:: Unified

Dark Energy Models

Luís M.G. Beça ::

Universidade do Porto

Departamento de Física

:: Unified Dark Energy Models ::

Author:
Luís M.G. Beça

Supervisor:
Prof. Pedro P. Avelino

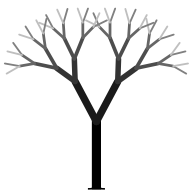
January 2008

Universidade do Porto

Departamento de Física

:: Unified Dark Energy Models ::

Tese submetida à Faculdade de Ciências
da Universidade do Porto para obtenção do
Grau de Doutor em Física



To my
Mom & Dad

Acknowledgments

First and foremost, I would like to thank Pedro Avelino, for his outstanding patience and support. His continuous encouragements have had a significant impact on my life, for which I'm forever in his debt. My sincere gratitude to Carlos Martins and Paulo Maurício for their invaluable help in several stages of this work. Also, I am in debt to my fellow colleagues Joana Oliveira, Josinaldo Menezes, and João Penedones for their friendship and support. A very special thanks to Carlos Herdeiro for sharing, countless times, his unique views, and to Caroline Santos for the many useful pointers along the way. Likewise, I am grateful to all members of Centro de Física do Porto for maintaining a friendly, easy-going, work environment and, in particular, to Miguel Costa. Also, my gratitude to Fundação para a Ciência e Tecnologia for funding this work under grant SFRH/BD/9302/2002.

To my Mom and Dad, for keeping up with my foul moods, always with love and support. To my dear friends (you know who you are) and family for always being just a phone call away. And finally, to Daniela, my deepest gratitude and never ending love for having changed my life in wonderful and unexpected ways.

Thank you all.

Luís M.G. Beça

Jan 2008

Resumo

Estuda-se a possibilidade, do ponto de vista fenomonológico, de unificar a matéria escura e a energia escura numa só componente isentrópica ('quartessence' canónica), no contexto da Relatividade Geral. Em particular, estuda-se o gás de Chaplygin generalizado (gCg) como protótipo de quartessence e determinam-se constrangimentos de ordem zero e linear. Conclui-se que o gCg se tem de comportar de uma forma muito semelhante a ΛCDM, um resultado que tem sido interpretado como o fim da energia escura unificada (EEU). Argumenta-se que esta conclusão é, em grande parte, prematura. Ao analisar-se o início do regime não-linear, põe-se em causa a validade dos métodos tradicionais perturbativos usados no âmbito da quartessence. Mostra-se, com efeito, que o colapso não linear de pequena escala, pode afectar o comportamento de larga escala do Universo de uma forma que estes métodos não levam em conta. Conclui-se que somente resolvendo as equações de Einstein por completo, se pode obter um veredicto final sobre a EEU. Algumas ideias simples sobre como contornar esta dificuldade são sugeridas.

Palavras-chave: Energia escura, Matéria escura, Quartessence, Gás de Chaplygin, Cosmologia, Relatividade Geral, Perturbações não-lineares

Résumé

La possibilité d'unifier la matière noire et l'énergie noire en un seul fluide isentropique ('quatressence' canonique) dans le contexte de la Relativité Générale est étudié du point de vue phénoménologique. En particulier, on étudie le gaz de Chaplygin généralisé (gCg) comme un prototype de la quatressence et on détermine les constraintes d'ordre zéro et linéaire. On peut conclure que le comportement du gCg est très semblable à ΛCDM, un résultat qui a été interpreté souvent comme la fin de l'énergie noire unifiée (ENU). On argument que cette conclusion est assez prémature. En analisant le commencement du régime non-linéaire, la validité des métodes perturbatives traditionelles utilisées au contexte de la quatressence, est mise en doute. On montre que l'agglomération non-linéaire, même à petite échelle, affecte le comportement à large échelle de l'Universe d'une façon que les métodes traditionelles ne sont pas capables de tenir en compte. On peut conclure que pour obtenir un veredict final sur l'ENU, on doit résoudre complètement (ça veut dire sans aucune approximation) les équations d'Einstein. Quelques simples idées pour traiter ce problème sons suggérées.

Mots-clés: Énergie Noire, Matière Noire, Quatressence, Gaz de Chaplygin, Cosmologie, Relativitée Genérale, Perturbations Non-Lineaires

Abstract

We study, from a phenomenological perspective, the possibility of unifying dark matter and dark energy into a single isentropic fluid (canonic quartessence), in the framework of General Relativity. In particular, we study the generalized Chaplygin gas (gCg) as a prototype for quartessence and determine background and linear constrains. We find that the gCg has to behave in a manner very similar to ΛCDM, a result widely seen as the end of unified dark energy (UDE). We argue that this is mostly a premature conclusion. By analyzing the onset of the non-linear regime, we bring into serious question the validity of traditional perturbative methods in the context of UDE. We show that non-linear clustering, even on small scales, affects the average pressure of the Universe in a manner that traditional methods fail to take into account. We conclude that only by solving the full Einstein field equations, can a definite answer be obtained regarding the validity of the UDE hypothesis. Some simple ideas on how to improve this situation are subsequently offered.

Key-words: Dark energy, Dark Matter, Quartessence, Chaplygin gas, Cosmology, General Relativity, Non-Linear Perturbations

Notations and Units

Throughout this thesis, we employ a $(-+++)$ signature for the spacetime metric $g_{\mu\nu}$. Greek indices run over spacetime coordinates, Latin indices run over space coordinates. $\mathcal{M}$ represents the inhomogeneous Universe, $\langle\mathcal{M}\rangle$, the average large scale background. If the global and local dynamics in $\mathcal{M}$ are the same, the symbol $\overline{\mathcal{M}}$ is used for the background instead of $\langle\mathcal{M}\rangle$. We follow the convention that quantities in $\overline{\mathcal{M}}$ are denoted with a bar on top, like so $\bar{Q}$, and no bar for quantities in $\mathcal{M}$. However, if Q is obviously a background quantity like the scale factor a or the Hubble parameter H (or $\mathcal{H}$, if conformal time is used), no bar is employed. The same symbol x^μ for coordinates in $\mathcal{M}$ and $\overline{\mathcal{M}}$ is used; there is no need to carefully distinguish $\bar{x}^\mu$ from x^μ because it is always possible to drag the coordinates from one manifold to the other by means of any diffeomorphism linking the two. Unless otherwise stated, natural units are used throughout.

Preface

Today, the macroscopic aspects of the Universe seem well understood. Multiple observations strongly suggest that we live in a (nearly) flat Universe, presently undergoing an accelerating phase [1, 2, 3, 4, 5, 6]. In the context of General Relativity, this acceleration can only be explained by the presence of an 'exotic' dark energy component violating the strong energy condition. Exactly what constitutes this energy, no one knows; it stands as one of the biggest mysteries in contemporary Cosmology. Theoretically, the simplest way to achieve this acceleration is through a cosmological constant [7]; unfortunately, at the present, we have no clear understanding of how this 'vacuum energy' arises. There are many other possible constructs, though, that can achieve this same large scale dynamics; however, they are mostly *canonic* (e.g. quintessence [8, 9]) and *non*-canonic (e.g. k-essence [10], phantom energy [11, 12], tachyons [13, 14], vacuum metamorphosis [15, 16], etc.) generalizations of the cosmological constant that also lack any solid foundations. This makes it hard to meaningfully compare them. This is not to say, of course, that some do not have better 'traits' than others; quintessence, after all, is often described as an enhanced cosmological constant. Nevertheless, these improvements are hardly rooted in any current fundamental understanding we have. Thus, barring any significant progress at the 'fundamental physics' front, we really have no way of discerning which alternatives are indeed *better*. Even worse is the fact that no 'direct' detection of dark energy has ever been made; we're not even sure if it *really* exists. As it happens, most theoretical physicists these days don't actually believe that General Relativity is the final word on gravity. Thus, it is quite possible that dark energy doesn't actually exist, in other words, gravity itself may be causing the

observed acceleration [17]. Indeed, by modifying General Relativity directly, it is possible to reproduce the current acceleration of the Universe without involving any dark energy. (There are many ways to achieve this but normally they all involve extra-dimensions. *Branes* [18], for instance, are a famous example.) At the moment, however, most modifications are still far too 'ad hoc' to be particularly pleasing, and are hardily fundamental well motivated, anyway. By tinkering with the large scale behavior of gravity, though, we haven't really closed any Pandora's Box regarding dark energy, we only pried it open even more. For our part, in this thesis, we will always work within the confines of General Relativity, by treating dark energy as a real entity.

There are also multiple observations that strongly suggest that most 'matter' in the Universe, an essential ingredient for structure formation, is in a dark *non*-baryonic form [3, 19, 4, 6, 20]. Also here, the exact nature of this non-baryonic dark matter is not known and possibilities abound; these commonly include all sorts of non-standard particles, from axions to the neutralino, and so forth. Other more exotic possibilities, involve direct modifications to General Relativity; these go from the latest revised relativist versions of MOND [21], to a non-symmetric modified gravity by Moffat [22, 23, 24]. In all of these, dark matter is traded for a much more complicated TeVeS (tensor, vector, scalar) gravity source. Again, we won't consider this route by always staying within General Relativity.

It is perhaps the greatest achievement of 20th-century Cosmology, that we are now in possession of a fairly accurate and complete inventory of the energy content of the Universe. Roughly speaking, we know today that only about 4% of the Universe is made of ordinary baryonic matter (of which roughly a quarter is actually visible), 26% non-baryonic dark matter and the remaining 70%, dark energy (and an insignificant amount of radiation). It is quite astonishing to realize that the vast majority of the Universe has yet to be 'seen'; then again, we could take this as a sign that something is wrong with General Relativity. 21st-century Cosmology is thus hard-pressed to explain what these unseen components *are*; this turns out to be a very difficult thing to do. At the most basic level (assuming they exist),

for instance, we're not even sure if dark energy and dark matter are fundamentally *different* from one another (a widespread *belief*, nonetheless). In reality, we only know that macroscopically they play different roles but, a priori, this doesn't actually force them to be different. It is certainly conceivable, at least theoretically so, that both could share a common origin. Given the current state of affairs, with no direct detection foreseeable in the near future, and no shortage of candidates also, the possibility that dark energy and dark matter are, somehow, just different *manifestations* of a single entity, should not be discarded lightly. In fact, it only stands to reason that such a possibility should be throughly investigated. A negative outcome would, at least, signal the fundamental difference between dark energy and dark matter. Either way, something of value could be gained.

