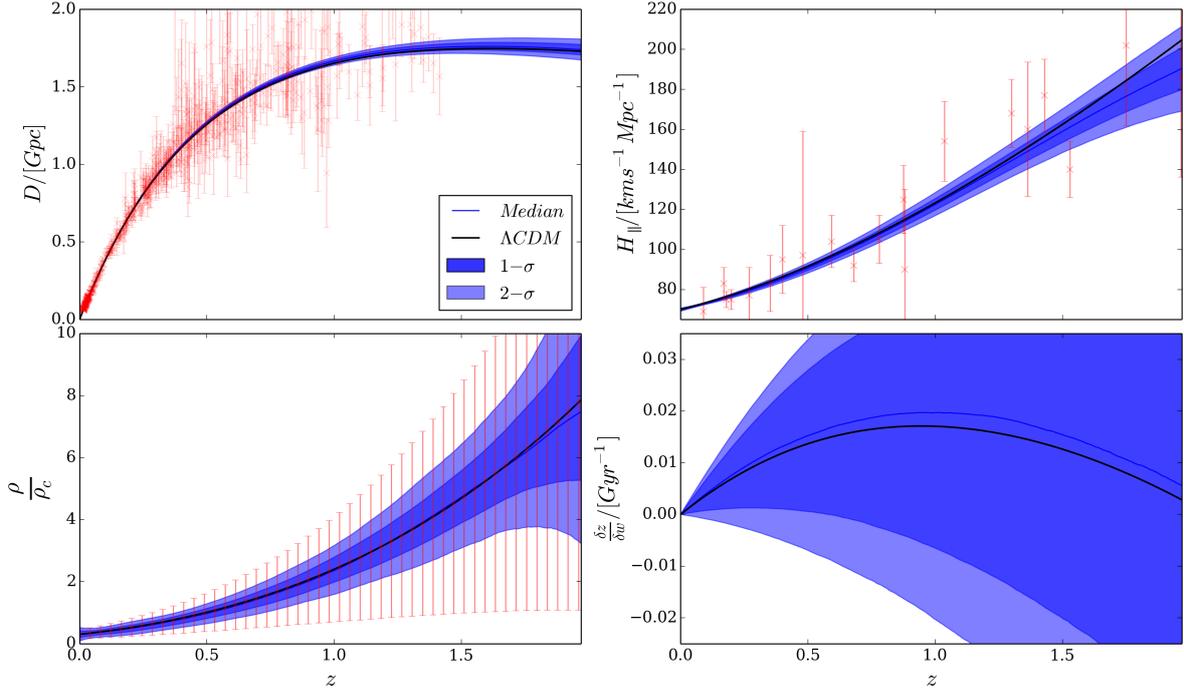


Figure 6.5: Observables on the PLC0 for currently available data. Top Left: Posterior distribution of $D(z)$. Top Right: Posterior distribution of $H(z)$. Bottom Left: The posterior distribution of $\rho(z)$. Bottom Right: The posterior distribution of $\frac{\delta z}{\delta w}$.

We initially guess a value of $\alpha = 1$ (i.e. the uncertainty in $\rho(z)$ grows linearly with redshift). The values of σ_f and α that maintain an overdispersed prior distribution are then determined by direct supervision using an iterative procedure. The procedure also requires an initial value for the length scale parameter l . For this we use the same value as the optimised value from the test simulation above. We then perform a trial run of the algorithm with ten chains of 2500 samples each. Afterwards we compare samples from the prior over $\rho(z)$ to the posterior and adjust the parameters until we find an acceptance rate of between 25-35% while also maintaining an overdispersed prior⁴. After each trial period we update the prior mean function by computing the median of the posterior samples. This procedure is repeated until the mean and contours of the reconstructed $\rho(z)$ function converge to within a tolerance of about 10%. We confirmed that initialising the procedure with a different prior mean function (i.e. that of the LTB model discussed in Example 5.1.1) does not significantly alter the result, it only alters the number of iterations required during the trial period.

This procedure might appear obscure and ad-hoc. However we found our results to be quite robust against the specified prior. Slightly altering the values of σ_f , α and l can affect the

⁴We found that the values $\sigma_f = 0.4$ and $\alpha = 0.8$ maintain an overdispersed prior when using $\beta = 0.35$ in the proposal (5.55).

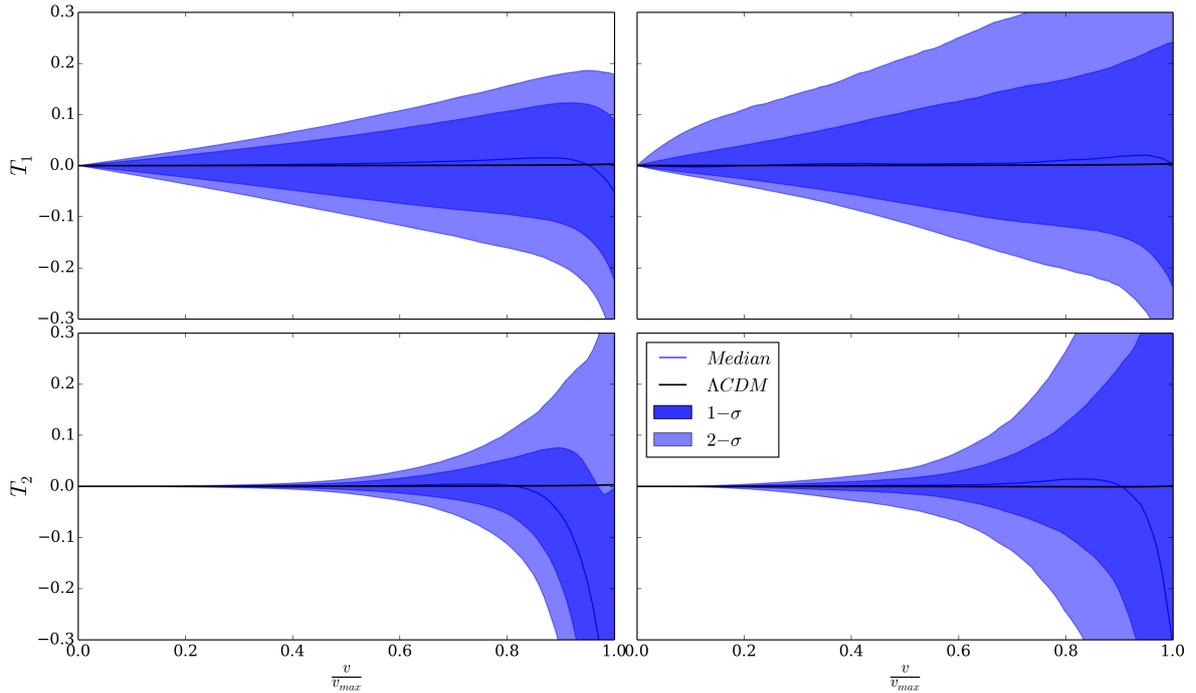


Figure 6.6: The quantities T_1 and T_2 for currently available data plotted as functions of the normalised affine parameter $\frac{v}{v_{max}}$. The figures on the left (right) correspond to the PLC0 (PLCF).

overall acceptance rate of the MCMC but it does significantly change the final distributions of quantities T_1 and T_2 . It should be noted however that this supervised process is very time consuming. It would also not be very convincing if it yielded a negative result i.e. found T_1 and T_2 significantly different from zero. More robust priors will be required to ultimately test the Copernican principle. Possible alternatives are to use sieve priors as discussed in [137] or to introduce additional noise variance hyperparameters which should then be inferred and marginalised over as discussed in 5.3. However, given the quality of current data, it is unlikely that such priors would alter our conclusions. Note that this procedure could be avoided if density data were available. Whether we use such data for inference or only to set the prior over $\rho(z)$ will depend on the intended application.

The data used for inference are Union 2.1 [4] $\mu(z)$ data, the cosmic chronometer $H_{||}(z)$ data [86, 87] and the value of t_{min} derived from nucleochronology in [127]. Since nucleochronology only places a lower limit on the age of the Universe we only allow this data point to contribute to the likelihood when $t_0 \leq t_{min}$ (i.e. using (5.28)). The results are summarised in Figures 6.5 - 6.8. Note that the tight constraints on $H_{||}(z)$ at low redshift are artificial and result from not marginalising over the choice of H_0 employed when calibrating the supernovae absolute

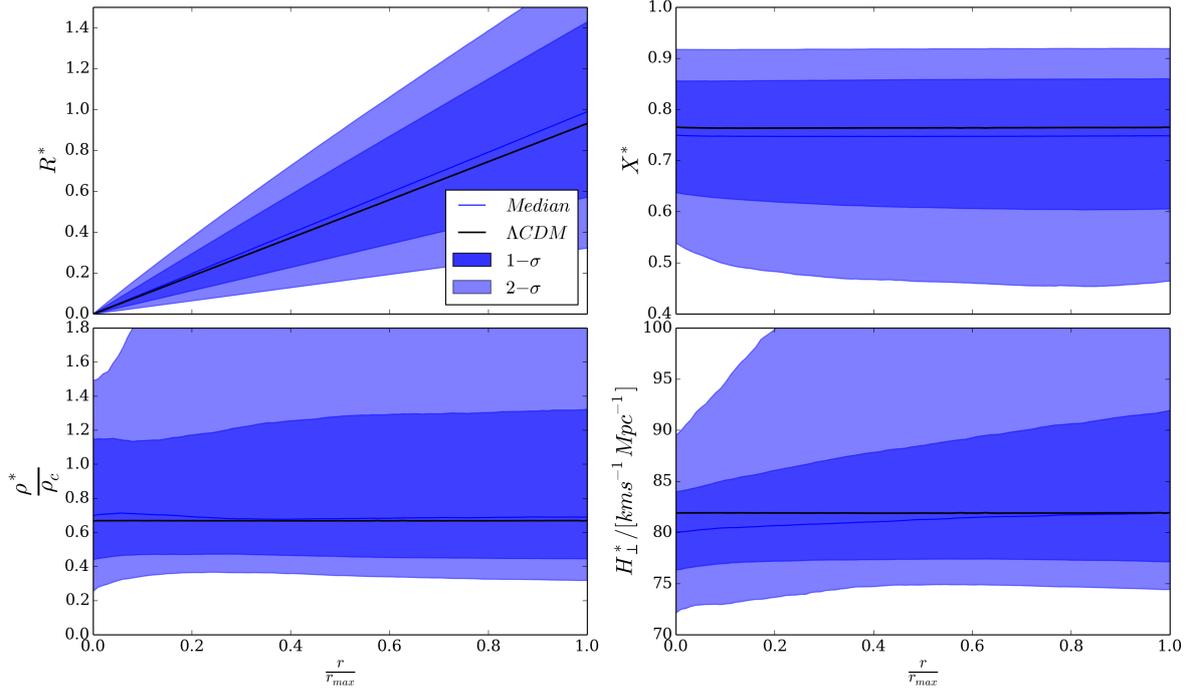


Figure 6.7: Posterior distributions on a constant time slice at $t^* = 10\text{Gyr}$ (Σ_{t^*}) for current data. These are plotted as functions of the normalised comoving radial coordinate r . Top Left: The metric function R . Bottom Left: The density on Σ_{t^*} plotted in units of the critical density ρ_c of our reference ΛCDM model (see Table 2.1). Top Right: The metric function X . Bottom Right: The transverse expansion rate H_\perp on Σ_{t^*} .

magnitude. Figure 6.6 shows that current data are perfectly compatible with ΛCDM but that large variations in the quantities T_1 and T_2 are allowed. In fact we found that, without an upper limit on t_0 , the maximum value of Λ in the simulation is basically unconstrained. This can lead to values of $\Omega_{K0} < -1$ which in turn can lead to numerical instabilities when computing t_0 using the elliptic function (2.67). We therefore also placed a hard lower limit on the dimensionless curvature parameter by rejecting samples with $\Omega_{K0} < -1$. Even with this restriction we still found that Λ was not very well bounded from above. To obtain the contour plots shown in Figure 6.8 we also implemented the prior $\Lambda < 0.23 \text{ Gpc}^{-2}$. Note that this is more than twice the value expected for a ΛCDM model defined by the parameters shown in Table 2.1. A similar restriction was enforced to obtain the contours shown in Figure 6.11. This restriction should be kept in mind when interpreting the constraints on the parameters shown in these two figures. These figures clearly indicate the inability of current data to confirm or refute the validity of the CP. Moreover, the space of allowed cosmological models is vast, with current data being perfectly compatible with $\Omega_{\Lambda 0} = 0$ models. We also see that $\Omega_{m0} \lesssim 0.6$ at $2 - \sigma$ confidence

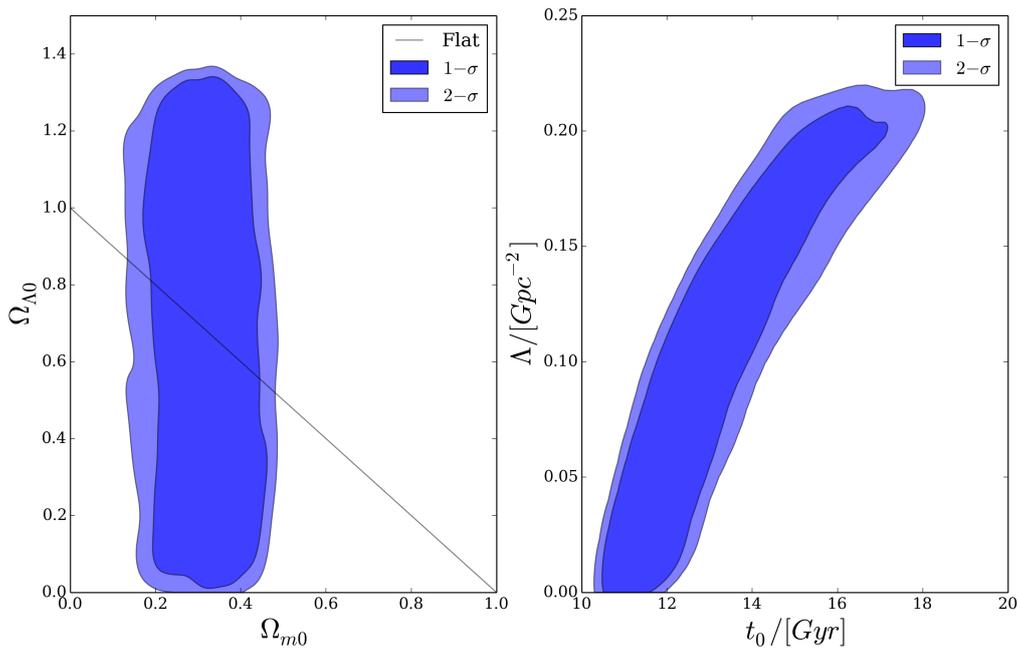


Figure 6.8: Contour plots of $\Omega_{\Lambda 0}$ vs. Ω_{m0} (left) and Λ vs. t_0 (right) at the vertex of the PLC0 for currently available data.

level. Thus there is a large class of models with non-negligible curvature which fit the data with high confidence. In particular void LTB models do not seem to be ruled out (see §6.4 for further qualification of this statement). Our results from simulated data in the previous section suggest that an accurate determination of both t_0 and ρ_0 could in principle rule out these models⁵. However, there we assumed that we also had an upper limit on the age of the Universe, something which is more difficult to obtain in practice.