Unified Dark Energy (UDE for short) models (sometimes also called unified dark matter models) are thus built upon the simplifying hypothesis that a *single* component (dubbed 'quartessence') simultaneously accounts for both dark energy and dark matter. Historically, the idea of UDE has sprung from the unusual proprieties of the Chaplygin gas [25, 26], an exotic fluid with an equation of state $p = -A/\varepsilon$, where A is a positive constant: it turns out that this fluid behaves as (pressureless) matter very early in the history of the Universe, and as a cosmological constant much later (smoothly transitioning between the two), a dual behaviour highly suggestive of a unified description of dark energy/dark matter. This gas is also special in that it can be 'motivated' in the context of string theory by considering a d-brane in a spacetime of $d + 2$ dimensions. Then, the Nambu-Goto action can be seen as describing a 'Newtonian' fluid with the above equation of state [27, 28], the negative pressure interpreted as the brane tension. This interpretation has made the Chaplygin gas the currently preferred prototype for quartessence; on the other hand, it is hardly a fundamental interpretation, so we shouldn't give it too much importance. [29] generalizes this fluid to a broader class parameterized by $p = -A/\varepsilon^\alpha$ where α is a constant. Unfortunately, the pressure of this so-called generalized Chaplygin gas (gCg) is no longer easily interpretable as a d-brane tension; consequently, some of the initial appeal is lost. Nevertheless, this does not make it any less useful to us, especially in the context of a phenomenological analysis. This is because the gCg

covers, in a continuous fashion, a wide gamut of quartessence models, from ΛCDM (a limiting case of UDE corresponding to $\alpha = 0$, as we will later show in detail) to the original Chaplygin gas. The real question then becomes whether or not the gCg (and, hence, quartessence in general) constitutes a viable alternative to an already macroscopically successful (non-unified) ΛCDM model of the Universe.

This work is an attempt to answer this question *phenomenologically*. Several tests are discussed, mainly of zero and linear order, and how they constrain the parameter space of the gCg. It turns out that an α close to zero is significantly favored by current observations, a result widely seen as indicating the failure of UDE. That this is mostly a premature conclusion, is perhaps one of the most relevant contributions we make to this subject. Our reasoning is grounded in the realization that non-linear effects are of critical importance for all UDE models. In broad terms, this is related to the fact that the average pressure $\langle p \rangle \equiv \langle -A/\varepsilon^{\alpha} \rangle$ of an inhomogeneous gCg manifold $\mathcal{M}$ *is not* $-A/\langle \varepsilon \rangle^{\alpha}$ (where $\langle\ \rangle$ represents a suitable spatial average) unless, of course, the perturbations are very small. This simple observation highlights the fact that an average gCg universe, i.e., $\langle \mathcal{M} \rangle$ doesn't behave, in general, as a spatially homogeneous gCg. This poses a serious problem for traditional perturbative methods for the simple reason that we need to know upfront how the background evolves in order to build perturbations on top of it. Unfortunately, in the case of the inhomogeneous gCg we simply don't know this; to find out, we would have to solve the full Einstein field equations and subsequently smooth any solutions we could muster. It goes without saying that this is a notoriously difficult task by any standard; if we could do it in general, perturbative methods would hardly have any 'raison d'être'. It is, of course, always possible to start out with a homogeneous gCg background and perturb it, like most of us have; the real problem is that we may be perturbing the *wrong* background. By building perturbations on top a homogenous gCg background, we are effectively ignoring the potential effect that non-linear small scale clustering may have on the actual background as a whole. (This is strongly suggested by the fact that the average equation of state of the inhomogeneous gCg differs from the homogenous one.) It is true that in 'most' cosmological models, non-linear small scale clustering can be 'swept under the rug'; there is a 'natural'

expectation that this should not significantly affect the very large Universe. Nevertheless, the Chaplygin gas does seems to be special in this way (and UDE, by extension), in that its small scale structure *does* affect the equation of state of the average universe. Unfortunately, this all adds up to the fact that background and linear tests are simply not enough to validate or disprove the UDE hypothesis, making the complete analysis of these models unexpectedly complicated. Considerable work has yet to be done in order to establish their ultimate fate.

The layout of this work is as follows: In CHAPTER 1, we summarize the theoretical framework of the Standard Model of Cosmology and briefly discuss the 'usual suspects' for dark energy and dark matter. The concept of quartessence is then introduced as a compelling alternative. In CHAPTER 2, we investigate the homogeneous background proprieties of UDE and constrain the gCg parameters $(\mathcal{A}, \alpha)$ using supernovae Type Ia luminosity distances. We also discuss implementing quartessence as an isentropic scalar field obeying a particular Lagrangian. This, in turn, will lead us to the conclusion that the gCg is totally equivalent (to any order) to an ordinary ΛCDM model (as far as gravity is concerned) when $\alpha \to 0$, a fact that plays an important role. The question of what is meant by a *single* ('atomic') fluid is also carefully addressed. In CHAPTER 3, we set up all the necessary machinery to describe linear perturbations in order to study large scale structure formation in the context of UDE models. A formal demonstration of the equivalence to 1st order between ΛCDM and the $\alpha = 0$ generalized Chaplygin gas is given as an illustration of this framework. The crucial role baryons play in the formation of structure is then carefully highlighted. Using this machinery, a gCg model (plus baryons) is constrained against the mass power spectrum obtained from the 2dF 100k Galaxy Redshift survey. Linear instabilities are also briefly touched upon. In CHAPTER 4, we discuss the so-called averaging problem in the context of UDE and how even the large scale universe may be affected by the non-linear small scale clustering that occurs in the quartessence component. The argument is made, both qualitatively and quantitatively, that non-linear effects cannot be safely ignored (except when $\alpha = 0$). Thus, the majority of background and linear results obtained without taking into consideration the effect of non-linearities are put into serious question. The need

for a full order treatment is highlighted. A few simple ideas on how to improve this situation are subsequently offered. Finally, we end this thesis by summarizing, in bullet form, the main results obtained in the course of this work.

Contents

Acknowledgments i

Resumo iii

Résumé v

Abstract vii

Notation and Units ix

Preface xi

1 Introduction **1**

1.1 The Cosmological Principle 3

1.1.1 Maximal Symmetry 4

1.1.2 Constant Curvature 5

1.1.3 Background Kinematics 8

1.1.4 Cosmography vs. Cosmology 14

1.2 The Field Equations 15

1.2.1 The Canonical Lagrangian 16

1.2.2 Gravity 17

1.2.3 Energy-Momentum Tensor 18

1.2.4 Quintessence as a (Perfect) Fluid 20

1.2.5 The Energy Conditions 20

1.2.6 Isentropic Fluids 22
1.3 The Friedmann equations 23
1.3.1 The Flatness Problem 26
1.3.2 The Horizon Problem 26
1.4 Dark Energy in the Past 27
1.5 Dark Energy in the Present 29
1.6 Dark Matter Today 30
1.7 Quartessence 33
1.8 Cardassian Expansion 34

2 Background UDE Cosmology **37**
2.1 The gCg as a Scalar Field 39
2.2 The gCg $\alpha \to 0$ Limit 41
2.3 A Truly 'Atomic' Fluid? 42
2.4 UDE Background Tests 44
2.4.1 Background Model 45
2.4.2 Old vs. New Samples 46
2.4.3 Supernovae Statistics 47
2.5 Results and Comments 48

3 UDE — Linear Evolution **53**
3.1 Quartessence Sound Speed 54
3.2 Cosmological Perturbations 56
3.2.1 Gauge Ambiguities 59
3.2.2 Gauge Freedom 60
3.2.3 (Initially Unperturbed) Synchronous Gauge 60

3.2.4 Perturbation Types 62
3.2.5 Perturbed Stress-Energy Tensor 65
3.2.6 Perturbed Field Equations 67
3.2.7 ΛCDM linear equivalence to the $\alpha = 0$ gCg 68
3.3 Baryons + gCg — The Model 69
3.3.1 Statistics of Scalar Perturbations 72
3.3.2 Results 76
3.4 Conclusions 80

4 UDE — Non-Linear Regime **83**
4.1 All Quiet on the Western Front? 83
4.2 Qualitative Approach 87
4.2.1 Case I: the gGg with $\bar{c}_s^2 > 0$ 87
4.2.2 Case II: UDE with $\bar{c}_s^2 < 0$ 88
4.3 Quantitative Approach 89
4.3.1 Mass Dispersion 90
4.3.2 $\langle p \rangle$ in the Non-Linear Regime 91
4.4 Conclusions and Future Prospects 93
4.5 A Possible Way Out? 93

A Thesis X-Ray **97**

B Numerical Code **101**

Chapter 1

Introduction

"*... entities should not be multiplied beyond necessity.*"
William of Ockham, 14th Century

Observations have been steadily piling over the years that can only be 'explained', in the context of General Relativity, by the presence of so-called *dark* forms of energy; ordinary *baryonic* matter and radiation are simply not enough. In broad terms, a dark energy component (violating the strong energy condition) is required to explain the recent acceleration of the Universe, and a (cold) dark matter component to account for the amount of structure observed in various scales (plus a few other things). Unfortunately, these components have never been observed directly in any way (hence their 'dark' moniker). It is also very unlikely that a direct detection will occur in the near future. In fact, they may not *exist* at all; this, however, only seems possible if General Relativity is somehow flawed, in other words, if gravity is a much more complex interaction than Einstein previously thought (on large scales). It is certainly possible to engineer dark energy and dark matter 'out of the picture', by modifying General Relativity directly; often this process involves extra-dimensions with strange topologies and/or fiddling with TeVeS gravity sources. At the moment, however, most modifications are far too 'ad hoc' to be particularly pleasing, and nobody seems in any real hurry to give up on General Relativity just

yet (at least for large scales). In this thesis, we will always work within the confines of General Relativity, in other words, we'll be assuming that dark energy/dark matter are real entities. On the other hand, if they do exist, it is particularly vexing to find out that roughly 96% of the Universe should be in this dark form! That such a large chunk of the Universe has yet to be 'seen', is truly mind boggling. (We can certainly spin this around and take it as a sign that General Relativity is somehow flawed.) Thus, contemporary Cosmology is hard-pressed to explain what these components *are*. In the absence of a 'smoking gun', however, this is a very hard thing to do; there are simply to many ways to wrap phenomenological theories around circumstantial evidence. Also, our understanding of fundamental physics is not sufficiently advanced to safely guide us through this uncharted territory, let alone suggest an optimum route. In the meantime, we are reduced to somewhat 'arbitrary' discussions of what a canonic (i.e. best possible) model for dark energy/dark matter should be.

Certainly, in this regard, the realization that dark energy/dark matter do not have to be, a priori, *independent* entities, should play a significant role. Indeed, there is no compelling observational reason to suppose that they are. It is only because dark matter and dark energy play distinct roles in the background, that we frequently perceive them as being different. However, this doesn't actually force them to be different, it only *suggests* that. The Chaplygin gas, for instance, became famous precisely for being able to mimic both dark energy *and* dark matter (depending on the local density), and still be just *one* form of energy—a weird one, granted, but so can be said of quintessence, k-essence, etc. Thus, we should keep an open mind to the possibility that dark energy and dark matter are just different *manifestations* of a single underlying field and not different entities *per se*. Models built around this simplifying hypothesis are the main subject of this thesis. Obviously, if this is true or not, can only be definitely settled by some sort of 'direct' observation, an unlikely event by all accounts. Still, on the theoretical front, the idea of a unified description of dark energy/dark matter has incredible heuristic potential and should be thoroughly investigated. All things being equal, having to explain in a fundamental way just one exotic form of energy, is obviously much better than

having to explain two. In case of failure, we would at least come back justified in treating dark energy/dark matter as an independent pair. Either way, something of value could be gained.

In this chapter, we briefly cover the foundations of the Standard Model of Cosmology as well as review some of the 'usual suspects' for dark energy and dark matter. The (generalized) Chaplygin gas is also introduced (as a compelling alternative) and, in turn, the broader concept of *quartessence*. These ideas, however, are mostly presented from a phenomenological point of view; nowhere do we try to establish them from 'first principles', a task that is better left to our fellow theoretical physicists. Presently, there is no such thing as a 'fundamental motivation' behind quartessence; we can only hope that one may be found in the future. At best, there is an 'ad hoc' string theory interpretation of the Chaplygin gas as a d-brane tension, but this is hardly fundamental, anyway; it also does not seem to apply to more general forms of quartessence. It is much too early to say, with any degree of certainty, if unified models will ultimately stand on their own. In this work, we are merely interested in gauging how viable these models are from a phenomenological perspective, as opposed to implementing them in a fundamental way.

1.1 The Cosmological Principle

The idea that the Universe is pretty much 'the same everywhere', a stance known as the Cosmological Principle, is of central importance for contemporary Cosmology. However, 'looking the same everywhere' is an informal concept only made precise through the notion of a *maximally symmetric manifold*. A maximally symmetric manifold is a manifold with homogeneous and isotropic geometric properties. Here, homogeneity and isotropy mean that the metric is invariant under (suitable) translations and rotations, respectively. Note that isotropy and homogeneity are independent concepts; one does not imply the other. For instance, a homogeneous manifold can easily be anisotropic if the anisotropy is the same in every point. Nevertheless, isotropy about every point *does* imply homogeneity.

1.1.1 Maximal Symmetry

On a more technical level, a maximally symmetric space is characterized by having the maximum possible number of *Killing vectors*, i.e. $\frac{1}{2}n(n+1)$ where n is the dimension of the manifold. The integral curves of each Killing field $\mathcal{K}$ describe a one-parameter family of diffeomorphisms (1-1 smooth maps) $\varphi_\lambda : \mathcal{M} \to \mathcal{M}$ that *Lie drag* the metric tensor $g_{\mu\nu}$. This is just a very abstract way of saying that $g_{\mu\nu}$ is essentially 'the same' object from point to point along the integral curves (in the sense that the pushforward $\varphi_* \, g_{\mu\nu} = g_{\mu\nu}$). In other words, they describe *symmetries* of the metric tensor or isometries of the geometry. In fact, if we adapt a coordinate system to the integral curves of $\mathcal{K}$ (meaning that λ is the only non-constant coordinate along them), we will find that $g_{\mu\nu}(\lambda) = g_{\mu\nu}(\lambda + k)$ where k is a constant. But this is how we usually spot symmetries in the first place, by searching for coordinates where the components of the object field remain invariant under some coordinate translation $\lambda \to \lambda + k$. Thus, Killing vectors are just a way of describing metric symmetries in a coordinate independent way.

Now, consider that if by 'everywhere' above we really meant the *entire* spacetime, we would be basically describing a static Universe simply because, by construction, the time slices would have to look the same—this is sometimes called the 'Perfect' Cosmological Principle. However, a static configuration is incompatible with the Hubble flow of faraway galaxies and thus, 'everywhere' really means just the 'space part' of spacetime. Let $\overline{\mathcal{M}}$ be such a manifold obeying this 'scaled down' principle; we'll refer to it as the *background manifold* or the background universe. This manifold is naturally foliated into a (trivial) fiber-bundle $\mathbf{R} \times \overline{\Sigma}$ where $\mathbf{R}$ represents the threading and $\overline{\Sigma}$, the maximally symmetric 3-slice, with metric

$$d\bar{s}^2 = -\mathrm{d}t^2 + R^2(t) d\bar{\sigma}^2 \, . \tag{1.1}$$

Here t is the cosmic time, $R(t)$ is the scale factor and $d\bar{\sigma}^2$ is the 3-metric in the $\overline{\Sigma}$ slice, expressed as

$$d\bar{\sigma}^2 = \bar{\sigma}_{ij}(u) \, \mathrm{d}u^i \mathrm{d}u^j \, , \tag{1.2}$$

where $u \equiv u^k$ are whatever *comoving* coordinates you like. $\overline{\Sigma}$ is sometimes called the comoving slice. Note the absence of cross terms $\mathrm{d}t\,\mathrm{d}u^i$ implying that the threading is orthogonal to the slices. Also note that the t-threads are geodesics making any local comoving observer (those with constant u^k) an *inertial* one. Only a comoving observer will think that the background looks isotropic. In fact, the Earth is not a comoving observer, which is why we observe a dipole anisotropy in the cosmic background radiation in the first place as a result of a conventional Doppler effect. It is also interesting to note that these locally inertial comoving observers are able to carry synchronized clocks that keep synchronized forever in $\overline{\Sigma}$. This is a defining trait impossible to attain on many other curved manifolds. It is, of course, always possible to synchronize clocks in a small enough region of a general $\mathcal{M}$ through cumbersome signal sending techniques; essentially this just amounts to choosing local observers with zero relative velocities and a common time origin (that is, 'comoving' observers in a small region). Fortunately, in $\overline{\mathcal{M}}$ things are much easier (in homogeneous spaces, really). For one, we don't have to rely on signal sending to synchronize clocks. In fact, any comoving observer can set his own proper clock in tune to some cosmic field $\bar{\varphi}$; given that $\bar{\varphi}$ will evolve exactly the same way everywhere in $\overline{\Sigma}$, every proper clock can be made to tick at a common rate. They may still have a different time origin, though. However, comoving observers can all agree to start their clocks at a common cosmic event like the Big Bang or some other thing like $\bar{\varphi}$ reaching a certain value.

1.1.2 Constant Curvature

If a manifold is maximally symmetric then the *curvature* must be same in every point and in every direction. This requirement greatly reduces the number of possible maximally symmetric spaces by constraining the Riemann tensor. Let us show this by setting a local inertial base at a given point p. This base is not unique, of course. There are plenty others related to it by inertial transformations at p (that is, ordinary rotations or 'Lorentz rotations' according to whether the signature is Euclidian or Lorentzian, respectively). Maximal symmetry requires that whatever

the inertial base used, the Riemann tensor components should be the same; if they changed, the space wouldn't be isotropic. There are only a few inertially invariant tensors that can be used to build this Riemann tensor, like the metric tensor or the Kronecker delta. In fact,

$$\bar{R}_{\hat{\rho}\hat{\sigma}\,\hat{\mu}\hat{\nu}} \propto \bar{g}_{\hat{\rho}\hat{\mu}}\,\bar{g}_{\hat{\sigma}\hat{\nu}} - \bar{g}_{\hat{\rho}\hat{\nu}}\,\bar{g}_{\hat{\sigma}\hat{\mu}}, \tag{1.3}$$

turns out to be the only inertially invariant construction that displays the same set of index symmetries as the Riemann tensor does [30] (here the hat refers to an inertial base at a given point p). But because this is a tensor relation, it is also valid at p in any other coordinate system. Moreover, in a maximally symmetric space 'all points are created equal', as Sean Carroll so succinctly puts it, and thus it is valid everywhere else. Contracting both sides yields the constant of proportionality and we get

$$\bar{R}_{\rho\sigma\,\mu\nu} = \frac{\bar{R}}{n(n-1)}(\bar{g}_{\rho\mu}\,\bar{g}_{\sigma\nu} - \bar{g}_{\rho\nu}\,\bar{g}_{\sigma\mu})\,, \tag{1.4}$$

where $\bar{R}$ is the *constant* curvature Ricci scalar (not to be confused with the scale factor in (1.1)). This expression, of course, relates only to intrinsic curvature and so doesn't limit the global structure of the manifold; we won't bother with such fine details by always assuming a trivial global topology. The actual value of $\bar{R}$ is not very important; it just represents an overall scaling of the underlying space. Its sign, however, *is* and gives rise to the classification of *positive*, *zero* and *negative* curvature spaces. In the case of interest to us, $\overline{\Sigma}$ (an Euclidean 3-space) reduces, respectively, to a 3-sphere, flat ordinary $\mathbb{R}^3$ or a 3-hyperboloid and the Ricci tensor to

$${}^{(3)}\bar{R}_{ij} = 2k\,\bar{\sigma}_{ij} \tag{1.5}$$

where $k = {}^{(3)}\bar{R}/6$. Hence, by construction, $\overline{\Sigma}$ is automatically *spherically symmetric* meaning that it can be foliated by 2-spheres. We take advantage of this and write the metric of the comoving slice in the form

$$d\bar{\sigma}^2 = \bar{\sigma}_{ij}\,\mathrm{d}u^i\mathrm{d}u^j = e^{2\beta(\tilde{r})}\,\mathrm{d}\tilde{r}^2 + \tilde{r}^2 d\Omega^2\,, \tag{1.6}$$

where $\tilde{r}$ is some radial coordinate and $d\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ is the usual metric on the 2-sphere. (By the way, this is just the space part of the vacuum Schwarzschild solution.) The non-zero components of the Ricci tensor for such a metric turn out to be

$$\begin{aligned} {}^{(3)}\bar{R}_{11} &= \frac{2}{\tilde{r}}\,\partial_1\beta \qquad {}^{(3)}\bar{R}_{22} = e^{-2\beta}(\tilde{r}\partial_1\beta - 1) + 1 \\ {}^{(3)}\bar{R}_{33} &= \bar{R}_{22}\sin^2\theta\,, \end{aligned} \tag{1.7}$$

which we set equal to (1.5) and solve for β. We get

$$\beta(\tilde{r}) = -\frac{1}{2}\ln(1 - k\tilde{r}^2)\,, \tag{1.8}$$

which, in turn, lets us to write (1.1) as

$$d\bar{s}^2 = -\mathrm{d}t^2 + R^2(t)\left[\frac{\mathrm{d}\tilde{r}^2}{1 - k\tilde{r}^2} + \tilde{r}^2 d\Omega^2\right]\,, \tag{1.9}$$

called the Friedmann-Robertson-Walker (FRW) metric form. Note how k here sets the curvature and thus the 'size' of the spacial slices. Note also how the following substitutions $\tilde{r} \to \lambda\tilde{r}$, $k \to \lambda^{-2}k$, $R(t) \to \lambda^{-1}R(t)$ where λ is a constant, leave (1.9) invariant. The choice $\lambda = \sqrt{|k|}$ is quite popular since it normalizes the value of the curvature k to $\{-1, 0, +1\}$, but otherwise forces us to work with a dimensionless radial coordinate $\tilde{r}$ and a scale factor with dimensions of length. The benefits of working with a normalized k, however, are not that significant and we prefer instead to make $\lambda = R_0$ (where the index refers to the present time) and trade dimensions: the scale factor $a = R/R_0$ is now dimensionless and $r = R_0\,\tilde{r}$ acquires the dimensions of [L]:

$$d\bar{s}^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1 - \kappa r^2} + r^2 d\Omega^2\right]\,. \tag{1.10}$$