6.3 Forecasts

In this section we investigate, in a fairly idealised way, which data would best be able to disentangle homogeneous models from models which allow for large scale radial inhomogeneities. Note that all systematics and possible correlations in the data have been neglected. We show the results of two simulations. The aim of the first simulation was to determine whether $D(z)$, $H_{\parallel}(z)$ and t_{min} data with the accuracy expected from future surveys could in principle be used to test the CP. These are the observables for which we can hope to generate reasonably model independent data from upcoming future surveys. Out of these three observables the value of t_{min} probably requires the most a priori assumptions about the Universe. In particular, since

⁵This is part of the reason for choosing such optimistic error bars on the simulated data.

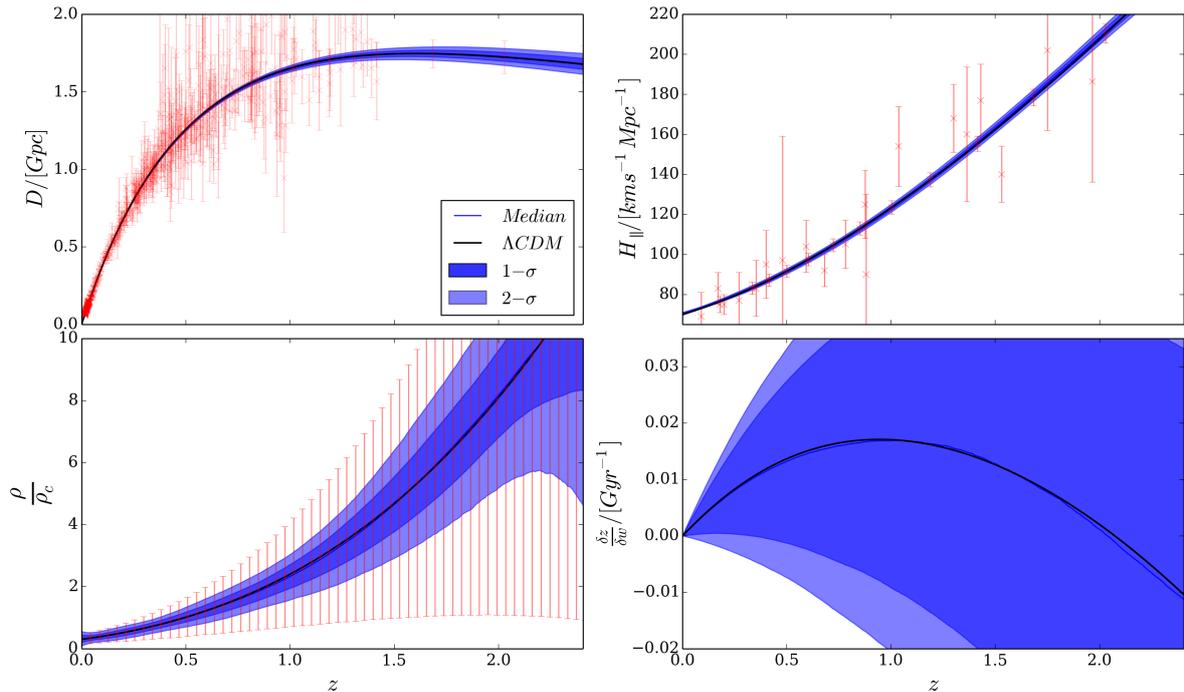


Figure 6.9: Observables on the PLC0 for forecast data. Top Left: Posterior distribution of $D(z)$. Top Right: Posterior distribution of $H(z)$. Bottom Left: The posterior distribution of $\rho(z)$. Bottom Right: The posterior distribution of $\frac{\delta z}{\delta w}$. Note that the simulated data points for $D(z)$ and $H_{\parallel}(z)$ have been appended to the data that are currently available.

an estimate of the time interval between the Big Bang and the beginning of structure formation is required, we need to make certain assumptions about the nature of the early Universe. This is not really in line with the bottom up approach we are pursuing. In the second simulation we therefore used redshift drift $\frac{\delta z}{\delta w}$ data instead of the t_{min} data point. As we will see such data have the potential to significantly constrain violations of the CP.

In both simulations we have used the forecast data given in [80] to get an estimate of the expected precision in $D(z)$ and $H_{\parallel}(z)$ data from upcoming surveys. We use approximately the same redshift values and relative uncertainty as shown in Figure 8 of [80]. We use twelve simulated data points for each $H_{\parallel}(z)$ and $D(z)$ which correspond, approximately, to a $\frac{\sigma_H}{H}$ of between 0.01 (at $z = 0.4$) and 0.02 (at $z = 2.4$) and a $\frac{\sigma_D}{D}$ of between 0.02 (at $z = 0.4$) and 0.06 (at $z = 2.4$). We treat these as independent data points which we centre on our reference Λ CDM model defined by the parameters listed in Table 2.1. We used these data in combination with the Union 2.1 [4] and cosmic chronometer [87] data (the combined data sets are plotted in Figure 6.9). For a value of t_{min} we have assumed a $1 - \sigma$ lower bound on the age of the universe at 10% and centred t_{min} on the t_0 value of our reference Λ CDM model. This is already quite

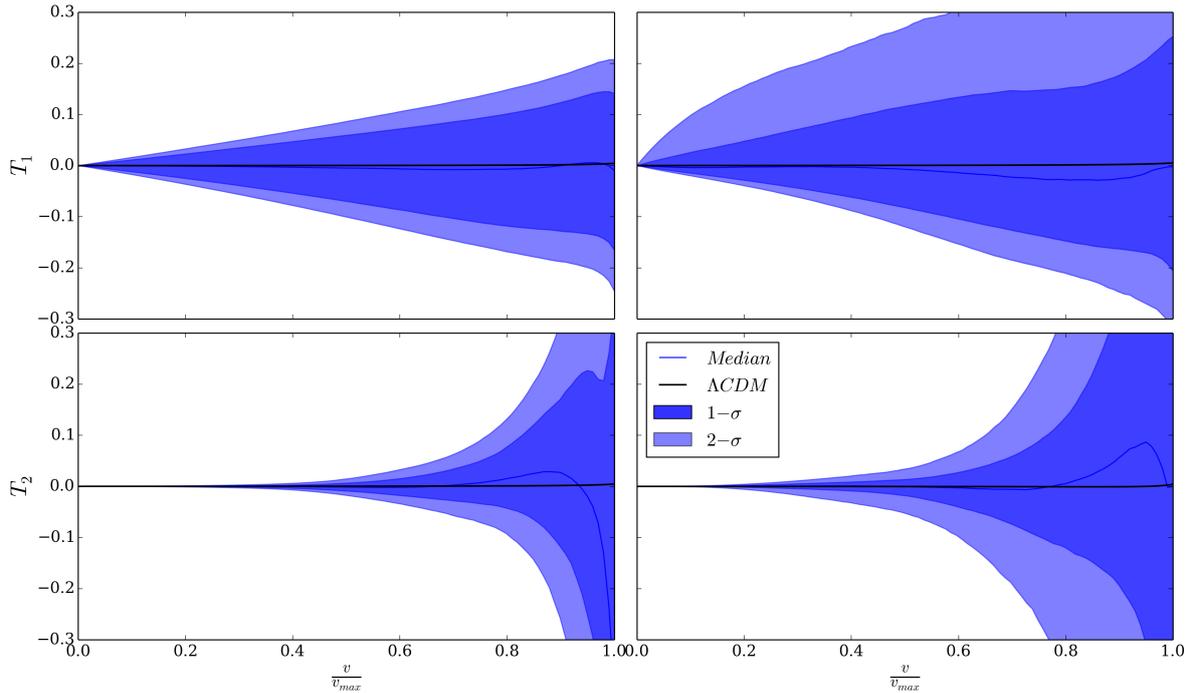


Figure 6.10: The quantities T_1 and T_2 for forecast data plotted as functions of the normalised affine parameter $\frac{v}{v_{max}}$. The figures on the left (right) correspond to the PLC0 (PLCF).

optimistic. However, since HI surveys such as the SKA promise to observe the universe right up to the dark ages, it is at least conceivable. We use the same likelihood function as in §6.2 and set the prior over $\rho(z)$ in exactly the same way. The results, summarised in Figures 6.9 - 6.11, are not very encouraging. Although the reconstructed functions on the PLC0 show significant improvement, neither the reconstructed distributions of T_1 and T_2 , nor the joint constraints on Ω_{m0} vs. $\Omega_{\Lambda0}$ are significantly improved. We still found that without an upper bound on t_0 the maximum value of Λ is, not surprisingly, effectively unconstrained. Thus we still had to enforce the priors $\Omega_{K0} > -1$ and $\Lambda < 0.23 \text{ Gpc}^{-2}$. Clearly, if we are going to test the CP using these data, we have to also find a way to place an upper bound on t_0 . However, our results on simulated test data in §6.1 seem to indicate that even a very precise observation of t_0 will still allow models with non-negligible curvature to fit the data with high confidence. The reason for this is that an observation of t_0 is only able to constrain the values of ρ and $H_{||}$, and hence Ω_{K0} , at the vertex of the cone. The best fit constant curvature models (i.e. FLRW) seem to prefer $\Omega_{K0} \approx 0$. This is not the case for Λ LTB models which have an extra functional degree of freedom which allows them to simultaneously fit all the data with Ω_{K0} significantly different from zero. To constrain this extra degree of freedom using estimates of the age of the Universe we would need to know the age of the Universe as a function of the redshift i.e. $t(z)$ data.

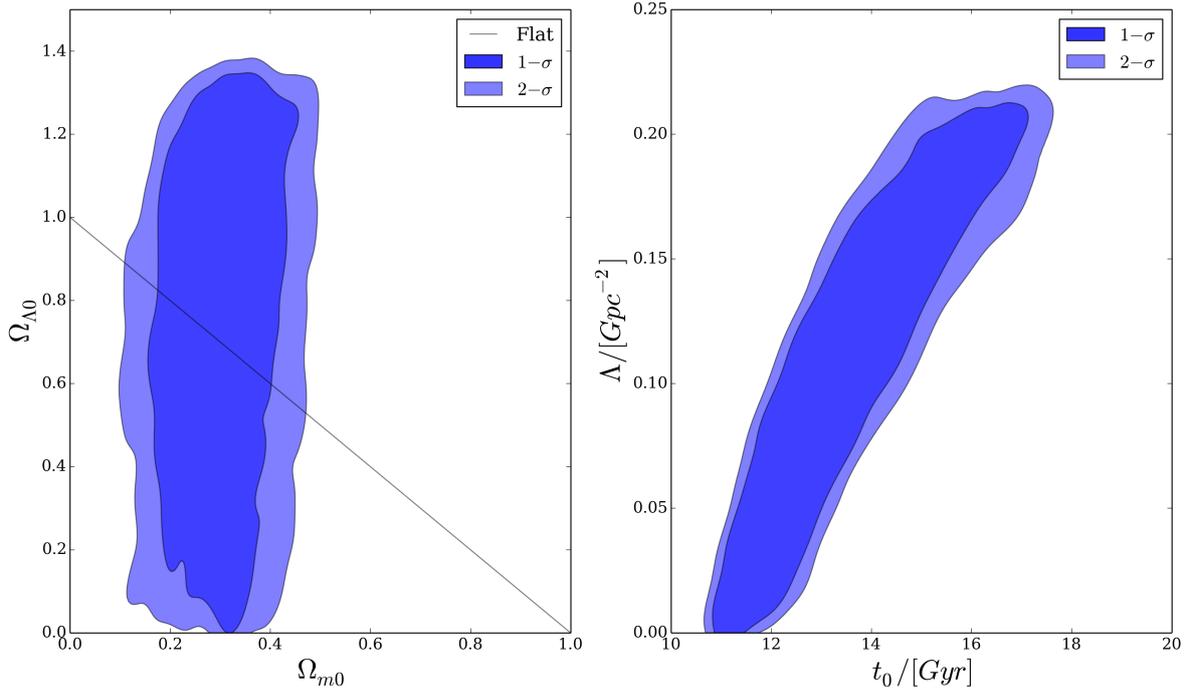


Figure 6.11: Contour plots of $\Omega_{\Lambda 0}$ vs. $\Omega_{m 0}$ (left) and Λ vs. t_0 (right) at the vertex of the PLC0 for forecast data.

This possibility will be investigated in future research. However, at the moment it is not clear whether $t(z)$ data could be obtained in a model independent way. Redshift drift data on the other hand are in principle completely model independent.