Naturally, the curvature $\kappa = k/R_0^2$ is no longer normalized. This is the form we'll be using. The Christoffel symbols [31, 32] for this metric are simple, albeit tedious,

to obtain. For later convenience we set them here:

$$\begin{aligned}
\bar{\Gamma}^0_{11} &= \frac{a\dot{a}}{1-\kappa r^2} & \bar{\Gamma}^1_{11} &= \frac{\kappa r}{1-\kappa r^2} \\
\bar{\Gamma}^0_{22} &= a\dot{a}r^2 & \bar{\Gamma}^0_{33} &= a\dot{a}r^2\sin^2\theta \\
\bar{\Gamma}^1_{22} &= -r(1-\kappa r^2) & \bar{\Gamma}^1_{33} &= -r(1-\kappa r^2)\sin^2\theta \\
\bar{\Gamma}^2_{12} &= \Gamma^3_{13} = \frac{1}{r} & \bar{\Gamma}^1_{01} &= \bar{\Gamma}^2_{02} = \bar{\Gamma}^3_{03} = \frac{\dot{a}}{a} \\
\bar{\Gamma}^2_{33} &= -\sin\theta\cos\theta & \bar{\Gamma}^3_{23} &= \cot\theta\,,
\end{aligned} \tag{1.11}$$

where $\dot{a} \equiv da/dt$ (all others are either zero or related to these by symmetry). It follows that the non-zero components of the Ricci tensor are

$$\begin{aligned}
\bar{R}_{00} &= -3\frac{\ddot{a}}{a} \\
\bar{R}_{11} &= \frac{a\ddot{a}+2\dot{a}^2+2\kappa}{1-\kappa r^2} \\
\bar{R}_{22} &= r^2(a\ddot{a}+2\dot{a}^2+2\kappa) \\
\bar{R}_{33} &= \bar{R}_{22}\sin^2\theta\,,
\end{aligned} \tag{1.12}$$

and the Ricci scalar is

$$\bar{R} = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2}\right]. \tag{1.13}$$

1.1.3 Background Kinematics

There are many kinematic effects that follow directly from the FRW metric (1.10). To see a few these, let us consider two comoving test particles (faraway galaxies if you like), one at the origin, the other at (r,θ,ϕ). Then, the 'instantaneous' distance between them is given by

$$d = a(t)\int_0^r \frac{dr'}{\sqrt{1-\kappa r'^2}} = a(t) f_\kappa(r)\,, \tag{1.14}$$

which increases in proportion to the scale factor. Here, depending on the 3-curvature of the slices we have that

$$\sqrt{|\kappa|}\, f_\kappa = \begin{cases} \sin^{-1}(\sqrt{|\kappa|}r) & \kappa > 0\,, \\ \sqrt{|\kappa|}r & \kappa = 0\,, \\ \sinh^{-1}(\sqrt{|\kappa|}r) & \kappa < 0\,. \end{cases} \tag{1.15}$$

(1.14) is often called the 'proper' distance between test particles, although strictly speaking it is not the result of a *proper* measurement. A measurement is called proper when it is made in a local rest-frame, using proper clocks and rulers. No single comoving observer is capable of measuring d directly. Instead, d is measured using an infinite array of proper rulers lined up in a slice of constant proper time (which is why the word proper gets stuck sometimes). This notion of distance leads to a 'recessional' velocity

$$v \equiv \dot{d} = Hd\,, \tag{1.16}$$

called the Hubble law, where $H = \dot{a}/a$ is the Hubble parameter. Note that v here is not a properly measured quantity either, so having a faster than light speed $v > c$ when $d > d_H = c/H$ (called the Hubble distance, Hubble length or Hubble scale) is no cause for alarm. v only has proper physical meaning as a relative velocity for objects that are 'infinitesimally' close, i.e., inside the same local flat patch. (Recall that in a curved manifold, there is no invariant way to compare tensors at different points; we can only compare them *locally* at a given point and its immediate flat vicinity. Hence, the concept of a relative velocity between distant points has no proper physical meaning, which is why a superluminal v is no big deal; while c is a local invariant, v has no local meaning. Incidentally, the 'size' of the inertial flat patches is determined by the 4-curvature radius of the manifold, which happens to be of the order of the Hubble length [33]. Thus, for $d \ll d_H$, we can pretend that galaxies are receding from each other with a relative velocity proportional to their distance, which is what Hubble originally discovered. Nevertheless, this is just a convenient way of seeing things; galaxies aren't really receding from each other, but rather it's the metric that is changing.) It follows that comoving test particles will 'accelerate' relative to each other by

$$\ddot{d} = (\dot{H} + H^2)d = -qH^2 d\,, \tag{1.17}$$

where $q = -a\ddot{a}/\dot{a}^2$ is called the 'deceleration' parameter. Again, this is not a proper quantity. Obviously, H and q are key cosmological quantities that need to be measured somehow. Unfortunately, there is no *direct* way of measuring instantaneous distances of this sort. In fact, distant objects like galaxies are mainly observable through the light they emit, which naturally takes a fine amount of time to reach us. We cannot, therefore, perform measurements along hypersurfaces of constant time, but only along null paths traveling from the past toward us, i.e, the past light cone. Below, we'll discuss some alternative notions of distance that can, in principle, be measured directly and how they relate to H and q.

Let us now inquire about the geodesic (inertial) motion of free particles in $\overline{\mathcal{M}}$. We'll start by introducing the following Killing *tensor* [34]

$$\bar{k}_{\mu\nu} = a^2(\bar{g}_{\mu\nu} + \bar{U}_\mu \bar{U}_\nu)\,, \tag{1.18}$$

where $\bar{U}^\mu = (1,0,0,0)$ is the 4-velocity of comoving observers. We remind the reader that a Killing tensor is simply a symmetric covariant tensor obeying $\nabla_{(\mu} k_{\nu_1\nu_2\ldots\nu_l)} = 0$, where the parenthesis denote symmetrization. You may recognize this as a generalized Killing vector $\nabla_{(\mu} k_{\nu)} = 0$. The point is that the quantity $k_{\nu_1\nu_2\ldots\nu_l}\, p^{\nu_1} p^{\nu_2} \cdots p^{\nu_l}$, where $p^\mu \equiv dx^\mu/d\lambda$ is the 4-momentum of the particle and λ the affine parameter, remains constant along geodesics, in other words

$$p^\mu \nabla_\mu (k_{\nu_1\nu_2\ldots\nu_l}\, p^{\nu_1} p^{\nu_2} \cdots p^{\nu_l}) = 0\,. \tag{1.19}$$

This is not too difficult to prove if one remembers that $p^\mu \nabla_\mu\, p^\nu = 0$ for geodesics and that $p^{\nu_1} p^{\nu_2} \cdots p^{\nu_l}$ is a symmetric tensor. It's also not very difficult to confirm that (1.18) is indeed a Killing tensor. It follows that $\bar{k}_{\mu\nu}\, \bar{p}^\mu \bar{p}^\nu = a^2\,[\,\bar{p}_\mu\, \bar{p}^\mu + (\bar{p}^\mu\, \bar{U}_\mu)^2]$ will be a constant along geodesics. Since for massive particles we have that $\bar{p}_\mu\, \bar{p}^\mu = -m^2$ or $(\bar{p}^0)^2 = m^2 + |\bar{\mathbf{p}}|^2$ where $|\bar{\mathbf{p}}|^2 = \bar{g}_{ij}\, \bar{p}^i \bar{p}^j$ is the ordinary 3-momentum and $\bar{p}^\mu\, \bar{U}_\mu = -\bar{p}^0$ we conclude that

$$|\bar{\mathbf{p}}| \;\propto \frac{1}{a}\,, \tag{1.20}$$

in other words, massive free particles slow down with respect to the comoving grid as the Universe expands (their peculiar velocities tend to zero). A similar thing

happens to photons; they don't slow down, of course, but they do lose some of their energy with the expansion. This is simple to obtain; since now $\bar{p}_\mu \bar{p}^\mu = 0$, it follows that $-\bar{p}^\mu \bar{U}_\mu = \bar{p}^0 \propto a^{-1}$, which is just the photon's energy (or frequency ω if $\hbar = 1$) as measured by a comoving observer. This energy loss is related to the fact that in $\overline{\mathcal{M}}$ there is no timelike Killing vector, thus no notion of a conserved energy [34]. Consequently, a photon emitted at an earlier time with frequency ω will be observed with a lower frequency $\omega_0 = (a/a_0)\,\omega$ at a later time; equivalently, the fractional change $z \equiv \Delta\lambda/\lambda$ in the proper wavelength, called the *redshift*, will be $1 + z = a_0/a$. From an observational point of view, the redshift of an object is extremely useful because it tells us when photons were emitted and how faraway their source was at the time. It thus acts as a measure of time and distance. Of course, z is only an *observable* quantity for events that take place after recombination (when photons decouple and become free), but in principle we can still use it to tag earlier events; we just won't be able to measure it directly, just as we can't measure the proper distance to faraway objects.

We're now in a better position to discuss a few alternative notions of distance that can, in principle, be measured directly. First, we introduce the so-called luminosity distance

$$d_L^2 = \frac{\mathcal{L}}{4\pi\mathcal{F}} \tag{1.21}$$

where $\mathcal{L}$ is the absolute luminosity of the source and $\mathcal{F}$ is the flux measured by the observer (i.e., the energy per unit time per unit area). In Euclidean space, (1.21) is just the familiar inverse square law: the luminosity spreads itself equally across every spherical surface that is concentric with the source. In a FRW universe, this is still true, but now we have to take into account that photons do not only redshift by a factor of $1 + z$ but also hit each sphere less frequently (due to the expansion) by another $1 + z$ factor. Thus, we conclude that in $\overline{\mathcal{M}}$

$$\begin{aligned} d_L &= (1+z)a_0 r = (1+z)a_0\, f_\kappa^{-1}\Big(a_0^{-1}\int_0^z \frac{dz'}{H(z')}\Big)\,, \\ &= H_0^{-1}\Big[z + \frac{1}{2}(1-q_0)z^2 + \cdots\Big]\,, \end{aligned} \tag{1.22}$$

where we have used the fact that the area of the surface touching the detector is $4\pi a_0^2 r^2$, and also that for null paths $f_\kappa(r) = \int dt/a$. The third equality comes from Taylor expanding the scale factor about the present day [35] and is valid for redshifts $z \lesssim 0.3$.