In the second simulation we used the 12 redshift drift data points shown in Figure 6.12. It is difficult to give a concrete answer for how accurately future surveys will be able to measure the redshift drift. The accuracy with which this quantity can be measured depends not only on the instrument but also on the objects being observed. Because of their rich spectra, quasars (QSO's) in the Ly α forest currently seem to be some of the best candidates for the job (see [157] for a discussion). These QSO's are incredibly bright and can therefore be observed out to high redshifts (typically between $2 < z < 5$). As a result their peculiar motions do not significantly contaminate the redshift determination. However, at $\frac{\delta z}{\delta w} \approx 10^{-8}$ per century, we still need extremely accurate determinations of the redshift to measure this quantity in practice. Our chances for detecting it obviously increase as more time elapses. Current estimates [157] with the expected sensitivity from the ELT's CODEX spectrograph [74] seem to suggest that it would take between 20-30 years to measure the redshift drift to within the 10% level, albeit in relatively wide redshift bins. Here we will stick to our highly idealised experiment and ignore the error in the redshift measurement (i.e. we assume the redshift bins are very narrow). We have therefore assumed that each of our 12 data points have a fixed uncertainty of 0.01 Gyr^{-1}

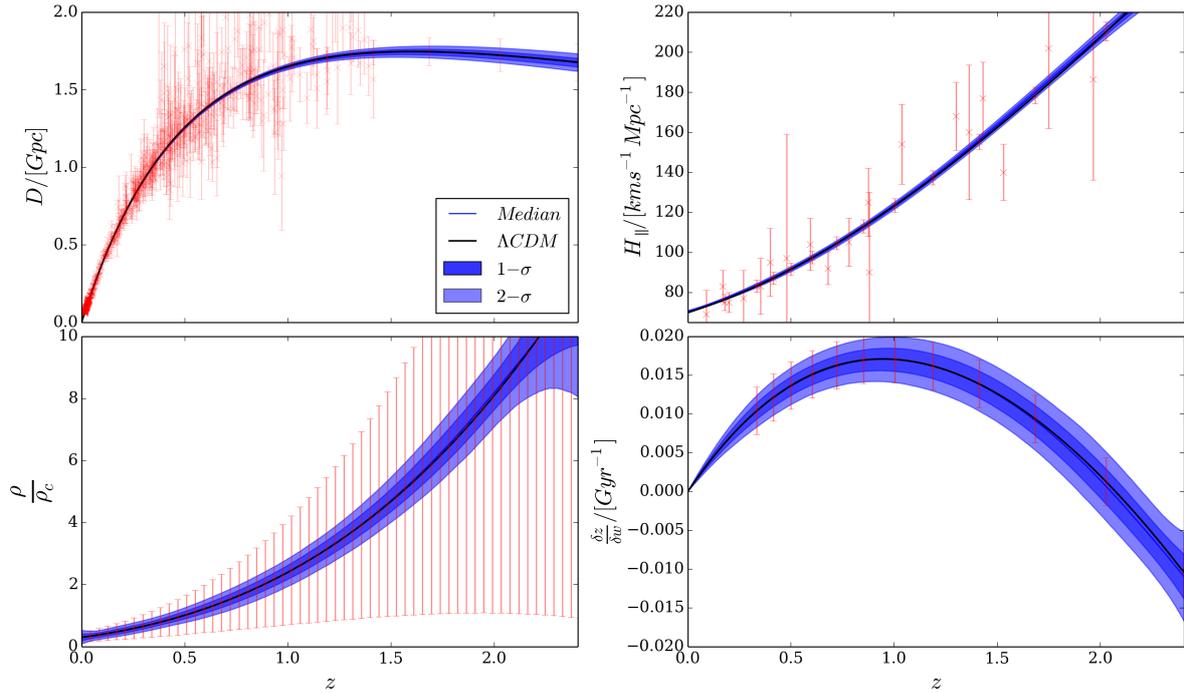


Figure 6.12: Observables on the PLC0 for forecast data including redshift drift. Top Left: Posterior distribution of $D(z)$. Top Right: Posterior distribution of $H(z)$. Bottom Left: The posterior distribution of $\rho(z)$. Bottom Right: The posterior distribution of $\frac{\delta z}{\delta w}$. Note that the simulated data points for $D(z)$ and $H_{\parallel}(z)$ have been appended to the data that are currently available. The simulated redshift drift data points are at the same redshift values as those for $D(z)$ and $H_{\parallel}(z)$.

which translates to approximately 20% relative error at $z = 1$. The redshift values are the same as those for the forecast $D(z)$ and $H_{\parallel}(z)$ data. Note that, although it is unlikely that we will obtain such accurate observations of the redshift drift at $z \leq 2$, data at higher redshifts will be more constraining. This we deduce from Figure 6.1 which shows that the uncertainty rapidly grows with redshift. More careful considerations are required though because incorporating these high redshift data will require that we specify mean functions for $H_{\parallel}(z)$ and $\rho(z)$ during GPR.

With these considerations in mind, we now summarise the results shown in Figures 6.12 - 6.14. It is interesting to note that the reconstructed distributions of $D(z)$ and $H_{\parallel}(z)$ are not significantly affected. The reconstructed distribution of $\rho(z)$ shows significant improvement, especially at high redshifts. This is dwarfed by the resulting improvement in the distributions of T_1 and T_2 shown in Figure 6.13. The quantity T_1 in particular shows significant improvements on both the PLC0 and the PLCF. Perhaps the most surprising are the remarkable constraints on the

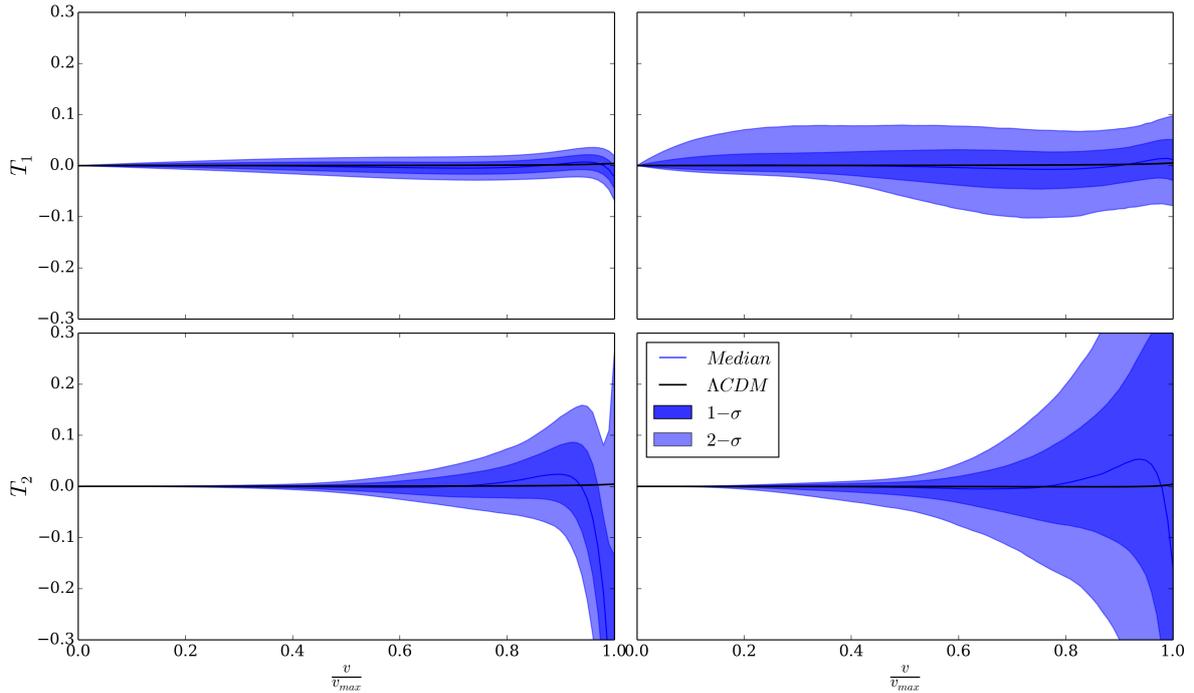


Figure 6.13: The quantities T_1 and T_2 for forecast data including redshift drift plotted as functions of the normalised affine parameter $\frac{v}{v_{max}}$. The figures on the left (right) correspond to the PLC0 (PLCF).

values of Λ and t_0 shown in Figure 6.14. The value of Λ is constrained to within approximately 15% at the $2 - \sigma$ level. We also see that redshift drift data places stronger constraints on t_0 from below than from above. A model independent upper bound on t_0 will therefore still provide valuable constraints. Finally, note that the constraints on the density parameter Ω_{m0} have remained about the same. Low redshift density data are therefore still required to further constrain models with non-negligible curvature. Note that, as with the simulated test data in §6.1, we did not have to enforce the priors $\Omega_{K0} > -1$ or $\Lambda < 0.23$.

6.4 Discussion

Certain valuable deductions can be made from the highly idealised experiment performed above. First among these is that accurate observations of $H_{||}(z)$ and $D(z)$ only place strong constraints on the curvature in models of the Universe in which the curvature is assumed to be constant. As we have shown for ALTB models in particular, the curvature is not strongly constrained by these data in models with additional functional degrees of freedom. Moreover, while t_0 data can be used to constrain the value of Λ , these constraints are degenerate with the values of ρ and $H_{||}$ at the vertex of the PLC0. As illustrated by our test simulation (see Figure 6.4), a

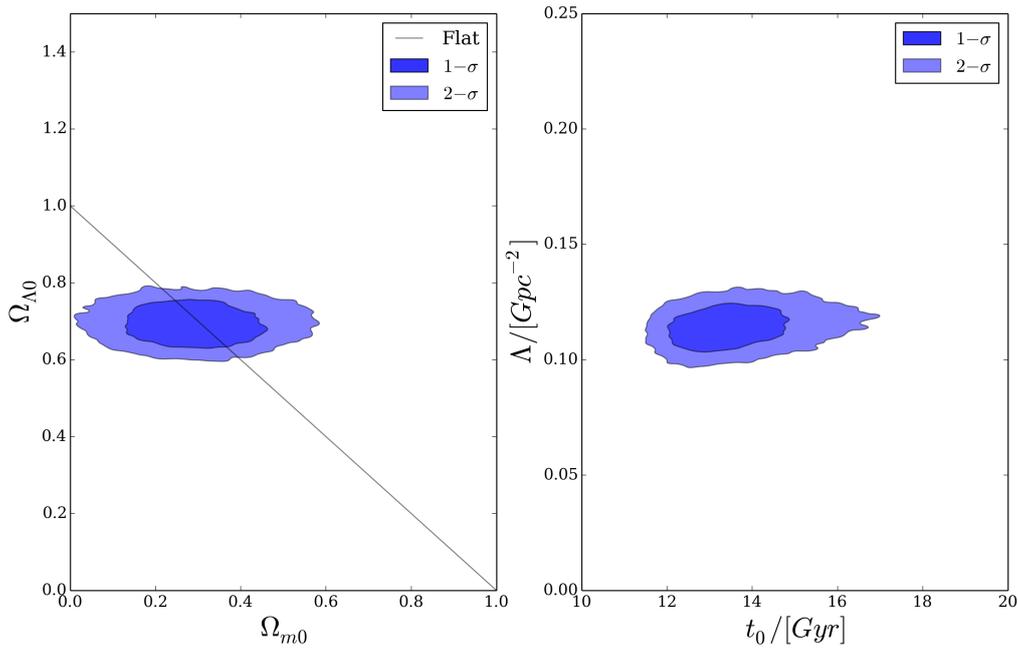


Figure 6.14: Contour plot of $\Omega_{\Lambda 0}$ vs. $\Omega_{m 0}$ (left) and Λ vs. t_0 (right) at the vertex of the PLC0 for forecast data including redshift drift.

fairly accurate estimate of the density at low redshifts would be required to further constrain the curvature in ALTB models. We highlighted some of the difficulties involved in obtaining such data directly from galaxy number counts in §3.5. The Compton scattering of CMB photons by hot intra-cluster gas (known as the Sunyaev-Zel’dovich effect [158]) provides another promising probe of radial inhomogeneity down the PLC0. The kinematic contribution to this effect, thought to be caused by the peculiar motion of galaxy clusters relative to the CMB and known as the kinematic Sunyaev-Zel’dovich (kSZ) effect [159], can be mimicked in a universe where the observer is situated near the centre of a Gpc scale under/over-density (see [160, 161] for example). This effectively happens because of the systematic red/blue-shifts of photons resulting from the change in gravitational potential as they traverse the inhomogeneity. Thus the observed magnitude of the kSZ effect places constraints on the relative over/under-density of the inhomogeneity and can therefore constrain the uncertainty in ρ at the vertex of the PLC0. This possibility will be investigated in future research. The kSZ effect does not, however, constrain the mean value of the density. Another difficulty with using the kSZ in this way is highlighted by the result proved on §A viz. in spherical symmetry the assumption that dust and radiation are comoving necessarily presupposes that the Universe is FLRW. Thus to incorporate kSZ data in a meaningful way requires a thorough understanding of the dynamics in

multi-fluid models. The dust + radiation case in spherical symmetry was investigated numerically in [162]. These authors illustrate that dust and radiation can't be assumed to be comoving without inadvertently fine tuning the radiation profile to be homogeneous. The implication is that it is not possible to separate out the kinematic and gravitational contributions to the kSZ effect in radially inhomogeneous models of the Universe. This makes it difficult to rule out LTB void models by combining only kSZ data with with the data that are already available. It is therefore very exciting to see attempts to analyse the CMB (see [163] for example) and the BAO (see [79] for example) in a model independent way. Such data would be invaluable to the observational cosmology programme. Indeed there are some who claim that CMB and kSZ data have already confirmed the CP (see [164, 165] for example). However, in light of the current uncertainty regarding the dynamics of radiation in inhomogeneous, multi-fluid models, we believe this claim to be premature.

The possibility of incorporating redshift drift data into the algorithm seems very promising. However, $D(z)$, $H_{\parallel}(z)$ and $\frac{\delta z}{\delta w}$ data by themselves do not seem to rule out models with non-zero curvature. It should be kept in mind that the contour plot of Ω_{m0} vs. $\Omega_{\Lambda0}$ shown in Figure 6.14 only shows the constraints on these parameters at the vertex of the PLC0. The variation of the curvature parameter with redshift is related to the quantity T_2 which is only strongly constrained at $\frac{v}{v_{max}} \leq 0.5$, or roughly $z \leq 0.4$ in our reference Λ CDM model defined by the parameters shown in Table 2.1. On the other hand, the quantity T_1 which measures the relative difference between the transverse and longitudinal expansion rates and so is related to the matter shear, is well constrained on the PLC0 over the whole domain of the problem (i.e. up to a maximum redshift of about $z_{max} = 2.4$). Besides our discussion in §3.3 and §3.4, we have given little to no indication of what the expected values of these quantities are in perturbed FLRW models of the Universe. This is a question that we cannot currently provide concrete answers to (however see [77] who suggests a limit of $|T_2| \leq 10^{-5}$). We can however suggest a possible way to do this.