Next, we define the angular distance

$$d_A = \frac{D}{\theta}\,, \tag{1.23}$$

where D is the proper diameter of the object and θ its angular size. Just as with the luminosity distance above, the idea is to construct a notion of distance that displays the same familiar geometric properties of Euclidean space, at least in a 'small' flat vicinity around the observer, in this case, the usual variation of the angular size with distance. From the FRW metric (1.10), we immediately conclude that

$$\begin{aligned} d_A &= ar = d_L(1+z)^{-2}\,, \\ &= H_0^{-1}\Big[z - \frac{1}{2}(3+q_0)z^2 + \cdots\Big]\,. \end{aligned} \tag{1.24}$$

Now, using (1.22) and (1.24) to infer H_0 and q_0 hinges on the existence of so-called standard candles and rulers (in other words, objects with known luminosities and sizes) and our ability to use them. Regarding standard rulers, the lack of reliable objects with known proper sizes is notorious. In recent years, however, a champion ruler has emerged from the cosmic microwave background (CMB). The temperature autocorrelation function [36] measures how the CMB temperature in two different directions of the sky fluctuates; naturally, this variation depends on the angular separation and the power spectrum of this autocorrelation is observed to have a series of 'acoustic' peaks. It turns out, that the first acoustic peak is roughly determined by the *sound horizon* at recombination, i.e., the maximum distance a sound wave in the baryon-radiation fluid could have travel until recombination. This sound horizon is given by $l_s \simeq d_H(z_r \simeq 1100)$ and serves as a standard ruler. The remarkable thing about this ruler is that its angular size almost only depends on the curvature of the $\overline{\Sigma}$ spatial slices. Hence, measuring the angular size of the first

acoustic peak has emerged as the leading and most direct method of determining the spatial curvature of the Universe. Recent data coming from the WMAP satellite experiment have shown that these slices are virtually flat [5]. On the note of standard candles now, Cepheid variables have been used for nearly a century. However, they are far too faint to be of any use for $z \gtrsim 0.1$. In the last decade or so, Type Ia supernovae have taken their place as a result of their extreme brightness; indeed, they have been observed up to a record $z = 1.7$ [37]. They appear to be good candles in so far as their luminosity profiles look relatively the same for all supernovae of this type (the cause for this uniformity, however, is not completely understood). They also seem to occur randomly in all types of galaxies. Measurements using this type of candle have produced one of the most spectacular results in the history of Cosmology [1, 4]: the Universe seems to be accelerating ($q_0 < 0$). This discovery took almost everyone by surprise. What is causing this acceleration is one of the biggest mysteries we have today. Considering that in a matter-dominated Universe, the gravitational self-attraction of matter naturally slows down the expansion, such a recent acceleration implies that a substantial amount of dark energy must have begun dominating over matter close to today. If this is a coincidence or not, is still an open question and a hotly debated one at that.

Despite all of this, however, 'distances' to faraway galaxies are still not known with the precision necessary for an accurate measurement of H_0; currently, the Hubble constant is believed to be between 65 to 80 km s^{-1}Mpc^{-1}. This uncertainty is usually parameterized by writing

$$H_0 = 100h \text{ km } s^{-1}\text{Mpc}^{-1} \simeq \frac{h}{3000} \text{ Mpc}^{-1} , \tag{1.25}$$

where the second equality uses units where $c = 1$; the most recent calculations using data from WMAP yielded $h \simeq 0.71$; for a review see Jackson's [38]. Hence, the Hubble length today is

$$d_H(t_0) = 2998h^{-1} \text{ Mpc} , \tag{1.26}$$

which roughly determines the 'size' of our local flat patch. Incidentally, the Hubble length is also frequently called the Hubble 'horizon' or the Hubble radius. This poor

terminology stems from the fact that the 'particle horizon' is normally of comparable size to d_H. However, they are intrinsically different things; while the Hubble length is a dynamical scale characterizing the rate of expansion, the former is determined by kinematic considerations alone. Recall that the particle horizon is defined as the maximum distance free photons can travel in a given amount of time, normally starting at the Big Bang and ending today (you can nitpick and replace Big Bang by recombination, if you wish); it therefore determines the size of the observable Universe. It also represents the typical size of causally connected regions. Thus, a priori, there is no reason why the particle horizon should be of comparable size to that of a local flat patch. In fact, if we drop the strong energy condition, the particle horizon can grow much bigger than d_H [33]; the observable Universe doesn't have to neatly fit inside the Hubble scale all the time.

1.1.4 Cosmography vs. Cosmology

This is as far as the Cosmological Principle will take us. It determines the 'kinematics' of the background manifold $\overline{\mathcal{M}}$ (a cosmography), not its dynamics (a cosmology). To go further we need a 'theory of manifold dynamics'. General Relativity (based on the Equivalence Principle) is such a theory and throughout this thesis we work consistently within this framework. In the following sections, we'll explore the consequences of plugging the Cosmological Principle into General Relativity. We stress, however, that the Cosmological Principle is in itself a distinct hypothesis from any dynamical theory and it is to our advantage to keep this present.

We end this section with two cautionary observations: First is the usual observation that the Cosmological Principle is meant to apply on large enough scales (over 100 Mpc) where obvious local inhomogeneities are averaged out much akin to how a gas is approximated by a fluid. The principle is firmly anchored in a variety of observations, the most important being the incredible isotropy of the CMB radiation. However, it is not set in stone. For instance, we cannot guarantee if it applies everywhere outside the observable Universe; actually, in the context of inflation, we

don't expect it to. Second, it is common to say that $\overline{\mathcal{M}}$ is curved or flat according to whether the spacial slices are curved or flat. Obviously, this is a language abuse given that a flat $\overline{\Sigma}$ will generally have a non-zero 4-curvature (by being embedded in a higher dimensional curved space). A famous 'reverse' example is the Milne Universe (an empty space with $k = -1$) where the 4-curvature is actually zero, but the spacial slices are hyperbolic and therefore curved.

1.2 The Field Equations

'Nature loves variational principles'. The actual solutions of *any* physical system (at least, at the microscopic, non-dissipative, level) always seems to extremize a given action functional. For example, most classical field theory solutions are critical 'points' of the action

$$S = \int d^4x \, \mathcal{L}(\Phi^i, \nabla_\mu \Phi^i) \,, \tag{1.27}$$

where $\mathcal{L}$ is the Lagrangian characteristic of the theory, $\{\Phi^i\}$ is the set of dynamical variables (here i labels fields, not components, and Φ^i can be *any* tensor field in $\mathcal{M}$, not just a scalar) and $d^4x = \mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3$ is the volume element. Note carefully that although d^4x looks like a 4-form, it actually transforms as a density (of weight 1), not as a scalar [39]. This implies that $\mathcal{L}$ is a scalar *density*, not a scalar field. Even so, we typically write

$$\mathcal{L} = \sqrt{-g}\widehat{\mathcal{L}} \,, \tag{1.28}$$

where $\widehat{\mathcal{L}}$ is now an 'honest' scalar field given that $\sqrt{-g}\, d^4x$ transforms as one. As usual, if Φ^i is an extremum of S, then under a small variation of the type

$$\begin{aligned} \Phi^i &\rightarrow \Phi^i + \delta\Phi^i \,, \\ \nabla_\mu \Phi^i &\rightarrow \nabla_\mu \Phi^i + \delta(\nabla_\mu \Phi^i) = \nabla_\mu \Phi^i + \nabla_\mu(\delta\Phi^i) \,, \end{aligned} \tag{1.29}$$

the action doesn't vary to first order, i.e., $\delta S = 0$. This fact translates into an Euler-Lagrange equation of motion constraining the extrema fields

$$\frac{\partial \widehat{\mathcal{L}}}{\partial \Phi^i} - \nabla_\mu \frac{\partial \widehat{\mathcal{L}}}{\partial(\nabla_\mu \Phi^i)} = 0\,, \tag{1.30}$$

that ultimately get picked by Nature.

1.2.1 The Canonical Lagrangian

As a simple example of this formalism, consider the case of a classical scalar field ϕ governed by the so-called *canonical* Lagrangian

$$\widehat{\mathcal{L}} = -\frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi - V(\phi)\,, \tag{1.31}$$

where V is some scalar potential. (Note that since $\nabla_\mu \phi = \partial_\mu \phi$, we could have used partial derivatives instead; however, this is generally regarded as a 'bad practice'.) (1.31), of course, doesn't describe *every* imaginable scalar field, but it is, nonetheless, a good starting point. In a cosmological context, ϕ is commonly called a 'Quintessence' field or just 'Quintessence'. Applying (1.30) to this scalar, we find that it obeys

$$\Box\phi - \frac{dV}{d\phi} = 0\,, \tag{1.32}$$

where $\Box = \nabla^\mu \nabla_\mu = g^{\mu\nu} \nabla_\mu \nabla_\nu$ is the covariant d'Alembertian. For later convenience, we introduce here the quantity

$$X = -\frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi = -\frac{1}{2} \nabla^\mu \phi \nabla_\mu \phi\,, \tag{1.33}$$

which is commonly refered to as the *kinetic energy* of ϕ (regardless of ϕ being a canonic field or not). Note that in a locally inertial frame, (1.33) reduces to $\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}(\nabla\phi)^2$; one can therefore interpret this as a plausible generalization of a point particle kinetic energy, thus justifying the name given to X. Note also that X at a point $p \in \mathcal{M}$ is a function of $\nabla^\mu \phi$ (and $g_{\mu\nu}$, of course), but *not* ϕ. You might find this perplexing at first, but given that $\nabla^\mu \phi$ is a *vector*, it is defined at p independently of the actual value $\phi(p)$. This is very similar to how in Hamiltonian mechanics, p and x are treated as independent variables.

1.2.2 Gravity

General Relativity is another example of a classical field theory, albeit a much more complex one. In this case, the dynamical variable is the metric tensor $g_{\mu\nu}$ and the Lagrangian is given by

$$\widehat{\mathcal{L}} = \frac{1}{16\pi G}\widehat{\mathcal{L}}_H + \widehat{\mathcal{L}}_M \tag{1.34}$$

where $\widehat{\mathcal{L}}_H = R$ is due to Hilbert. The second term is loosely called the 'matter' term and we'll discuss it shortly. (Note that due to requirement of metric compatibility $\nabla_\sigma g_{\mu\nu} = 0$ [32], this Lagrangian cannot be written in terms of covariant derivatives of $g_{\mu\nu}$.) The field equations can be obtained straightforwardly by varying the action directly. It turns out that the inverse metric $g^{\mu\nu}$ is better suited for this purpose, not $g_{\mu\nu}$. Thus, making use of $R = g^{\mu\nu}R_{\mu\nu}$ and treating δ as a 'derivative operator', we get

$$\delta S_H = \int d^4x \, \left(R\delta\sqrt{-g} + \sqrt{-g}R_{\mu\nu}\delta g^{\mu\nu} + \sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} \right) , \tag{1.35}$$

from which we ultimately aim to factor out $\delta g^{\mu\nu}$. The middle steps to achieve this are not very interesting and we spare the reader some lengthy details by just quoting the results (see, for instance, [34]): The third term in the integrand ends up not contributing at all, while the first term expands nicely to

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu} , \tag{1.36}$$

yielding

$$\delta S_H = \int d^4x\sqrt{-g}\left[R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \right]\delta g^{\mu\nu} . \tag{1.37}$$

The tensor inside the square brackets is called the Einstein tensor and is usually denoted by $G_{\mu\nu}$; it is a purely geometric entity related to the curvature of the manifold $\mathcal{M}$. Einstein was delighted by the fact that $\nabla_\mu G^{\mu\nu} = 0$, regardless of the actual manifold, an identity called Bianchi's identity. This identity ends up playing a crucial role in energy 'conservation'. It follows that the critical points of General Relativity are given by

$$\frac{1}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} = \frac{1}{16\pi G}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \right) + \frac{1}{\sqrt{-g}}\frac{\delta S_M}{\delta g^{\mu\nu}} = 0 , \tag{1.38}$$

where

$$T_{\mu\nu} = -2\frac{1}{\sqrt{-g}}\frac{\delta S_M}{\delta g^{\mu\nu}}\,, \tag{1.39}$$

is called the *energy-momentum* tensor of the 'matter field'. Collecting everything, the field equations can thus be neatly arranged as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}\,. \tag{1.40}$$

1.2.3 Energy-Momentum Tensor

We stress that the 'energy-momentum' name for (1.39) is warranted; by definition, it is automatically a symmetric, gauge invariant, *conserved* tensor (courtesy of Bianchi's identity) with the dimensions of an energy density. Let's apply it to the case of the canonical scalar field (1.31). Now, however, we vary the action in relation to the inverse metric, *not* ϕ. We end up with

$$\begin{aligned}\delta S_\phi &= \int d^4x \left[\sqrt{-g}\left(-\frac{1}{2}\delta g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi\right) + \delta\sqrt{-g}\,(X - V(\phi))\right]\,, \\ &= \int d^4x\,\sqrt{-g}\delta g^{\mu\nu}\left[-\frac{1}{2}\nabla_\mu\phi\nabla_\nu\phi - \frac{1}{2}g_{\mu\nu}\,(X - V(\phi))\right]\end{aligned} \tag{1.41}$$

and, therefore

$$T_{\mu\nu}(\phi) = \nabla_\mu\phi\nabla_\nu\phi + (X - V(\phi))g_{\mu\nu}\,. \tag{1.42}$$

Recall that in any locally inertial frame, T^{00} represents the energy density, T^{0i} the energy flux density (which equals the momentum density T^{i0}) and T^{ij} the spatial stress (see, for instance, the 'bible' [40]). Thus, the momentum density associated with ϕ reduces to

$$T^{0i} = -\dot{\phi}\,\partial_i\phi\,. \tag{1.43}$$

Also recall that *rest*-frames are an important class of inertial frames characterized by the fact that they measure zero momentum density (it goes without saying that

these are almost always 'local' frames in the sense that measurements made in these frames can only be attached proper physical meaning in small enough regions). The absence of momentum can be interpreted straightforwardly as saying that the rest-frame comoves, at least momentarily, with the 'center-of-mass' of a local region of the field. (These frames almost always exist in practice, although there are a few situations where they don't. A plane electromagnetic wave, for instance, cannot have a comoving rest-frame, otherwise momentum-less photons would exist. This is not a problem for background radiation, though. In $\overline{\mathcal{M}}$, we'll be comoving with the background slice, not the photons themselves, in such a way that their momentum density is null—this just means that they are equally moving in all directions 'canceling' each others momenta.) Because we are interested in a time evolving field, the rest-frame for quintessence is thus characterized by

$$\partial_i \phi = 0\,. \tag{1.44}$$

We emphasize that this is a *local* equation and therefore does *not* imply a homogeneous field; this assertion is only true in the background universe where the FRW rest-frame extends globally with proper meaning. It follows that the rest-energy density T^{00} and pressure T_{ii} for quintessence are given by

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi)\,, \qquad P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,, \tag{1.45}$$

and zero stress T_{ij}. Note carefully, however, that these quantities are not true scalars; even though the potential $V(\phi)$ is a scalar, the term $\frac{1}{2}\dot{\phi}^2$ is not. Nevertheless, we can still define 'proper' quantities

$$\varepsilon_\phi = X + V(\phi)\,, \qquad p_\phi = X - V(\phi)\,, \tag{1.46}$$

that reduce to ρ_ϕ and P_ϕ in the rest-frame and *are* manifestly scalars. Because these are *properly* defined invariant quantities, they are know rather unimaginatively as the 'proper' energy density and pressure of the scalar field. Finally, we would like to draw the reader's attention to the curious fact that the canonical Lagrangian (1.31) equals the proper pressure of the quintessence field. Later, we'll generalize this to a broader class of interest.

1.2.4 Quintessence as a (Perfect) Fluid

In the rest-frame, quintessence appears as a stressless isotropic form of energy; this sounds very familiar to a *perfect fluid*. In fact, they are very much the *same*. We can show this explicitly by making the following identifications

$$u_\mu = \frac{\nabla_\mu \phi}{\sqrt{2X}}\,, \quad \varepsilon = \varepsilon_\phi\,, \quad p = p_\phi\,. \tag{1.47}$$

and subsequently plugging them into (1.42). We obtain

$$T^{\mu\nu} = (\varepsilon + p)u^\mu u^\nu + p g^{\mu\nu}\,, \tag{1.48}$$

i.e., the energy-momentum tensor of ϕ written in the form of a perfect fluid. Recall that in (1.48), u^μ is the 4-velocity field describing the motion of the fluid and that ε and p, are its proper energy density and pressure, respectively. Given that, in general, the fluid elements all move in relation to each other, a myriad of locally inertial observers, each comoving with a fluid element, are necessary to measure p and ε globally. Note also that the word 'perfect' here is used as a synonym for isotropy in the inertial rest-frame; nothing else is implied by it, namely a particular relation between pressure and density.

1.2.5 The Energy Conditions

We stress, however, that this hydrodynamical analogy between quintessence and a perfect fluid is only possible if (1.47) really *is* a 4-velocity by which we mean a normalized timelike vector $u^\mu u_\mu = -1$. This implies that $\nabla_\mu \phi$ has to be timelike, forcing the kinetic term $X > 0$; one can interpret this as a way of guaranteeing that fluid elements do not flow outside the light-cone. In practice, this is ensured by means of an energy condition. Table (1.1) summarizes the energy conditions that are commonly used to define 'reasonable' fields. They are stated under the assumption that the stress-energy tensor of the field can be diagonalized, i.e., written in the form [41]

$$T^{\mu\nu} = \varepsilon\, \hat{e}^\mu_{(0)} \otimes \hat{e}^\nu_{(0)} + \sum_i p_i\, \hat{e}^\mu_{(i)} \otimes \hat{e}^\nu_{(i)} \tag{1.49}$$

Table 1.1: Energy Conditions.

Name	Statement	Conditions
Weak	$T_{\mu\nu}\, t^\mu t^\nu \geqslant 0$	$\varepsilon \geqslant 0, \quad \varepsilon + p_i > 0$
Null	$T_{\mu\nu}\, k^\mu k^\nu \geqslant 0$	$\varepsilon + p_i \geqslant 0$
Strong	$\left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu}\right) t^\mu t^\nu \geqslant 0$	$\varepsilon + \sum_i p_i \geqslant 0, \quad \varepsilon + p_i \geqslant 0$
Dominant	$-T^\nu_\mu\, t^\mu$ future directed	$\varepsilon \geqslant \lvert p_i \rvert$

where ε and p_i are the energy density and pressure eigenvalues and $\hat{e}_{(\mu)}$, the eigenvector tetrad (here parenthesis distinguish base indices from component indices). One might think that because $T^{\mu\nu}$ is a real symmetric matrix, such eigentetrads would always exist. Unfortunately, this isn't true; given that the metric $g_{\mu\nu}$ is not positive definite, the linear map $T^\mu_{\ \nu} : V \to V$ need not be diagonal (see [42] for a detailed explanation). An example is the null fluid (the above mentioned plane wave) which only has 3 independent eigenvectors (one null and two spacelike), not 4. Nevertheless, it is generally believed that 'reasonable' fields do have 4-dimensional eigenbases.

We remind the reader that in Table (1.1), $T_{\mu\nu}\, t^\mu t^\nu$ and $-T^\mu_{\ \nu}\, t^\nu$ are, respectively, the energy and momentum density a local observer with 4-velocity t^μ will measure (if you forget why, just write them in an inertial frame comoving with the observer). Also there, k^μ is a null vector. Again, see [41, 42] for a more detailed explanation. Here, we merely gloss over the physical meaning of the energy conditions. For instance, one can interpret the dominant energy condition as the statement that the speed of energy flow is always less than the speed of light. In the case of quintessence, this forces $X > 0$ and $V > 0$, as one can easily check, making (1.47) timelike and, by extension, the hydrodynamical analogy possible. In fact, any field obeying the dominant condition can be 'treated as a fluid' (though not necessarily a perfect one). Fluids are generally easier to work with than fields, so we will sometimes say that fluids are a 'high-level' description of 'matter', while fields are 'low-level'. The weak energy condition, on the other hand, only states that the proper energy of the field

should be positive. Note therefore, that because it implies a positive proper energy, the weak energy condition is automatically included by the dominant one. However, the former, by itself, only implies $X > 0$ and $V > -X$ for quintessence. Also, it turns out that it is the strong energy condition that is responsible for making gravity 'attractive' by focusing congruences.

Although ordinary forms of matter do obey some of these energy conditions, they are hardly set in stone. It turns out that quantum fields can generally violate any one of them. For instance, the famous 'Casimir effect' violates the weak energy condition by making the energy density in the region between two closely held conducting plates negative. Also, during *inflation* the strong energy condition is broken. Incidentally, 'dark energy' must equally break the strong energy condition, if it is to accelerate the expansion of the Universe today. Still, in a cosmological context, it is standard to at least assume the dominant condition. This is expressed by requiring

$$|\omega| \leqslant 1\,, \tag{1.50}$$

where $p = \omega\varepsilon$, but even this is a conservative starting point. Phantom energy [11, 12], for example, is a speculative form of energy where $\omega < -1$. Additionally, note that the strong energy condition is broken when $\omega < -1/3$.

1.2.6 Isentropic Fluids

'Matter' in the background universe is routinely described as a 'high-level' fluid. This fluid has to be perfect so as to share the same isotropic proprieties of $\overline{\mathcal{M}}$. It also has to comove with the background slice; this means that in the FRW frame, its 4-velocity is $\bar{u}^\mu = (1, 0, 0, 0)$ everywhere, which greatly simplifies (1.42). It must also obey energy-momentum conservation $\bar{\nabla}_\mu \bar{T}^{\mu\nu} = 0$; thus, grabbing the necessary symbols from (1.11), the time part leads to

$$\begin{aligned} \bar{\nabla}_\mu \bar{T}^{\mu 0} &= \partial_\mu \bar{T}^{\mu 0} + \bar{\Gamma}^\mu_{\mu\sigma}\, \bar{T}^{\sigma 0} + \bar{\Gamma}^0_{\mu\sigma}\, \bar{T}^{\mu\sigma} \\ &= \dot{\bar{\varepsilon}} + 3\frac{\dot{a}}{a}(\bar{\varepsilon} + \bar{p}) = 0\,, \end{aligned} \tag{1.51}$$

usually rewritten as

$$\frac{\dot{\bar{\varepsilon}}}{\bar{\varepsilon}} = -3(1+\bar{\omega})H\,, \tag{1.52}$$

while the space part leads to $\partial_i \bar{p} = 0$, a null pressure gradient, as it should given the homogeneous and isotropic nature of $\overline{\mathcal{M}}$. To go any further than this, we need to know exactly what the equation of state parameter $\bar{\omega}$ is, in other words, its functional dependence. There are many possibilities for this: one useful class is defined by $\bar{\omega} \equiv \bar{\omega}(\varepsilon)$ strictly as a function of the proper energy density. Mukhanov calls these fluids *isentropic* [33]; the simplest examples are actually constant-$\bar{\omega}$ fluids for which (1.52) reduces to

$$\bar{\varepsilon} = \bar{\varepsilon}_0 \Big(\frac{a_0}{a}\Big)^{3(1+\bar{\omega})}\,. \tag{1.53}$$

This covers vacuum energy, dust (pressureless non-relativistic matter) and radiation when $\bar{\omega} = -1, 0, 1/3$ respectively. We still lack the evolution of a though, so we don't know everything yet.