There are a wealth of cosmological data (e.g. CMB [15], BAO [18], weak lensing [19], redshift-space distortions [20] or galaxy clusters [21]) which, because of their inherent dependence on FLRW perturbation theory, we could not incorporate into the algorithm. These data place strong constraints on some of the non-background parameters of the concordance model of cosmology, in particular the curvature perturbation $\mathcal{P}_{\zeta}(k_0)$ and the matter perturbations amplitude σ_8 (see for example §13 of [25]). In order to constrain the allowed variations in T_1 and T_2 , expressions for their correlation functions should be derived in terms of the standard perturbation parameters. The constraints on these parameters can then be used, by directly

sampling their posteriors for example, to get an estimate of the expected variation in T_1 and T_2 . Such an analysis is essential to test the CP within the framework presented in this thesis. We should also remark on the possibility of incorporating physically motivated priors into the algorithm. An example is placing a hard prior on the length scale parameter l during GPR. Since the algorithm is only meant to test for large scale inhomogeneities, variations in the functions $H_{\parallel}(z)$ and $\rho(z)$ below a certain scale can and should be excluded from the outset. This is part of the reason for not marginalising over the hyperparameters during GPR. Note that GPR is completely ignorant of the physical laws which govern the underlying process. There are therefore certain physical considerations which could be brought to bear to further constrain the class of allowable cosmological models. However, since a priori assumptions about the nature of the Universe, especially the early Universe, are exactly the kind of assumptions we are trying to avoid, this is an aspect that will require careful consideration during possible future applications of the algorithm.

Concluding remarks

The primary goal of this thesis was to develop a general and extendible framework which can, in the spirit of the observational cosmology programme [3], be used to infer the geometry of the universe from data. This would provide the most direct and model independent test of the cosmological principle. However, there are a number of difficulties which currently prevent us from implementing this approach in complete generality i.e without making any symmetry assumptions about the space-time. We have therefore restricted the analysis to the class of spherically symmetric models which may include a cosmological constant viz. the Λ LTB model. Note that, if we also relaxed the assumption of isotropy, there would be multiple ways, in addition to violations of the CP, in which T_1 and T_2 could deviate from zero. While it is possible that the CP could be violated and that there are contributions to Λ from incorrect averaging procedures, these effects should, in accordance with Occam's razor, be tested for separately. If neither effect can adequately account for observations separately then we may consider testing them in tandem. Thus, since Λ LTB is the simplest generalisation of Λ CDM upon abandoning the Copernican principle, the framework we have presented is the logical starting point from which to test the assumption of homogeneity on large scales.

Our results in §6.2 illustrate that current data are not able to confirm or refute the validity of the CP. This we expected from the outset and therefore tried to provide some additional insight into which data would be required to ultimately test the CP. In §6.3 we performed a series of idealised experiments that provide at least partial answers to this question. In §6.4 we established that it would be very difficult to verify the CP using only $D(z)$, $H_{\parallel}(z)$ and t_0 data. The main difficulty here stems from the fact that an observation of t_0 only constrains the combination of Λ , H_{\parallel} and ρ at the vertex of the PLC0. We hinted that an estimate of the age of the Universe as a function of redshift $t(z)$ would be required to constrain the extra functional degree of freedom present in Λ LTB models. However, as discussed in §3.6, methods which use

the ages of astrophysical objects to derive bounds on the age of the Universe presuppose that we know the time that elapsed between the Big Bang and the beginning of structure formation. We therefore investigated the possibility of using redshift drift data instead.

Redshift drift data are slightly at odds with the spirit of the original observational cosmology programme. The reason for this is that, to obtain redshift drift data, we need to know derivatives transverse to the PLC0. The only way to measure these derivatives in practice is to compare the redshift on two different PLC's. Thus redshift drift is not really an on-the-lightcone observable. However, at a separation of only 20 - 30 years, these PLC's are, for all effective purposes, an infinitesimal distance apart. Although redshift drift data show great promise, there are considerable difficulties involved in obtaining such data in practice. In particular, since these data will have to be reported in redshift bins, there will be non-negligible uncertainty in the value of the redshift (see the discussion in §3.2). Such horizontal error bars could be incorporated simply by introducing an additional sampling layer in which we sample $z = z(z_{obs})$ before sampling $H_{||}(z)$ and $\rho(z)$ (an idea that we already exploit to deal with uncertainties in the $v(z)$ relation). This could also be used to investigate the effects of redshift remapping in ALTB models.

Because of the highly idealised nature of our investigation, the results shown in §6.3 should be considered as a best case scenario. The constraints would be significantly degraded if we accounted for all the possible sources of uncertainty (e.g. systematic in the data or uncertainties in the reported redshifts). Thus it is likely that we will have to supplement the D , $H_{||}$ and $\frac{\delta z}{\delta w}$ data with additional data to ultimately test the CP. As discussed in §6.4, low redshift density data and an upper bound on the age of the Universe could be very valuable to the observational cosmology programme. Establishing the required precision in these data is however non-trivial. It should be kept in mind that the formalism we have presented still relies on a number of assumptions. The most fundamental underlying assumption is the theory of gravity employed viz. GR. If GR is violated on cosmological scales, the tests we have presented will have little meaning for the CP as it could result in T_1 and T_2 being different from zero even when the CP holds¹ Moreover, cosmological tests of gravity theories are necessarily degenerate with both the symmetry assumptions of the cosmological model and the assumptions underlying astrophysical data generating processes. We are of the opinion that GR is the most robust and well tested assumption in this formalism, albeit on much smaller scales than we are interested in. Finding statistically significant deviations of T_1 and T_2 from zero would most probably indicate a breakdown of either the symmetry or astrophysical assumptions. Even though we have advocated

¹We should note that the quantity T_2 is in principle independent of GR. However, since we use the field equations to reconstruct $D(z)$ from samples of $H_{||}(z)$ and $\rho(z)$, the T_2 test we present does depend on the validity of GR.

this formalism as a means to test the former, it inevitably also tests the underlying astrophysical assumptions. This degeneracy can, to some extent, be broken by carefully constructing consistency relations which test different aspects of the model under scrutiny. The quantities T_1 and T_2 , for example, test different aspects of the Λ CDM model. Whereas $T_1 \neq 0$ when deformations of the fluid are allowed (i.e. $\sigma_{ab} \neq 0$), $T_2 \neq 0$ when the spatial curvature is not constant. Although these effects are not completely unrelated, our astrophysical assumptions will affect them differently. An example of this is provided by the fact that redshift drift data strongly constrains T_1 but are not able to place tight constraints on T_2 at high redshifts. In such a scenario, because of the degeneracy between symmetry and astrophysical assumptions, we would need high confidence in the redshift drift data to conclude that the Universe is indeed homogeneous on large scales. Establishing the same result using a different, independent data set, $t(z)$ say, would make the conclusion more robust.² In light of the current uncertainties regarding the nature of the cosmological constant, we should strive to scrutinize the Λ CDM model by testing as many consistency relations, with as many independent data sets, as possible. However, since Λ CDM is the simplest model compatible with current data, it remains the preferred model with which to describe the background dynamics of the Universe.

Shifting the focus away from testing the CP, the above ideas could also be exploited to test the compatibility between certain astrophysical assumptions and the concordance model. To see this, suppose we ran the algorithm with all the cosmological data we currently have at our disposal, including those that rely on FLRW perturbation theory. In this case finding a value of either T_1 or T_2 different from zero would indicate an incompatibility between some of the data. Such an incompatibility would not manifest in an analysis which presupposes a Λ CDM model. Finding that the distributions of T_1 and T_2 are compatible with zero, however, suggests that the data, and therefore the astrophysical assumptions, are compatible with each other and the concordance model. The time scales involved in obtaining model independent data such as, for example, redshift drift, might necessitate relying on these sorts of tests, sometimes referred to as null tests, for some time to come. Note that the distributions of T_1 and T_2 are not our only guiding factors. As we hinted at in §3.5, it should be possible to work backwards from the

²Of course it is incredibly difficult, if not impossible, to prove the CP by showing that the quantities T_1 and T_2 (or any other consistency relations for that matter) are close enough to zero. The difficulty lies in determining how close to zero they should be for the CP to be valid. The best we can do is to derive expressions for T_1 and T_2 in terms of the perturbed FLRW metric up to a specific order and then use data to estimate the maximum allowed deviation of T_1 and T_2 at that order. If the contours of T_1 and T_2 that we reconstruct using the formalism developed in this thesis are consistent with those allowed in a perturbed FLRW model, then we still have only shown that the CP is not violated within the regime in which the perturbed form of the FLRW metric, at the specified order, provides an adequate description of the real Universe.

reconstructed distribution of $\rho(z)$ to find the number density of galaxies $n(z)$. We may then test our assumptions (e.g. the mass to luminosity ratio of galaxies or the shape of the selection function) by comparing $n(z)$ to the observed number density $n_{obs}(z)$. Such an analysis can provide valuable insights into the current model dependence of certain cosmological data sets.

Appendix A

Spherically symmetric models

In this appendix we present some details regarding spherically symmetric models of the Universe. We will start by establishing that when CDM and radiation are comoving in a spherically symmetric universe then it must have maximally symmetric spatial sections (this result is also established numerically in [162]). This highlights a potential difficulty with incorporating radiation into the observational formalism. We will then outline the procedure used to rewrite the the EFE's in the form shown in §2.3.2. Finally we will describe the transformations required to rewrite the integral (2.65) in terms of the Carlson elliptic integral (2.66). Finally we will the parametric solution used in 5.1.1 to fit the the LTB model to data.

A.1 Comoving dust and radiation

The energy content of the universe is assumed to consist of two perfect fluids, one with pressure (radiation) and one without (dust). Assuming that dust moves with 4-velocity u^a , we may describe these fluids with energy momentum tensors (EMT's) of the form

$$T_D^{ab} = \rho_D u^a u^b, \quad (\text{A.1})$$

$$T_R^{ab} = \rho_R \hat{u}^a \hat{u}^b + P h^{ab}, \quad (\text{A.2})$$

where the subscripts D and R label dust and radiation components respectively and \hat{u}^a is the 4-velocity of radiation with respect to dust. Denoting the boost factor by $\gamma = (1 - v_a v^a)^{-\frac{1}{2}}$ we can relate \hat{u}^a to u^a via

$$\hat{u}^a = \gamma(u^a + v^a); \quad \hat{u}_a \hat{u}^a = -1; \quad v^a v_a > 0; \quad v_a u^a = 0, \quad (\text{A.3})$$

where v^a the peculiar bulk velocity of the radiation fluid. To respect local energy/momentum conservation the total EMT $T^{ab} = T_R^{ab} + T_D^{ab}$ must satisfy

$$\nabla_a T^a_b = 0. \quad (\text{A.4})$$

However since the fluids don't interact in any way other than gravitationally, the EMT's of both fluids must also satisfy the conservation equations separately. Introducing standard spherical coordinates $x^a = [t, r, \theta, \phi]$ the metric for a spatially spherically symmetric space-time can be written as

$$ds^2 = -N(t, r)^2 dt^2 + X(t, r)^2 dr^2 + R(t, r)^2 d\Omega^2. \quad (\text{A.5})$$

For simplicity we analyse the problem in the frame of dust. We have verified that, with a little more effort, the result can also be established in the frame of radiation. Note that, since the EMT inherits the symmetries of the spacetime, we know that the peculiar velocity will only have components in the radial direction i.e $v^a = [0, v, 0, 0]$. This can be confirmed by substituting the general form of v^a into the total MCE (2.27) as defined in §2.2. Substituting (A.5) into the MCE for dust we find that

$$\rho_D \frac{N'}{N} = 0, \quad \Rightarrow \quad N = N(t), \quad (\text{A.6})$$

since we assume $\rho_D \neq 0$. Choosing t to measure proper time we may set $N(t) = 1$ in the frame of dust. The $G^0_1 = \kappa T^0_1$ component of the field equations (note including Λ) with the metric (A.5) gives

$$2 \frac{\dot{R}'}{N^2 R} - 2 \frac{N' \dot{R}}{N^3 R} - 2 \frac{\dot{X} R'}{N^2 R X} = \frac{4}{3} \gamma^2 \kappa N \rho_R v, \quad (\text{A.7})$$

$$2 \frac{\dot{R}'}{R} - 2 \frac{\dot{X} R'}{R X} = \frac{4}{3} \gamma^2 \kappa \rho_R v, \quad (\text{A.8})$$

where we have used the fact that $N = 1$ to obtain the last equation. Supposing that the fluids are comoving (i.e. $v = 0$), we obtain the usual LTB first integral i.e.

$$\frac{\dot{X}}{X} = \frac{\dot{R}'}{R'}, \quad \Rightarrow \quad X(t, r) = g(r) R'(t, r), \quad (\text{A.9})$$

with $g(r)$ a function of integration (note $g(r) \neq 0$ otherwise the coordinate system is not well defined). Under the assumption that the fluids are comoving, the MCE for radiation reads

$$\frac{1}{3} \rho'_R + \frac{4}{3} \rho_R N^3 N' = 0. \quad (\text{A.10})$$

However since $N' = 0$ it must be that $\rho'_R = 0$. This is rather peculiar, setting $v = 0$ forces the radiation profile to be homogeneous. Computing the ECE of radiation we find

$$\frac{1}{3} \dot{\rho}_R - \frac{16}{3} \rho_R N^2 \dot{N} - \frac{4}{3} N^3 \left(\dot{\rho}_R + \rho_R \left(\frac{\dot{X}}{X} + 2 \frac{\dot{R}}{R} \right) \right) = 0. \quad (\text{A.11})$$

Taking the derivative of (A.11) with respect to r , we find, using $N' = 0$ and commuting the partial derivatives, that

$$\left(\frac{\dot{X}}{X} + 2 \frac{\dot{R}}{R} \right)' = 0 \quad \Rightarrow \quad \frac{\dot{X}}{X} + 2 \frac{\dot{R}}{R} = C(t) \quad (\text{A.12})$$

for some function $C(t)$. This equation integrates to

$$X(t, r)R^2(t, r) = F(r) \exp\left(\int C(t)dt\right) \quad (\text{A.13})$$

for some function $F(r)$. However we have from the field equations that $X(t, r) = g(r)R'(t, r)$ so that equation (A.13) can be written as

$$R'(t, r)R^2(t, r) = \tilde{F}(r) \exp\left(\int_{t_0}^t C(t)dt\right) \Rightarrow \frac{1}{3}R^3(t, r) = \left[\exp\left(\int C(t)dt\right)\right] \left[\int \tilde{F}(r)dr\right] \quad (\text{A.14})$$

where $\tilde{F}(r) = F(r)/g(r)$. Thus $R(t, r)$ is a separable function of t and r . If this is the case then eq. (A.9) also implies that $X(t, r)$ is a separable function of t and r . When the metric can be written as

$$ds^2 = -dt^2 + X(t, r)^2 dr^2 + R(t, r)^2 d\Omega^2, \quad (\text{A.15})$$

it is easy to show that separability of the metric functions implies FLRW. To see this, assume that the scale factors are separable i.e. $X(t, r) \equiv \xi(t)x(r)$ and $R(t, r) \equiv a(t)\alpha(r)$. Then by eq. (A.9) we have that

$$X(t, r) = g(r)R'(t, r) \Rightarrow X(t, r) = g(r)\alpha'(r)a(t). \quad (\text{A.16})$$

Substituting (A.16) into (A.15) we have

$$ds^2 = -dt^2 + a(t)^2 (g(r)^2 \alpha'(r)^2 dr^2 + \alpha(r)^2 d\Omega^2). \quad (\text{A.17})$$

Next introduce a new radial coordinate such that (note we have not yet used the gauge freedom in r)

$$d\tilde{r} = \alpha'(r)dr, \Rightarrow \tilde{r} = \alpha(r) \quad \text{when} \quad \alpha(0) = 0. \quad (\text{A.18})$$

Dropping tildes we find that

$$ds^2 = -dt^2 + a(t)^2 (g(r)dr^2 + r^2 d\Omega^2), \quad (\text{A.19})$$

which has maximal spatial symmetries and is completely equivalent to the FLRW form of the metric. We have established that a spherically symmetric universe containing only dust and radiation is necessarily FLRW when the fluids are comoving. This is an interesting result which crucially depends on the assumption that the fluids do not interact. Although it might at first appear surprising, it is actually quite intuitive. Since we have forced the radiation to be comoving with dust, we have forced the radiation to undergo geodesic motion. This is only possible if pressure gradients vanish¹. Since we only use the conservation equations

¹Note this idea generalises to more asymmetrical spacetimes.

and one of the off diagonal EFE's (i.e. $G_1^0 = \kappa T_1^0$) the result holds regardless of whether we include a cosmological constant or not. The fact that one additional derivative of the ECE (A.11) combined with the MCE (A.10) establishes the result, seems to suggest that it is the local conservation of energy and momentum which enforces the symmetries. This will be a dynamical effect which takes some time. It would be interesting to investigate if, or on which scale, this effect is significant at the time of decoupling.

The above description in terms of two non-interacting fluids is probably not a very realistic. However, it does suggest caution on our part. In the real Universe we have to account for a number of other particle species, the most abundant of which are baryons. Now baryons and radiation do in fact interact and will therefore not satisfy the conservation equations separately. However, the above argument actually generalises to any two non-interacting fluids, as long as one of them has vanishing pressure. Since we routinely describe the dynamics of the background Universe under this approximation, we have to be aware of the fact that our idealisation might have subtle consequences.

A.2 Single spherically symmetric fluid

Here we outline the procedure used to rewrite the field equations in terms of the function $M(t, r)$ (2.56) in §2.3.2. Substituting the metric (A.5) into the field equations gives the following system of equations

$$2\frac{R''}{RX^2} - 2\frac{\dot{R}\dot{X}}{N^2RX} - 2\frac{R'X'}{RX^3} - \frac{1}{R^2} - \left(\frac{\dot{R}}{NR}\right)^2 + \left(\frac{R'}{RX}\right)^2 + \Lambda = -\kappa\rho \quad (\text{A.20})$$

$$2\frac{\dot{R}'}{N^2R} - 2\frac{N'\dot{R}}{N^3R} - 2\frac{\dot{X}R'}{N^2XR} = 0 \quad (\text{A.21})$$

$$-2\frac{\ddot{R}}{N^2R} + 2\frac{\dot{R}\dot{N}}{RN^3} + 2\frac{R'N'}{X^2RN} - \frac{1}{R^2} - \left(\frac{\dot{R}}{NR}\right)^2 + \left(\frac{R'}{RX}\right)^2 + \Lambda = w\kappa\rho \quad (\text{A.22})$$

$$\begin{aligned} \frac{\dot{R}\dot{N}}{RN^3} - \frac{\ddot{R}}{RN^2} + \frac{R'N'}{X^2RN} - \frac{R'X'}{RX^3} + \frac{R''}{RX^2} - \frac{\dot{R}\dot{X}}{N^2RX} - \frac{N'X'}{NX^3} + \frac{N''}{NX^2} \\ - \frac{\ddot{X}}{N^2X} + \frac{\dot{N}\dot{X}}{N^3X} + \Lambda = w\kappa\rho \end{aligned} \quad (\text{A.23})$$

We now manipulate these equations as follows:

1. Multiply (A.20) by R^2R' , (A.21) by $R^2\dot{R}$ and replace the $2\frac{RR'\dot{R}\dot{X}}{N^2X}$ term on the LHS of (A.20) using the resulting expression to find

$$\kappa R^2R'\rho = M' \quad (\text{A.24})$$

with $M = M(t, r)$ defined by

$$M(t, r) = -\frac{RR'^2}{X^2} + R + \frac{R\dot{R}^2}{N^2} - \frac{1}{3}\Lambda R^3. \quad (\text{A.25})$$

2. Next multiply (A.22) by $R^2\dot{R}$, (A.21) by $\frac{N^2R^2R'}{X^2}$ and replace the $2\frac{RR'\dot{R}N'}{X^2N}$ on the RHS of (A.22) using the resulting expression. Comparing to $\dot{M}(t, r)$ shows that (A.22) can be written as

$$-\kappa R^2\dot{R}w\rho = \dot{M}. \quad (\text{A.26})$$

These are the same as (2.54) and (2.55). Note that this procedure can be generalised in the case of two fluids (not necessarily comoving) to write the field equations in terms of \dot{M} and M' with M defined exactly as in (A.25). However, even in the case of a single fluid analytic solutions are not easy to find. The difficulty stems from the fact that the function M still depends on both t and r . The form of these equations might still be suitable to solve with a relatively simple numerical scheme, something we plan to investigate in future research.

Specialising to the case of dust we get the ALTB solution discussed in §2.3.2. In this case, starting from

$$t(\tilde{R}) - t_B(r^*) = \frac{1}{H_{\perp 0}} \int_0^{\tilde{R}} \frac{dR}{R\sqrt{\Omega_m \left(\frac{R_0}{R}\right)^3 + \Omega_K \left(\frac{R_0}{R}\right)^2 + \Omega_\Lambda}}, \quad (\text{A.27})$$

we can expand R into the denominator and pull out a factor of R_0 giving

$$t(\tilde{R}) - t_B(r^*) = \frac{1}{H_{\perp 0}} \int_0^{\tilde{R}} \frac{dR}{R_0\sqrt{\Omega_m \left(\frac{R_0}{R}\right) + \Omega_K + \Omega_\Lambda \left(\frac{R}{R_0}\right)^2}}. \quad (\text{A.28})$$

Changing variables $a = \frac{R}{R_0}$, $\Rightarrow da = \frac{dR}{R_0}$ this can be written as

$$t(\tilde{a}) - t_B(r^*) = \frac{1}{H_{\perp 0}} \int_0^{\tilde{a}=\frac{\tilde{R}}{R_0}} \frac{\sqrt{a} da}{\sqrt{\Omega_m + \Omega_K a + \Omega_\Lambda a^3}}, \quad (\text{A.29})$$

where we have expanded a factor of \sqrt{a} into the denominator. The integrand is an elliptic function that can be recast into one of Carlson's symmetric forms viz.

$$R_J(x, y, z, p) = \frac{3}{2} \int_0^\infty \frac{dt}{(t+p)\sqrt{t+x}\sqrt{t+y}\sqrt{t+z}}. \quad (\text{A.30})$$

Here x, y and z are the (possibly complex) roots of a cubic polynomial in t . They must have positive real parts, only one of which can be zero at any given time. The transformations that recast (A.29) into this form are, firstly

$$b = \frac{1}{a} \quad \Rightarrow \quad da = -\frac{db}{b^2}$$

which gives

$$t(\tilde{a}) - t_B(r^*) = \frac{1}{H_{\perp 0}} \int_{\frac{1}{\tilde{a}}}^{\infty} \frac{db}{b\sqrt{\Omega_m b^3 + \Omega_K b^2 + \Omega_\Lambda}}. \quad (\text{A.31})$$

Next shifting $c = b - \frac{1}{\tilde{a}} = b - \frac{R_0}{R}$, and pulling out a factor of $\sqrt{\Omega_m}$ (thus we do not consider vacuum solutions), results in

$$t\left(\frac{R_0}{R}\right) - t_B(r^*) = \frac{1}{\sqrt{\Omega_m} H_{\perp 0}} \int_0^{\infty} \frac{dc}{\left(c + \frac{R_0}{R}\right) \sqrt{\left(c + \frac{R_0}{R}\right)^3 + \frac{\Omega_K}{\Omega_m} \left(c + \frac{R_0}{R}\right)^2 + \frac{\Omega_\Lambda}{\Omega_m}}}. \quad (\text{A.32})$$

At any fixed value of r the quantity inside the square root sign is a cubic polynomial which can be written in terms of the elliptic function R_J as

$$t\left(\frac{R_0}{R}\right) - t_B(r^*) = \frac{1}{H_{\perp 0}(r^*)} \frac{2R_J(-x_1, -x_2, -x_3, \frac{R_0(r^*)}{R(r^*)})}{3\sqrt{\Omega_m(r^*)}}, \quad (\text{A.33})$$

where the x_i are the roots of²

$$\left(x + \frac{R_0(r^*)}{R(r^*)}\right)^3 + \frac{\Omega_K}{\Omega_m} \left(x + \frac{R_0(r^*)}{R(r^*)}\right)^2 + \frac{\Omega_\Lambda}{\Omega_m} = 0. \quad (\text{A.34})$$

After fixing the gauge for R_0 , equation (A.33) allows us to get $t(R)$ at any value of r . This relation can be numerically inverted to get $R(t, r)$ to arbitrary precision. Once that is given we can also find $H_{\perp}(t, r)$ and hence also $\dot{R}(t, r)$. For the full solution we therefore also need to find $R'(t, r)$, in which case the field equations would allow us to compute all required higher order derivatives of R (such as we might need in computing the geodesic equations for example). The procedure to find $R'(t, r)$ is described in [61]. We will not need the details since we only use this form of the ALTB procedure to find t_0 . We do however employ the parametric LTB solution in Example 5.1.1 so we will briefly outline how to find it.

To find the LTB solution we go back to the Friedmann-like equation of the form (2.60) i.e.

$$H_{\perp}^2 = \left(\frac{\dot{R}}{R}\right)^2 = \frac{M}{R^3} + \frac{E}{R^2}, \quad (\text{A.35})$$

where $M = M(r)$ follows the fact that $w = 0$ in (A.26). The function $E(r)$ (where $g(r) = \frac{1}{\sqrt{1+E(r)}}$ with $g(r)$ as in (A.9)) is related to curvature and determines the kind of evolution of $R(t, r)$. We distinguish between the three types of evolution:

- if $E(r) < 0 \Rightarrow$ elliptic evolution
- if $E(r) = 0 \Rightarrow$ parabolic evolution
- if $E(r) > 0 \Rightarrow$ hyperbolic evolution

²We will refrain from writing down the lengthy expressions for the roots, they were found using Sympy, Python's symbolic math package

Here we will demonstrate the solution for the hyperbolic case since that is the case relevant to Example 5.1.1. First we write (A.35) in terms of the integral

$$t(R) - t_B(r) = \int_0^R \frac{d\tilde{R}}{\sqrt{M/\tilde{R} + E}}, \quad (\text{A.36})$$

where it should be understood that we integrate along worldlines of constant r and the function of integration $t_B(r)$ can again be identified as the bang time function. Introducing an additional parameter, $\eta = \eta(t, r)$ say, we try the trial solution $R(\eta) = C(r)(\cosh(\eta) - 1)$ for arbitrary positive definite function $C(r)$ and find the solution

$$H_{\perp 0}(t - t_B) = \frac{\Omega_{m0}}{2\Omega_{K0}^{\frac{3}{2}}}(\sinh(\eta) - \eta), \quad R = \frac{\Omega_{m0}}{2\Omega_{K0}}(\cosh(\eta) - 1), \quad (\text{A.37})$$

where Ω_{m0} , Ω_{K0} and $H_{\perp 0}$ depend only on r and are defined as in §2.3.2. This allows us to specify initial data for the model as described in Example 5.1.1.