1.3 The Friedmann equations

Let us now apply Einstein's equation to the background universe $\overline{\mathcal{M}}$. We begin by rewriting (1.40) in a slightly more convenient form

$$R_{\mu\nu} = 8\pi G\Big(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu}\, T\Big)\,, \tag{1.54}$$

where $T = T^{\mu}_{\ \mu}$ is the trace of the energy-momentum tensor. In the FRW frame, the background source behaves as a perfect fluid so we have that $\bar{T}^{\mu}_{\ \nu} = (-\bar{\varepsilon}, \bar{p}, \bar{p}, \bar{p})$ implying $\bar{T} = -\bar{\varepsilon} + 3\bar{p}$. Harkening back for the Ricci components in §(1.1.2), we get from the $\mu\nu = 00$ part of (1.54) that

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\bar{\varepsilon} + 3\bar{p})\,, \tag{1.55}$$

and from $\mu\nu = ii$ part

$$\frac{\ddot{a}}{a} + 2\Big(\frac{\dot{a}}{a}\Big)^2 + 2\frac{\kappa}{a^2} = 4\pi G(\bar{\varepsilon} - \bar{p})\,. \tag{1.56}$$

All other $\mu\nu$ lead to $0 = 0$ identities. (1.55) and (1.56) are called the *Friedmann equations* and they govern the evolution of the scale factor and, by extension, the dynamics of the background. Alas, they do not form a closed system; hence, we cannot solve them completely unless we bring some additional information. The assumption that the background source is isentropic normally takes care of this; the state equation $\bar{p} = \bar{p}(\bar{\varepsilon})$ then closes the system. If, one the other hand, the fluid is not isentropic, the equation governing the field has to be included, and also its relation between $\bar{p}$ and $\bar{\varepsilon}$. Suffice it to say that the isentropic case is generally much simpler and we will assume for now. We can use (1.55) to remove the second derivative in (1.56) and do a bit of cleanup to obtain

$$H^2 = \frac{8\pi G}{3}\bar{\varepsilon} - \frac{\kappa}{a^2}\,. \tag{1.57}$$

This is what most people call *the* Friedmann equation, by the way. It follows that if the $\overline{\Sigma}$ spatial slices are flat, then $\bar{\varepsilon} = \bar{\varepsilon}_c \equiv 3H^2/8\pi G$ which is called the *critical* density. It is usually simplest to measure energy densities as a fraction of this quantity, in other words, $\Omega = \bar{\varepsilon}/\bar{\varepsilon}_c$, or the density parameter. Thus, we rewrite (1.57) in the form

$$\Omega - 1 = \frac{\kappa}{a^2H^2}\,. \tag{1.58}$$

From this we immediately see that the curvature sign is determined by having the (total) energy density above ($\kappa > 0$) or below ($\kappa < 0$) the critical density. Recent data from the WMAP experiment place us extraordinarily close to the flat case with an $\Omega = 1.02 \pm 0.02$ [5]. That we are so close to being flat actually poses a curious problem called the *flatness* problem, discuss shortly in the following section.

Before we do so, however, it should be made clear that the energy density $\bar{\varepsilon}$ and pressure $\bar{p}$ above are really the sum of several components, not just one, as it might hastily seem. To make this explicit, we recast once again the Friedmann equation in the form

$$H^2 = H_0^2\left[\sum_i \Omega_{i0}\frac{\bar{\varepsilon}_i}{\bar{\varepsilon}_{i0}} + (1-\Omega_0)\left(\frac{a_0}{a}\right)^2\right]\,, \tag{1.59}$$

where now $\Omega = \sum_i \Omega_i$ and the subscript '0' denotes the present time. (Incidentally, we see from this that we can very well pretend that the curvature term $\Omega_\kappa = 1 - \Omega_0$

represents a fictitious component having $\bar{\omega} = -1/3$.) In doing this, we are implicitly assuming that the several components only interact gravitationally (i.e. they are minimally coupled).

Now, let us consider a class of models where all the components are constant-$\bar{\omega}$ isentropic fluids $\bar{p}_i = \bar{\omega}_i \bar{\varepsilon}_i$. (If you find it too narrow, consider the fact that our chances of distinguishing between several time-varying scalar fields are rather bleak at the moment [43, 44]. Hence, we might as well take the pragmatic stance of looking at these isentropic components as the *effective* versions of the underlying dynamic fields, if any.) The dynamics of this class can be roughly understood if we interpret the Friedmann equation as an energy integral of motion of a one-dimensional fictitious particle moving with an a coordinate. Consider (1.59) rewritten in the form

$$\dot{a}^2 - \sum_i \Omega_i a^{-(1+3\omega_i)} = \Omega_\kappa \,, \tag{1.60}$$

where we've used (1.53) and momentarily switched units so that $H_0 = a_0 = 1$. Comparing this to the standard energy equation of the fictitious particle, $E = K + V$, we see that the curvature term plays the role of the mechanical energy, $\dot{a}^2$, the kinetic energy and the remaining term, the potential felt by the particle. We can even push forward this Newtonian analogy by calculating the force $-dV/da$ felt by the particle and see the Raychaudhuri equation emerge. Each fluid then contributes with a partial force of

$$-dV_i/da = -(1 + 3\bar{\omega}_i)\Omega_i a^{-(2+3\omega_i)} \,, \tag{1.61}$$

so that fluids with $-1/3 < \bar{\omega} < 1$ (which cover all ordinary forms of energy like matter and radiation) decelerate the expansion, while fluids with $\bar{\omega} < -1/3$ (that violate the strong energy condition) accelerate it. Thus, there is ample room for all sorts of complicated dynamics. However, looking back at (1.53), we can see that different isentropic components will dominate over each other at different times. During these periods (called eras), we can greatly simplify things by pretending that only one fluid exists. Solving the Friedmann equation in these conditions becomes a trivial matter and one finds that $a \propto t^{2/3(1+\bar{\omega})}$ where $\bar{\omega}$ corresponds to the dominate

fluid at the particular time. It follows, for example, that in the radiation era $a \propto t^{1/2}$, while in the matter era $a \propto t^{2/3}$.

1.3.1 The Flatness Problem

Let us now look at (1.58) a bit more closely. If the Universe only contains ordinary forms of energy, then the scale factor can never accelerate $\ddot{a} < 0$ and therefore $\dot{a} = aH$ decreases monotonically with time. This means that Ω is repelled from unity, unless Ω is exactly one, in which case the universe is always flat. In other words, if $|\Omega - 1|$ is close to zero today, it must have been even closer in the past. How much closer, say at the time of nucleosynthesis when the Universe was about a second old? We can get a rough idea by assuming a radiation dominated background ($|\Omega - 1| \propto t$) and an age for the Universe of about 10^{17} s. Since we know that Ω today doesn't differ from unity, say by more than an order of magnitude, i.e. $|\Omega_0 - 1| < 10^{-1}$, it follows that

$$|\Omega(t_{nuc}) - 1| \lesssim 10^{-16}\,, \tag{1.62}$$

which constitutes an extraordinary constraint! An Ω outside this interval, at such an early age, will either lead to a closed universe that recollapses almost immediately or to an open universe that quickly enters the curvature phase and cools down below 3 K within the first few seconds of existence. The flatness condition $\Omega = 1$ is therefore an unstable critical point for all ordinary 'strong energy abiding' models. Thus, it is rather puzzling that the Universe has managed to survive for so long, if it only contains ordinary forms of energy like matter and radiation.

1.3.2 The Horizon Problem

The flatness problem is not the only problem affecting ordinary models. Another is the following: After removing the dipole anisotropy of the cosmic microwave background (caused by the Earth's peculiar motion in the FRW frame), we are

left with a residual anisotropy that is less than one part in 10^{-4}. It can be shown that the comoving horizon scale at the time of last scattering (i.e., the maximum theoretical distance photons could have travel up until decoupling), is given by $180\Omega_0^{-1/2}h^{-1}\,\mathrm{Mpc}$ [45], which typically subtends about one degree in the sky. This means that regions separated by more than one degree in the sky were never in causal contact prior to last scattering. So how come is the CMB so uniform all over? This strong uniformity has to be forced upon ordinary models, at a very early stage.

A similar situation exists with nucleosynthesis [46, 47]. Because the nuclear reactions responsible for the formation of light elements like H, He, etc. are highly non-linear processes, the primordial abundances of such elements would be inevitably affected by the presence of any anisotropies in the energy distribution at that time. Consequently, we would not be able to reproduce the observed abundances of such primordial elements (undoubtedly one of the great successes of Cosmology), unless at the time of nucleosynthesis the Universe was already an extremely homogeneous and isotropic place to begin with. Once again this has to be imposed on ordinary models.

1.4 Dark Energy in the Past

There are other problems we will discuss, but for now let us focus on the ones already listed. In essence they come down to the following: ordinary models of the universe have to be *extremely* fine-tuned into a flat homogeneous and isotropic state, at a very early stage, in order to match current observations. In general, that the parameters of a given model have to be somewhat tuned in order to reproduce a given data set, is nothing out of the ordinary; it's just the nature of the game (unless, of course, our model possessed some sort of fundamental mechanism 'built in' to provide those very same parameters; then it would just be a question of checking if observations were matched or not, without any tuning involved). *Fine*-tuning, however, is another game. That such fine-tunning is necessary for ordinary models

has lead to the introduction of *inflation* [48, 49]. In the abstract, inflation simply refers to any period where the scale factor is accelerating $\ddot{a} > 0$. An equivalent description is the following

$$\frac{d}{dt}\left(\frac{H^{-1}}{a}\right) < 0\,, \tag{1.63}$$

which many people seem to prefer as it gives a slightly more physical interpretation of inflation as the shrinking of the comoving Hubble length. During this period, Ω is 'naturally' driven towards unity rather that away, as one can easily check from (1.58). This solves the flatness problem if inflation occurs sufficiently early in the history of the Universe (before nucleosynthesis, that is) and for a sufficient amount of time. It also 'solves' the horizon problem, as inflation has the 'side-effect' of smoothing things out, allowing our patch of the Universe to have originated from a tiny homogeneous region that was well inside the Hubble scale before inflation started (see [35, 45] for a more detailed account). As a bonus, it also helps getting rid of unwanted relics (like the infamous monopoles, etc. and other exotic remnants from eventual phase transitions in the early Universe), if these form well before inflation kicks in.

Implementing inflation requires the use of exotic fields that violate the strong energy condition; such fields are normally dubbed *dark energy* and are dynamic in nature. The easiest way to implement inflation is by means of a canonic scalar field subject to a simple potential like a power law or an exponential. (Fields such as these are everywhere in modern particle physics, describing all sorts of weird particles, like the illusive Higgs particle, etc.) Nowadays, however, models involving the interplay of two scalars (called hybrid models) are considerably more popular than the single-scalar approach; it is even possible to implement inflation with higher order fields, but these are mostly 'proof-of-concept' theories rather than being very useful (see [50] for a review). Regardless of how inflation is implemented, dark energy plays a crucial role in the very early Universe.