Bibliography

- [1] H. L. Bester, J. Larena, P. J. van der Walt, and N. T. Bishop. What's inside the cone? Numerically reconstructing the metric from observations. *JCAP*, 2:009, February 2014. doi: 10.1088/1475-7516/2014/02/009.
- [2] H. L. Bester, J. Larena, and N. T. Bishop. Towards the geometry of the Universe from data. *MNRAS*, 453:2364–2377, November 2015. doi: 10.1093/mnras/stv1672.
- [3] G. F. R. Ellis, S. D. Nel, R. Maartens, W. R. Stoeger, and A. P. Whitman. Ideal observational cosmology. *Physics Reports*, 124:315–417, 1985. doi: 10.1016/0370-1573(85)90030-4.
- [4] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, K. S. Dawson, A. Dey, M. Doi, M. Donahue, S. Deustua, P. Eisenhardt, E. Ellingson, L. Faccioli, V. Fadeyev, H. K. Fakhouri, A. S. Fruchter, D. G. Gilbank, M. D. Gladders, G. Goldhaber, A. H. Gonzalez, A. Goobar, A. Gude, T. Hattori, H. Hoekstra, E. Hsiao, X. Huang, Y. Ihara, M. J. Jee, D. Johnston, N. Kashikawa, B. Koester, K. Konishi, M. Kowalski, E. V. Linder, L. Lubin, J. Melbourne, J. Meyers, T. Morokuma, F. Munshi, C. Mullis, T. Oda, N. Panagia, S. Perlmutter, M. Postman, T. Pritchard, J. Rhodes, P. Ripoche, P. Rosati, D. J. Schlegel, A. Spadafora, S. A. Stanford, V. Stanishev, D. Stern, M. Strovink, N. Takanashi, K. Tokita, M. Wagner, L. Wang, N. Yasuda, H. K. C. Yee, and T. Supernova Cosmology Project. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above $z > 1$ and Building an Early-type-hosted Supernova Sample. *ApJ*, 746:85, February 2012. doi: 10.1088/0004-637X/746/1/85.
- [5] P. Peter and J.P. Uzan. *Primordial cosmology*. Oxford graduate texts. Oxford University Press, 2009. ISBN 9780199209910. URL <https://books.google.co.za/books?id=FZXvAAAAAMAAJ>.
- [6] G. Lemaître. Expansion of the universe, A homogeneous universe of constant mass and

- increasing radius accounting for the radial velocity of extra-galactic nebulae. *MNRAS*, 91:483–490, March 1931. doi: 10.1093/mnras/91.5.483.
- [7] G. Lemaître. *The Primeval Atom: An Essay on Cosmogony*. Van Nostrand, 1950. URL <https://books.google.co.za/books?id=eSoFAAAAMAAJ>.
- [8] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929. doi: 10.1073/pnas.15.3.168.
- [9] E. Hubble and M. L. Humason. The Velocity-Distance Relation among Extra-Galactic Nebulae. *ApJ*, 74:43, July 1931. doi: 10.1086/143323.
- [10] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038, September 1998. doi: 10.1086/300499.
- [11] B. P. Schmidt et al. The High Z supernova search: Measuring cosmic deceleration and global curvature of the universe using type Ia supernovae. *Astrophys. J.*, 507:46–63, 1998. doi: 10.1086/306308.
- [12] S. Perlmutter et al. Measurements of Omega and Lambda from 42 high redshift supernovae. *Astrophys. J.*, 517:565–586, 1999. doi: 10.1086/307221.
- [13] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, July 1965. doi: 10.1086/148307.
- [14] J. L. Weiland, N. Odegard, R. S. Hill, E. Wollack, G. Hinshaw, M. R. Greason, N. Jarosik, L. Page, C. L. Bennett, J. Dunkley, B. Gold, M. Halpern, A. Kogut, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, K. M. Smith, D. N. Spergel, G. S. Tucker, and E. L. Wright. Seven-year wilkinson microwave anisotropy probe (wmap) observations: Planets and celestial calibration sources. *The Astrophysical Journal Supplement Series*, 192(2):19, 2011. URL <http://stacks.iop.org/0067-0049/192/i=2/a=19>.
- [15] P. A. R. Ade et al. Planck 2013 results. I. Overview of products and scientific results. *Astron. Astrophys.*, 571:A1, 2014. doi: 10.1051/0004-6361/201321529.
- [16] S. Alam, F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, E. Armengaud, É. Aubourg, S. Bailey, and et al. The Eleventh and Twelfth

- Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. *ApJS*, 219:12, July 2015. doi: 10.1088/0067-0049/219/1/12.
- [17] B. Bassett and R. Hlozek. *Baryon acoustic oscillations*, page 246. Cambridge Univ. Press, 2010.
- [18] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, A. J. Cuesta, L. A. N. da Costa, K. S. Dawson, R. de Putter, D. J. Eisenstein, J. E. Gunn, H. Guo, J.-C. Hamilton, P. Harding, S. Ho, K. Honscheid, E. Kazin, D. Kirkby, J.-P. Kneib, A. Labatie, C. Loomis, R. H. Lupton, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, C. K. McBride, K. T. Mehta, O. Mena, F. Montesano, D. Muna, R. C. Nichol, S. E. Nuza, M. D. Olmstead, D. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. Parejko, I. Pâris, W. J. Percival, P. Petitjean, F. Prada, B. Reid, N. A. Roe, A. J. Ross, N. P. Ross, L. Samushia, A. G. Sánchez, D. J. Schlegel, D. P. Schneider, C. G. Scóccola, H.-J. Seo, E. S. Sheldon, A. Simmons, R. A. Skibba, M. A. Strauss, M. E. C. Swanson, D. Thomas, J. L. Tinker, R. Tojeiro, M. V. Magaña, L. Verde, C. Wagner, D. A. Wake, B. A. Weaver, D. H. Weinberg, M. White, X. Xu, C. Yèche, I. Zehavi, and G.-B. Zhao. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample. *MNRAS*, 427:3435–3467, December 2012. doi: 10.1111/j.1365-2966.2012.22066.x.
- [19] M. Bartelmann and P. Schneider. Weak gravitational lensing. *Phys.Rept.*, 340:291–472, 2001. doi: 10.1016/S0370-1573(00)00082-X.
- [20] W. J. Percival, L. Samushia, A. J. Ross, C. Shapiro, and A. Raccanelli. Redshift-space distortions. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 369(1957):5058–5067, 2011. ISSN 1364-503X. doi: 10.1098/rsta.2011.0370.
- [21] S. W. Allen, A. E. Evrard, and A. B. Mantz. Cosmological Parameters from Observations of Galaxy Clusters. *ARA&A*, 49:409–470, September 2011. doi: 10.1146/annurev-astro-081710-102514.
- [22] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis. Modified Gravity and Cosmology. *Phys. Rept.*, 513:1–189, 2012. doi: 10.1016/j.physrep.2012.01.001.
- [23] R.M. Wald. *General Relativity*. Physics, astrophysics. University of Chicago Press, 1984. ISBN 9780226870328. URL <https://books.google.co.za/books?id=ibSdQgAACAAJ>.

- [24] E. Poisson. *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*. Cambridge University Press, 2004. ISBN 9781139451994. URL https://books.google.co.za/books?id=bk2XEgz_ML4C.
- [25] G.F.R. Ellis, R. Maartens, and M.A.H. MacCallum. *Relativistic Cosmology*. Cambridge University Press, 2012. ISBN 9780521381154. URL <http://books.google.co.za/books?id=FPRFi3L0h9kC>.
- [26] K. Bolejko, A. Krasinski, C. Hellaby, and M. N. Celerier. *Structures in the universe by exact methods: Formation, evolutions, interactions*. Cambridge Univ. Press, Cambridge, 2010.
- [27] R. D'Inverno. *Introducing Einstein's Relativity*. Clarendon Press, 1992. ISBN 9780198596868. URL <http://books.google.co.za/books?id=8nw5fIWhkI4C>.
- [28] J. Ehlers. Contributions to the relativistic mechanics of continuous media. *General Relativity and Gravitation*, 25:1225–1266, December 1993. doi: 10.1007/BF00759031.
- [29] G. F. R. Ellis. Relativistic cosmology. *Gen.Rel.Grav.*, 41:581–660, 2009. doi: 10.1007/s10714-009-0760-7.
- [30] G. F. R. Ellis and H. van Elst. Cosmological models: Cargese lectures 1998. *NATO Adv.Study Inst.Ser.C.Math.Phys.Sci.*, 541:1–116, 1999.
- [31] T. Gebbie and G. F. R. Ellis. GIC approach to cosmic background radiation anisotropies. Part 1. *Annals Phys.*, 282:285–320, 2006. doi: 10.1006/aphy.2000.6033.
- [32] W. R. Stoeger, S. D. Nel, R. Maartens, and G. F. R. Ellis. The fluid-ray tetrad formulation of Einstein's field equations. *Class. Quantum Grav.*, 9:493–507, 1992. doi: 10.1088/0264-9381/9/2/013.
- [33] W. R. Stoeger, G. F. R. Ellis, and S. D. Nel. Observational cosmology. III. Exact spherically symmetric dust solutions. *Class. Quantum Grav.*, 9:509–525, 1992. doi: 10.1088/0264-9381/9/2/014.
- [34] W. R. Stoeger, S. J. Stanley, D. Nel, and G. F. R. Ellis. Observational cosmology. IV. Perturbed spherically symmetric dust solutions. *Class. Quantum Grav.*, 9:1711–1723, 1992. doi: 10.1088/0264-9381/9/7/007.
- [35] W. R. Stoeger, S. J. Stanley, D. Nel, and G. F. R. Ellis. Observational cosmology. V. Solution of the first-order general perturbation equations. *Class. Quantum Grav.*, 9:1725–1751, 1992. doi: 10.1088/0264-9381/9/7/008.

- [36] W. R. Stoeger, G. F. R. Ellis, and C. Xu. Observational cosmology. VI. The microwave background and the Sachs-Wolfe effect. *Phys. Rev. D*, 49:1845–1853, 1994. doi: 10.1103/PhysRevD.49.1845.
- [37] R. Maartens, N. P. Humphreys, D. R. Matravers, and W. R. Stoeger. Inhomogeneous universes in observational coordinates. *Class. Quantum Grav.*, 13:253–264, 1996. doi: 10.1088/0264-9381/13/2/013.
- [38] W. R. Stoeger, M. E. Araujo, and T. Gebbie. The Limits on Cosmological Anisotropies and Inhomogeneities from COBE Data. *Astrophys. J.*, 476:435–+, 1997. doi: 10.1086/303633.
- [39] M. E. Araújo and W. R. Stoeger. Exact spherically symmetric dust solution of the field equations in observational coordinates with cosmological data functions. *Phys. Rev. D*, 60(10):104020, 1999. doi: 10.1103/PhysRevD.60.104020.
- [40] M. E. Araújo and W. R. Stoeger. The angular-diameter distance maximum and its redshift as constraints on $\Lambda \neq 0$ Friedmann-Lemaître-Robertson-Walker models. *Mon. Not. Roy. Astr. Soc.*, 394:438–442, 2009. doi: 10.1111/j.1365-2966.2008.14321.x.
- [41] M. E. Araújo and W. R. Stoeger. Using time drift of cosmological redshifts to find the mass-energy density of the Universe. *Phys. Rev. D*, 82(12):123513–+, 2010. doi: 10.1103/PhysRevD.82.123513.
- [42] M. E. Araújo and W. R. Stoeger. Finding a spherically symmetric cosmology from observations in observational coordinates – advantages and challenges. *JCAP*, 7:29–+, 2011. doi: 10.1088/1475-7516/2011/07/029.
- [43] H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers, and E. Herlt. *Exact Solutions of Einstein's Field Equations*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2003. ISBN 9781139435024. URL <https://books.google.co.za/books?id=SiWXP8FjTFEC>.
- [44] F. K. Manasse and C. W. Misner. Fermi Normal Coordinates and Some Basic Concepts in Differential Geometry. *Journal of Mathematical Physics*, 4:735–745, June 1963. doi: 10.1063/1.1724316.
- [45] J. L. Feng. Dark Matter Candidates from Particle Physics and Methods of Detection. *ARA&A*, 48:495–545, September 2010. doi: 10.1146/annurev-astro-082708-101659.

- [46] Clifford M. W. The confrontation between general relativity and experiment. *Living Reviews in Relativity*, 9(3), 2006. doi: 10.1007/lrr-2006-3. URL <http://www.livingreviews.org/lrr-2006-3>.
- [47] A. Hees, O. Minazzoli, and J. Larena. Some cosmological consequences of a breaking of the Einstein equivalence principle. *ArXiv e-prints*, April 2015.
- [48] S. Tsujikawa. Quintessence: A Review. *Class. Quant. Grav.*, 30:214003, 2013. doi: 10.1088/0264-9381/30/21/214003.
- [49] S. Weinberg. The Cosmological Constant Problems (Talk given at Dark Matter 2000, February, 2000). *ArXiv Astrophysics e-prints*, May 2000.
- [50] R. Durrer and R. Maartens. Dark energy and dark gravity: theory overview. *General Relativity and Gravitation*, 40:301–328, February 2008. doi: 10.1007/s10714-007-0549-5.
- [51] S. M. Carroll. The Cosmological constant. *Living Rev. Rel.*, 4:1, 2001. doi: 10.12942/lrr-2001-1.
- [52] C. Clarkson, T. Clifton, A. Coley, and R. Sung. Observational constraints on the averaged universe. *Phys. Rev. D*, 85(4):043506, February 2012. doi: 10.1103/PhysRevD.85.043506.
- [53] C. Clarkson, G. F. R. Ellis, J. Larena, and O. Umeh. Does the growth of structure affect our dynamical models of the universe? The averaging, backreaction and fitting problems in cosmology. *Rept.Prog.Phys.*, 74:112901, 2011. doi: 10.1088/0034-4885/74/11/112901.
- [54] T. Buchert and S. Räsänen. Backreaction in late-time cosmology. *Ann.Rev.Nucl.Part.Sci.*, 62:57–79, 2012. doi: 10.1146/annurev.nucl.012809.104435.
- [55] T. Clifton. The method of images in cosmology. *Classical and Quantum Gravity*, 31(17):175010, September 2014. doi: 10.1088/0264-9381/31/17/175010.
- [56] V. A. A. Sanghai and T. Clifton. Post-Newtonian cosmological modelling. *Phys. Rev. D*, 91(10):103532, May 2015. doi: 10.1103/PhysRevD.91.103532.
- [57] J.-P. Uzan. Dynamics of relativistic interacting gases: from a kinetic to a fluid description. *Classical and Quantum Gravity*, 15:1063–1088, April 1998. doi: 10.1088/0264-9381/15/4/025.
- [58] G. Lemaître. l’Univers en expansion. *Annales de la Société Scientifique de Bruxelles*, 53:51–+, 1933. For an English translation see: The Expanding Universe. *Gen. Rel. Grav*, 29(5):641, 1997.

- [59] R.C. Tolman. Effect of inhomogeneity on cosmological models. *Proc. Nat. Acad., Sci.* 20:169, 1934.
- [60] H. Bondi. Spherically symmetrical models in general relativity. *Mon. Not. Roy. Astr. Soc.*, 107:410–+, 1947.
- [61] W. Valkenburg. Complete solutions to the metric of spherically collapsing dust in an expanding spacetime with a cosmological constant. *Gen.Rel.Grav.*, 44:2449–2476, 2012. doi: 10.1007/s10714-012-1405-9.
- [62] R. Sachs. Gravitational Waves in General Relativity. VI. The Outgoing Radiation Condition. *Royal Society of London Proceedings Series A*, 264:309–338, November 1961. doi: 10.1098/rspa.1961.0202.
- [63] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner. Gravitational Waves in General Relativity. VII. Waves from Axi-Symmetric Isolated Systems. *Royal Society of London Proceedings Series A*, 269:21–52, August 1962. doi: 10.1098/rspa.1962.0161.
- [64] G. Temple. New Systems of Normal Co-ordinates for Relativistic Optics. *Royal Society of London Proceedings Series A*, 168:122–148, October 1938. doi: 10.1098/rspa.1938.0164.
- [65] J. Kristian and R. K. Sachs. Observations in Cosmology. *Astro. Phys.*, 143:379, February 1966. doi: 10.1086/148522.
- [66] S. Räsänen. Light propagation and the average expansion rate in near-frw universes. *Phys. Rev. D*, 85:083528, Apr 2012. doi: 10.1103/PhysRevD.85.083528. URL <http://link.aps.org/doi/10.1103/PhysRevD.85.083528>.
- [67] O. Umeh, C. Clarkson, and R. Maartens. Nonlinear relativistic corrections to cosmological distances, redshift and gravitational lensing magnification: Ii. derivation. *Classical and Quantum Gravity*, 31(20):205001, 2014. URL <http://stacks.iop.org/0264-9381/31/i=20/a=205001>.
- [68] A. Sandage. The Change of Redshift and Apparent Luminosity of Galaxies due to the Deceleration of Selected Expanding Universes. *ApJ*, 136:319, September 1962. doi: 10.1086/147385.
- [69] A. Loeb. Direct Measurement of Cosmological Parameters from the Cosmic Deceleration of Extragalactic Objects. *Astrophys.J.*, 499:L111–L114, 1998. doi: 10.1086/311375.

- [70] P. Corasaniti, D. Huterer, and A. Melchiorri. Exploring the Dark Energy Redshift Desert with the Sandage-Loeb Test. *Phys.Rev.*, D75:062001, 2007. doi: 10.1103/PhysRevD.75.062001.
- [71] J. Geng, Y. Li, J. Zhang, and X. Zhang. Redshift drift exploration for interacting dark energy. *Eur. Phys. J.*, C75(8):356, 2015. doi: 10.1140/epjc/s10052-015-3581-8.
- [72] A. Balbi and C. Quercellini. The time evolution of cosmological redshift as a test of dark energy. *MNRAS*, 382:1623–1629, December 2007. doi: 10.1111/j.1365-2966.2007.12407.x.
- [73] J. P. Uzan, C. Clarkson, and G. F. R. Ellis. Time drift of cosmological redshifts as a test of the copernican principle. *Phys. Rev. Lett.*, 100:191303, May 2008. doi: 10.1103/PhysRevLett.100.191303. URL <http://link.aps.org/doi/10.1103/PhysRevLett.100.191303>.
- [74] L. Pasquini, S. Cristiani, R. Garcia-Lopez, M. Haehnelt, and M. Mayor. Codex: an ultra-stable high resolution spectrograph for the e-elt. *The Messenger*, 140:20–21, 2010.
- [75] I. Boisse, F. Bouchy, B. Chazelas, S. Perruchot, F. Pepe, C. Lovis, and G. Hébrard. Consequences of spectrograph illumination for the accuracy of radial-velocimetry. In *European Physical Journal Web of Conferences*, volume 16 of *European Physical Journal Web of Conferences*, page 2003, July 2011. doi: 10.1051/epjconf/20111602003.
- [76] B. Bassett and R. Hlozek. *Baryon acoustic oscillations*, page 246. Cambridge Univ. Press, 2010.
- [77] R. Maartens. Is the Universe homogeneous? *Royal Society of London Philosophical Transactions Series A*, 369:5115–5137, December 2011. doi: 10.1098/rsta.2011.0289.
- [78] C. Clarkson, B. Bassett, and T. H. Lu. A general test of the Copernican Principle. *Phys.Rev.Lett.*, 101:011301, 2008. doi: 10.1103/PhysRevLett.101.011301.
- [79] S. February, C. Clarkson, and R. Maartens. Galaxy correlations and the BAO in a void universe: structure formation as a test of the Copernican Principle. *JCAP*, 1303:023, 2013. doi: 10.1088/1475-7516/2013/03/023.
- [80] P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos. Late-time cosmology with 21cm intensity mapping experiments. *Astrophys.J.*, 803(1):21, 2015. doi: 10.1088/0004-637X/803/1/21.

- [81] R. Jimenez and A. Loeb. Constraining Cosmological Parameters Based on Relative Galaxy Ages. *ApJ*, 573:37–42, July 2002. doi: 10.1086/340549.
- [82] J. Dunlop, J. Peacock, H. Spinrad, A. Dey, R. Jimenez, D. Stern, and R. Windhorst. A 3.5-Gyr-old galaxy at redshift 1.55. *Nature*, 381:581–584, June 1996. doi: 10.1038/381581a0.
- [83] A. Renzini. Stellar Population Diagnostics of Elliptical Galaxy Formation. *ARA&A*, 44: 141–192, September 2006. doi: 10.1146/annurev.astro.44.051905.092450.
- [84] M. Moresco, R. Jimenez, A. Cimatti, and L. Pozzetti. Constraining the expansion rate of the Universe using low-redshift ellipticals as cosmic chronometers. *J. Cosmology Astropart. Phys.*, 3:045, March 2011. doi: 10.1088/1475-7516/2011/03/045.
- [85] M. Moresco, A. Cimatti, R. Jimenez, L. Pozzetti, G. Zamorani, M. Bolzonella, J. Dunlop, F. Lamareille, M. Mignoli, H. Pearce, P. Rosati, D. Stern, L. Verde, E. Zucca, C. M. Carollo, T. Contini, J.-P. Kneib, O. Le Fèvre, S. J. Lilly, V. Mainieri, A. Renzini, M. Scoddegio, I. Balestra, R. Gobat, R. McLure, S. Bardelli, A. Bongiorno, K. Caputi, O. Cucciati, S. de la Torre, L. de Ravel, P. Franzetti, B. Garilli, A. Iovino, P. Kampeczyk, C. Knobel, K. Kovač, J.-F. Le Borgne, V. Le Brun, C. Maier, R. Pelló, Y. Peng, E. Perez-Montero, V. Presotto, J. D. Silverman, M. Tanaka, L. A. M. Tasca, L. Tresse, D. Vergani, O. Almaini, L. Barnes, R. Bordoloi, E. Bradshaw, A. Cappi, R. Chuter, M. Cirasuolo, G. Coppa, C. Diener, S. Foucaud, W. Hartley, M. Kamionkowski, A. M. Koekemoer, C. López-Sanjuan, H. J. McCracken, P. Nair, P. Oesch, A. Stanford, and N. Welikala. Improved constraints on the expansion rate of the Universe up to $z \sim 1.1$ from the spectroscopic evolution of cosmic chronometers. *J. Cosmology Astropart. Phys.*, 8:006, August 2012. doi: 10.1088/1475-7516/2012/08/006.
- [86] M. Moresco, L. Verde, L. Pozzetti, R. Jimenez, and A. Cimatti. New constraints on cosmological parameters and neutrino properties using the expansion rate of the Universe to $z \sim 1.75$. *J. Cosmology Astropart. Phys.*, 7:053, July 2012. doi: 10.1088/1475-7516/2012/07/053.
- [87] M. Moresco. Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at $z \sim 2$. *MNRAS*, 450:L16–L20, June 2015. doi: 10.1093/mnrasl/slv037.
- [88] G Bruzual and S Charlot. Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344(4):1000–1028, 2003.

- [89] R. de Putter, L. Verde, and R. Jimenez. Testing LTB Void Models Without the Cosmic Microwave Background or Large Scale Structure: New Constraints from Galaxy Ages. 2012.
- [90] S. February, J. Larena, C. Clarkson, and D. Pollney. Evolution of linear perturbations in spherically symmetric dust spacetimes. *Class. Quant. Grav.*, 31:175008, 2014. doi: 10.1088/0264-9381/31/17/175008.
- [91] P. Astier. The expansion of the universe observed with supernovae. *Reports on Progress in Physics*, 75:116901, 2012. doi: 10.1088/0034-4885/75/11/116901.
- [92] D. A. Howell, A. Conley, M. Della Valle, P. E. Nugent, S. Perlmutter, G. H. Marion, K. Krisciunas, C. Badenes, P. Mazzali, G. Aldering, P. Antilogus, E. Baron, A. Becker, C. Baltay, S. Benetti, S. Blondin, D. Branch, E. F. Brown, S. Deustua, A. Ealet, R. S. Ellis, D. Fouchez, W. Freedman, A. Gal-Yam, S. Jha, D. Kasen, R. Kessler, A. G. Kim, D. C. Leonard, W. Li, M. Livio, D. Maoz, F. Mannucci, T. Matheson, J. D. Neill, K. Nomoto, N. Panagia, K. Perrett, M. Phillips, D. Poznanski, R. Quimby, A. Rest, A. Riess, M. Sako, A. M. Soderberg, L. Strolger, R. Thomas, M. Turatto, S. van Dyk, and W. M. Wood-Vasey. Type Ia supernova science 2010-2020. *ArXiv e-prints*, March 2009.
- [93] I. M. H. Etherington. On the Definition of Distance in General Relativity. *Philosophical Magazine*, 15:761, 1933.
- [94] W. Baade and F. Zwicky. On Super-novae. *Proceedings of the National Academy of Science*, 20:254–259, May 1934. doi: 10.1073/pnas.20.5.254.
- [95] R. Minkowski. Supernovae and Supernova Remnants. *ARA&A*, 2:247, 1964. doi: 10.1146/annurev.aa.02.090164.001335.
- [96] J. C. Wheeler and R. Levreault. The peculiar Type I supernova in NGC 991. *ApJ*, 294:L17–L20, July 1985. doi: 10.1086/184500.
- [97] J. C. Wheeler. Resource Letter: OTS-1: Observations and theory of supernovae. *American Journal of Physics*, 71:11–22, January 2003. doi: 10.1119/1.1523076.
- [98] S. E. Woosley and T. A. Weaver. The physics of supernova explosions. *ARA&A*, 24:205–253, 1986. doi: 10.1146/annurev.aa.24.090186.001225.
- [99] A. V. Filippenko. Optical Spectra of Supernovae. *ARA&A*, 35:309–355, 1997. doi: 10.1146/annurev.astro.35.1.309.

- [100] S. Chandrasekhar. The Maximum Mass of Ideal White Dwarfs. *ApJ*, 74:81, July 1931. doi: 10.1086/143324.
- [101] F. Hoyle and W. A. Fowler. Nucleosynthesis in Supernovae. *ApJ*, 132:565, November 1960. doi: 10.1086/146963.
- [102] K. Nomoto, F.-K. Thielemann, and K. Yokoi. Accreting white dwarf models of Type I supernovae. III - Carbon deflagration supernovae. *ApJ*, 286:644–658, November 1984. doi: 10.1086/162639.
- [103] M. Livio. The Progenitors of Type Ia Supernovae. In J. C. Niemeyer and J. W. Truran, editors, *Type Ia Supernovae, Theory and Cosmology*, page 33, 2000.
- [104] D. A. Howell. Type Ia supernovae as stellar endpoints and cosmological tools. *Nature Communications*, 2:350, June 2011. doi: 10.1038/ncomms1344.
- [105] R. Amanullah, C. Lidman, D. Rubin, G. Aldering, P. Astier, K. Barbary, M. S. Burns, A. Conley, K. S. Dawson, S. E. Deustua, M. Doi, S. Fabbro, L. Faccioli, H. K. Fakhouri, G. Folatelli, A. S. Fruchter, H. Furusawa, G. Garavini, G. Goldhaber, A. Goobar, D. E. Groom, I. Hook, D. A. Howell, N. Kashikawa, A. G. Kim, R. A. Knop, M. Kowalski, E. Linder, J. Meyers, T. Morokuma, S. Nobili, J. Nordin, P. E. Nugent, L. Östman, R. Pain, N. Panagia, S. Perlmutter, J. Raux, P. Ruiz-Lapuente, A. L. Spadafora, M. Strovink, N. Suzuki, L. Wang, W. M. Wood-Vasey, N. Yasuda, and T. Supernova Cosmology Project. Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at $0.511 < z < 1.12$ and the Union2 Compilation. *ApJ*, 716:712–738, June 2010. doi: 10.1088/0004-637X/716/1/712.
- [106] J. Guy, P. Astier, S. Baumont, D. Hardin, R. Pain, N. Regnault, S. Basa, R. G. Carlberg, A. Conley, S. Fabbro, D. Fouchez, I. M. Hook, D. A. Howell, K. Perrett, C. J. Pritchett, J. Rich, M. Sullivan, P. Antilogus, E. Aubourg, G. Bazin, J. Bronder, M. Filiol, N. Palanque-Delabrouille, P. Ripoche, and V. Ruhlmann-Kleider. SALT2: using distant supernovae to improve the use of type Ia supernovae as distance indicators. *A&A*, 466: 11–21, April 2007. doi: 10.1051/0004-6361:20066930.
- [107] P. L. Kelly, M. Hicken, D. L. Burke, K. S. Mandel, and R. P. Kirshner. Hubble Residuals of Nearby Type Ia Supernovae are Correlated with Host Galaxy Masses. *ApJ*, 715:743–756, June 2010. doi: 10.1088/0004-637X/715/2/743.
- [108] M. Sullivan, A. Conley, D. A. Howell, J. D. Neill, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, J. Guy, D. Hardin, I. M. Hook, R. Pain, N. Palanque-Delabrouille,

- K. M. Perrett, C. J. Pritchett, N. Regnault, J. Rich, V. Ruhlmann-Kleider, S. Baumont, E. Hsiao, T. Kronborg, C. Lidman, S. Perlmutter, and E. S. Walker. The dependence of Type Ia Supernovae luminosities on their host galaxies. *MNRAS*, 406:782–802, August 2010. doi: 10.1111/j.1365-2966.2010.16731.x.
- [109] A. Conley, G. Goldhaber, L. Wang, G. Aldering, R. Amanullah, E. D. Commins, V. Fadeyev, G. Folatelli, G. Garavini, R. Gibbons, A. Goobar, D. E. Groom, I. Hook, D. A. Howell, A. G. Kim, R. A. Knop, M. Kowalski, N. Kuznetsova, C. Lidman, S. Nobili, P. E. Nugent, R. Pain, S. Perlmutter, E. Smith, A. L. Spadafora, V. Stanishev, M. Strovink, R. C. Thomas, W. M. Wood-Vasey, and Supernova Cosmology Project. Measurement of Ω_m , Ω_Λ from a Blind Analysis of Type Ia Supernovae with CMAGIC: Using Color Information to Verify the Acceleration of the Universe. *ApJ*, 644:1–20, June 2006. doi: 10.1086/503533.
- [110] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz. Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. *ApJ*, 688:709–728, December 2008. doi: 10.1086/591439.
- [111] M. B. Ribeiro and W. R. Stoeger. Relativistic Cosmology Number Counts and the Luminosity Function. *ApJ*, 592:1–16, July 2003. doi: 10.1086/375717.
- [112] V. V. L. Albani, A. S. Iribarrem, M. B. Ribeiro, and W. R. Stoeger. Differential Density Statistics of the Galaxy Distribution and the Luminosity Function. *ApJ*, 657:760–772, March 2007. doi: 10.1086/510520.
- [113] A. S. Iribarrem, A. R. Lopes, M. B. Ribeiro, and W. R. Stoeger. Relativistic cosmology number densities and the luminosity function. *A&A*, 539:A112, March 2012. doi: 10.1051/0004-6361/201117535.
- [114] A. Iribarrem, P. Andreani, C. Gruppioni, S. February, M. B. Ribeiro, S. Berta, E. Le Floch, B. Magnelli, R. Nordon, P. Popesso, F. Pozzi, and L. Riguccini. Cosmological model dependence of the galaxy luminosity function: far-infrared results in the Lemaître-Tolman-Bondi model. *A&A*, 558:A15, October 2013. doi: 10.1051/0004-6361/201321396.
- [115] A. Iribarrem, P. Andreani, S. February, C. Gruppioni, A. R. Lopes, M. B. Ribeiro, and W. R. Stoeger. Relativistic cosmology number densities in void-Lemaître-Tolman-Bondi models. *A&A*, 563:A20, March 2014. doi: 10.1051/0004-6361/201322507.
- [116] N. Kaiser. On the spatial correlations of Abell clusters. *ApJ*, 284:L9–L12, September 1984. doi: 10.1086/184341.

- [117] R. Johnston. Shedding light on the galaxy luminosity function. *A&A Rev.*, 19:41, August 2011. doi: 10.1007/s00159-011-0041-9.
- [118] W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974. doi: 10.1086/152650.
- [119] H. J. Mo and S. D. M. White. An Analytic model for the spatial clustering of dark matter halos. *Mon. Not. Roy. Astron. Soc.*, 282:347, 1996. doi: 10.1093/mnras/282.2.347.
- [120] M. Tegmark and P. J. E. Peebles. The Time Evolution of Bias. *ApJ*, 500:L79–L82, June 1998. doi: 10.1086/311426.
- [121] P. J. van der Walt and N. T. Bishop. Observational cosmology using characteristic numerical relativity: Characteristic formalism on null geodesics. *Phys. Rev. D*, 85(4):044016, February 2012. doi: 10.1103/PhysRevD.85.044016.
- [122] J. J. Cowan, F.-K. Thielemann, and J. W. Truran. Radioactive dating of the elements. *ARA&A*, 29:447–497, 1991. doi: 10.1146/annurev.aa.29.090191.002311.
- [123] B. Chaboyer, P. Demarque, P. J. Kernan, and L. M. Krauss. A Lower Limit on the Age of the Universe. *Science*, 271:957–961, February 1996. doi: 10.1126/science.271.5251.957.
- [124] B. Chaboyer. The age of the universe. *Phys. Rep.*, 307:23–30, December 1998. doi: 10.1016/S0370-1573(98)00054-4.
- [125] L. M. Krauss and B. Chaboyer. New Globular Cluster Age Estimates and Constraints on the Cosmic Equation of State and The Matter Density of the Universe. *ArXiv*, November 2001.
- [126] V. Hill, B. Plez, R. Cayrel, T. C. Beers, B. Nordström, J. Andersen, M. Spite, F. Spite, B. Barbuy, P. Bonifacio, E. Depagne, P. François, and F. Primas. First stars. I. The extreme r-element rich, iron-poor halo giant CS 31082-001. Implications for the r-process site(s) and radioactive cosmochronology. *A&A*, 387:560–579, May 2002. doi: 10.1051/0004-6361:20020434.
- [127] C. Sneden, A. McWilliam, G. W. Preston, J. J. Cowan, D. L. Burris, and B. J. Armosky. The Ultra-Metal-poor, Neutron-Capture-rich Giant Star CS 22892-052. *ApJ*, 467:819, August 1996. doi: 10.1086/177656.

- [128] N. T. Bishop, R. Gomez, P. R. Holvorcem, R. A. Matzner, P. Papadopoulos, and J. Winicour. Cauchy-characteristic evolution and waveforms. *Journal of Computational Physics*, 136(1):140 – 167, 1997. ISSN 0021-9991. doi: 10.1006/jcph.1997.5754. URL <http://www.sciencedirect.com/science/article/pii/S0021999197957545>.
- [129] N. T. Bishop, R. Gómez, L. Lehner, M. Maharaj, and J. Winicour. High-powered gravitational news. *Phys. Rev. D*, 56:6298–6309, November 1997. doi: 10.1103/PhysRevD.56.6298.
- [130] P. J. van der Walt and N. T. Bishop. Observational cosmology using characteristic numerical relativity. *Phys. Rev. D*, 82(8):084001, October 2010. doi: 10.1103/PhysRevD.82.084001.
- [131] P. J. van der Walt. *Numerical Relativity on Cosmological Past Null Cones*. PhD thesis, Rhodes University, 2012.
- [132] R.H. Pletcher, J.C. Tannehill, and D. Anderson. *Computational Fluid Mechanics and Heat Transfer, Second Edition*. Series in Computational and Physical Processes in Mechanics and Thermal Sciences. Taylor & Francis, 1997. ISBN 9781560320463. URL <https://books.google.co.za/books?id=ZJPbtHeilCgC>.
- [133] F. (Faming) Liang, Chuanhai Liu, and Raymond J. Carroll. *Advanced Markov chain Monte Carlo methods : learning from past samples*. Wiley series in computational statistics. Wiley, Chichester, West Sussex, U.K., 2010. ISBN ISBN: 9780470748268. Includes bibliographical references and index.
- [134] S. Brooks, A. Gelman, G. Jones, and X.L. Meng. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2011. ISBN 9781420079425. URL <https://books.google.co.za/books?id=qfRsAIKZ4rIC>.
- [135] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.
- [136] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970. doi: 10.1093/biomet/57.1.97.
- [137] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. Mcmc methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 08 2013. doi: 10.1214/13-STS421. URL <http://dx.doi.org/10.1214/13-STS421>.

- [138] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 02 1997. doi: 10.1214/aoap/1034625254. URL <http://dx.doi.org/10.1214/aoap/1034625254>.
- [139] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511, 1992. <http://www.stat.columbia.edu/~gelman/research/published/itsim.pdf>.
- [140] M. K. Cowles and B. P. Carlin. Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):pp. 883–904, 1996. ISSN 01621459. URL <http://www.jstor.org/stable/2291683>.
- [141] Z. Li and H. A. Scheraga. Monte carlo-minimization approach to the multiple-minima problem in protein folding. Technical report, Academy of Sciences of the United States of America, 1987. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC299132/>.
- [142] J. Garcia-Bellido and T. Haugbølle. Confronting lemaitre–tolman–bondi models with observational cosmology. *Journal of Cosmology and Astroparticle Physics*, 2008(04):003, 2008. URL <http://stacks.iop.org/1475-7516/2008/i=04/a=003>.
- [143] M. Redlich, K. Bolejko, S. Meyer, G. F. Lewis, and M. Bartelmann. Probing spatial homogeneity with LTB models: a detailed discussion. *A&A*, 570:A63, October 2014. doi: 10.1051/0004-6361/201424553.
- [144] M. Seikel, C. Clarkson, and M. Smith. Reconstruction of dark energy and expansion dynamics using Gaussian processes. *JCAP*, 1206:036, 2012. doi: 10.1088/1475-7516/2012/06/036.
- [145] A. Shafieloo, A. G. Kim, and E. V. Linder. Gaussian process cosmography. *Phys. Rev. D*, 85(12):123530, June 2012. doi: 10.1103/PhysRevD.85.123530.
- [146] V. C. Busti, C. Clarkson, and M. Seikel. Evidence for a lower value for H_0 from cosmic chronometers data? *MNRAS*, 441:L11–L15, June 2014. doi: 10.1093/mnrasl/slu035.
- [147] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. Adaptive computation and machine learning series. University Press Group Limited, 2006. ISBN 9780262182539. URL <http://books.google.ca/books?id=vWtwQgAACAAJ>.
- [148] M. A. Álvarez, D. Luengo, and N. D. Lawrence. Linear Latent Force Models using Gaussian Processes. *ArXiv e-prints*, July 2011.

- [149] J. Hartikainen, M. Seppanen, and S. Sarkka. State-Space Inference for Non-Linear Latent Force Models with Application to Satellite Orbit Prediction. *ArXiv e-prints*, June 2012.
- [150] M. A. Álvarez, D. Luengo, and N. D. Lawrence. Latent Force Models. *AISTATS: Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, 2009. URL <http://jmlr.csail.mit.edu/proceedings/papers/v5/alvarez09a/alvarez09a.pdf>.
- [151] M. K. Titsias and N. D. Lawrence. Bayesian Gaussian Process Latent Variable Model. *AISTATS: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 2010. URL http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS2010_TitsiasL10.pdf.
- [152] R. Turner, M. P. Deisenroth, and C. E. Rasmussen. State-Space Inference and Learning with Gaussian Processes. *AISTATS: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 2010. URL http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS2010_TurnerDR10.pdf.
- [153] N. D. Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *JMLR: Workshop and Conference Proceedings vol 6:1783-1816*, 2005. URL <http://jmlr.org/papers/volume6/lawrence05a/lawrence05a.pdf>.
- [154] R. Boloix-Tortosa, F. J. Payán-Somet, E. Arias-de-Reyna, and J. José Murillo-Fuentes. Proper Complex Gaussian Processes for Regression. *ArXiv e-prints*, February 2015.
- [155] D. W Scott. Scott’s rule. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):497–502, 2010.
- [156] H. L. Bester, J. Larena, and N. T. Bishop. Numerically reconstructing the geometry of the Universe from data. 2016. URL <http://inspirehep.net/record/1417086/files/arXiv:1601.07362.pdf>.
- [157] J Liske, A Grazian, E Vanzella, M Dessauges, M Viel, L Pasquini, M Haehnelt, S Cristiani, F Pepe, G Avila, et al. Cosmic dynamics in the era of extremely large telescopes. *Monthly notices of the royal astronomical society*, 386(3):1192–1218, 2008.
- [158] R. A. Sunyaev and Y. B. Zeldovich. Small-Scale Fluctuations of Relic Radiation. *Ap&SS*, 7:3–19, April 1970. doi: 10.1007/BF00653471.

- [159] R. A. Sunyaev and I. B. Zeldovich. The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement. *MNRAS*, 190:413–420, February 1980.
- [160] P. Bull, T. Clifton, and P. G. Ferreira. Kinematic sunyaev-zel’dovich effect as a test of general radial inhomogeneity in lemaître-tolman-bondi cosmology. *Phys. Rev. D*, 85:024002, Jan 2012. doi: 10.1103/PhysRevD.85.024002. URL <http://link.aps.org/doi/10.1103/PhysRevD.85.024002>.
- [161] J. García-Bellido and T. Haugbølle. Looking the void in the eyes - the kinematic Sunyaev Zeldovich effect in Lemaître Tolman Bondi models. *J. Cosmology Astropart. Phys.*, 9:016, September 2008. doi: 10.1088/1475-7516/2008/09/016.
- [162] W. C. Lim, M. Regis, and C. Clarkson. Spherically symmetric cosmological spacetimes with dust and radiation — numerical implementation. *JCAP*, 1310:010, 2013. doi: 10.1088/1475-7516/2013/10/010.
- [163] M. Vonlanthen, S. Rasanen, and R. Durrer. Model-independent cosmological constraints from the CMB. *JCAP*, 1008:023, 2010. doi: 10.1088/1475-7516/2010/08/023.
- [164] P. Zhang and A. Stebbins. Confirmation of the Copernican Principle at Gpc Radial Scale and above from the Kinetic Sunyaev-Zel’dovich Effect Power Spectrum. *Physical Review Letters*, 107(4):041301, July 2011. doi: 10.1103/PhysRevLett.107.041301.
- [165] J. P. Zibin and A. Moss. Nowhere to hide: closing in on cosmological homogeneity. 2014.