1.5 Dark Energy in the Present

Dark energy also plays a crucial role today. We know this mainly from measuring the luminosity distances to faraway supernovae. Just as with inflation in the past, there are many ways to achieve this present state of acceleration using different exotic fields. The simplest one is a constant field Λ obeying the following Lagrangian [7, 34]

$$\widehat{\mathcal{L}}_\Lambda = -\frac{\Lambda}{8\pi G}\,, \tag{1.64}$$

which is non-other than the famous cosmological constant. Obtaining the energy-momentum tensor associated with this field is a trivial matter; in the rest-frame, it describes a perfect fluid with

$$p_\Lambda = -\varepsilon_\Lambda = -\frac{\Lambda}{8\pi G}\,, \tag{1.65}$$

frequently interpreted as the energy of vacuum or the energy associated with space itself. Current CMB data is consistent with this field having $\Omega_{\Lambda 0} \approx 0.7$ or $\bar{\varepsilon}_{\Lambda 0} \approx 10^{-8}\,\mathrm{erg/cm^3}$. On dimensional grounds, however, one would expect this vacuum energy to be of the order of m_P^4 (where m_P is the Planck mass) or about $10^{112}\,\mathrm{erg/cm^3}$, which is 120 orders of magnitude greater that what is observationally required! Alas, there is no known mechanism to enforce such a minute vacuum energy, making it the worst case of fine-tunning in the whole of Physics. This conundrum is known as the 'cosmological constant problem' and we are nowhere near to solving it. Another related problem is the 'why now?' problem: Why does the cosmological constant start dominating over matter so close to the present day? It may well turn out to be that this is just a coincidence, with no profound meaning, but if one doesn't like coincidences (or anthropic justifications, for that matter), then an explanation must be found, instead of just imposing it. Incidentally, one can also ask why inflation starts when it starts; nobody really knows. These fine-tunning problems don't sit well with cosmologists (and physicists alike) and thus alternatives are actively searched for.

Noteworthy among the many proposed alternatives are canonic scalar fields, often dubbed *quintessence* [8, 9]. These models typically involve a single scalar field

but in some cases more than one (again, just as with inflation). So-called 'tracking' models [45] exist where one can obtain an energy density evolution that is reasonably independent of the initial conditions (essencially due to attractor dynamics). However, one still has to tweak some parameters in the scalar potential to obtain this behavior; hence, it can't really be claimed as a satisfactory solution. On the other hand, given that one has yet to see a scalar field in action, it is clear that all such models are not much better justified than the cosmological constant itself (despite frequent claims to the contrary). This is further compounded by the fact that, given some time dependence for the scale factor and an energy density, one is always able to *construct* a potential for a quintessence-type model that is able to reproduce them (see, for instance, [14]). One is therefore reminded of Ockham's razor and can legitimately ask if observational data provides any strong justification for them, as compared to the conceptually simpler cosmological constant. There are many more alternatives, of course, for example, k-essence [10], phantom energy [11, 12], tachyons [13, 14], vacuum metamorphosis [15, 16], etc. but these are mostly non-canonic generalizations of the cosmological constant, in other words, toy-models thrown around to see if they stick. Given our present knowledge of fundamental physics we can't really claim any of these alternatives as being much better justified than the other.

1.6 Dark Matter Today

The radiation fraction today is measured to be $\Omega_{r0}h^2 = 4.17 \times 10^{-5}$ (this includes the CMB photons plus three massless neutrinos consistent with the standard model of particle physics). Thus, the remaining 30% of the critical energy density must come from matter alone. Yet, our current best estimate for the baryon fraction is only of about $\Omega_{b0} \approx 0.04$, which is the combined effort of variety of methods, from nucleosynthesis to 'direct counting'. It follows, that most matter in the Universe must be in a *non*-baryonic form. About this non-baryonic matter only a few things are known. For one, it must interact very weakly with ordinary matter so as to

have escaped detection thus far. It must also be comprised of particles that have been non-relativistic ('cold') for a long time. If it was 'hot', dark matter would have free-streamed out of over dense regions a long time ago, suppressing the formation of structure to a degree inconsistent with what is observed today. In the meantime, virtually every known particle in the standard model has been ruled out as a candidate for dark matter. However, beyond the standard model, the story is completely different; oodles of candidates exists (see, for instance, [51]), from massive neutrinos to axions and even stranger particles. Suffice it to say that we are still far away from a definite picture.

The simplest observation suggesting the existence of dark matter actually comes from the rotation curves of spiral galaxies. Stars in these galaxies move approximately in a circular fashion around the center; their velocities are thus simply related to the amount of matter inside their orbits. What is observed, however, is that the amount of visible matter (inferred from the luminosity distribution) is not nearly enough to justify the rotation curves of virtually every spiral. Typically the visible mass decays with the distance to the center of the galaxy, but the velocities do not; they stabilize into a constant value that extends well into the rim. Assuming that gravity is Newtonian at these scales, the difference between the rotation-inferred and the luminosity-inferred matter distribution must then be attributed to the presence of some sort of dark matter. At the moment, the major problem with this picture is the following: Numerical simulations of certain types of galaxies involving CDM, typically predict cuspy dark matter distributions (i.e. distributions that spikes at the center) that have not been observed [52]. This is not a death sentence, however, as these simulations often make assumptions that may turn out to be incorrect. For instance, the relation between baryonic high density regions and CDM is not very well known. In particular, it is possible that such areas may involve additional forces besides gravity, which in turn could affect the cuspy dark matter profiles.

On the other hand, we should keep an open mind to the possibility that dark matter doesn't exist and that it is gravity that behaves differently from expected in these scales. Historically, the first modified gravity theory getting rid of dark matter

was MOND [53], and although it was a highly phenomenological hack of Newtonian gravity, it nonetheless confounded its detractors by hypnotically reproducing almost every spiral curve with a given set of parameters. The only thing preventing its widespread adoption was the fact that it wasn't a relativistic theory. However, this has since changed with the introduction of TeVeS gravity [21]. The latest successor of MOND in this modified gravity framework is due to Moffat [22, 24] and it is able to reproduce the old MOND results [23].

In the past year, however, observations from the Bullet Cluster [54, 55] have cast considerable doubt on the ability of any modified gravity theory to replace dark matter completely. Let us see why: In a nutshell, the Bullet cluster is the result of two colliding clusters of galaxies. Now, the vast majority of ordinary matter in galaxy clusters is not contained, as one might initially think, in the galaxies themselves but is rather typically dispersed in the intergalactic medium, in the form of hot X-ray emitting gas. As the two clusters crossed each other, the hot gas from each collided with the other; as a result, both were stripped of most of their ordinary mass, which was left behind. This has been observed by the Chandra X-ray telescope. On the other hand, most galaxies passed right through as if nothing had happened. Lensing of background galaxies was then used to infer the gravitational field of the Bullet cluster. What was found was nothing short of extraordinary (see the references above): the majority of the gravitational field trailed along with the colliding clusters, it didn't stay behind with the gas of ordinary matter. Of course, modified gravity theories can, in principle, give rise to all sorts of crazy gravity 'forces', but one thing we expect is for gravity to at least point in the direction of the source. Now, in the case of the Bullet cluster, if we assume there is no dark matter present, it means that the gravity source must have been left behind with the hot gas. Nevertheless, the majority of the gravitational potential (as inferred by lensing) trailed along with the clusters, and therefore does not point back to the supposed source! The modified gravity necessary to explain this configuration would have to be very strange indeed. This is not to say it's impossible; in fact, Brownstein & Moffat were able to do it quite recently [56]; their gravity, however, is 'unnaturally' skewed and the details come out as over complicated. On the contrary, dark matter

seems a much simpler and considerably more direct explanation for the observed cluster dynamics, not to say background dynamics that MOND (and successors) always had difficulty in reproducing [57, 58]. It is perhaps a bit early to say with absolute certainty that dark matter really exists, but the case for it now seems stronger than ever.

1.7 Quartessence

As we have just seen, the Standard Model of Cosmology (usually denoted by ΛCDM) only captures the 'high-level' details of the Universe. The 'low-level' innards, on the other hand, are quite sketchy, to say the least. There are many ways to implement the observed background dynamics, however, they all lack any proper foundations. Our understanding of fundamental physics is hardly at a stage were we can clearly single out a preferred implementation over any other. Even worse is the fact that what we know today about dark energy and dark matter is pieced together in a very circumstantial manner. For instance, we don't even know for a fact if dark energy and dark matter are fundamentally different from each other (which is, nonetheless, a widespread belief). Macroscopically, they certainly look and behave as if they were different but, a priori, this doesn't preclude them from sharing a common origin.

Quartessence models make the bold assumption that they do; in other words, dark matter and dark energy are interpreted as being different manifestations of a common field. For obvious reasons, they are also frequently called unified dark energy (UDE) models. Clearly, they constitute a phenomenologically interesting class. One only has to take a brief look at the history of Science to realize that unifying efforts are generally at the inception of all significant progress. People were first alerted to this possibility by the appearance of the Chaplygin gas [25, 26], discussed at length in the following chapters. This exotic fluid has the peculiar ability to behave as matter *and* as a cosmological constant depending on its local density value. In fact, the homogeneous Chaplygin gas 'interpolates' these two states in a continuous fashion. It first appeared in the context of string theory

as the tension of a d-brane in a spacetime of $d+2$ dimensions; the Nambu-Goto action can be seen to describe a 'Newtonian' fluid with an isentropic equation of state given by $p = -A/\varepsilon$ [27, 28]. At a toy-model level, one can trivially generalize this to a different dependence on the energy density, as described in [29], to $p = -A/\varepsilon^{\alpha}$. In doing so, however, we lose the ability to straightforwardly interpret the pressure of this generalized fluid as the tension of a d-brane. Such interpretation, however, was hardly of a fundamental nature to begin with, so we gladly trade it for a wider phenomenological cover. Anyway, as will be shown in the next chapter, this generalized Chaplygin gas can be implemented as a non-canonic isentropic scalar field obeying a particular Lagrangian. Another key property of this generalized gas is that in the limit of $\alpha \to 0$, it becomes totally equivalent (to any order) to a standard ΛCDM two-component model; gravity alone does not distinguish the two. We are thus interested in determining how viable UDE models are as an alternative to the standard model. Their failure would, in principle, constitute strong evidence for the different natures of dark energy and dark matter.

1.8 Cardassian Expansion

We end this chapter by mentioning a very simple alternative to dark energy called *cardassian* expansion [59, 60] by Freese et al. In this model there is only matter and the current background acceleration is caused not by a 'weird' dark energy component but rather by gravity itself (a modified version, that is). To see how a matter-only expansion can come about, let us start by consider a flat Friedmann model containing only matter. In this case, we have that

$$H^2 = \frac{8\pi G}{3}\bar{\varepsilon}_m \,, \tag{1.66}$$

and since $\bar{\varepsilon}_m \propto a^{-3}$ it follows that $\dot{a}^2 \propto a^{-1}$, i.e. a decreasing function of the scale factor. Now, if the scale factor is ever to accelerate, we must counter this a^{-1} term with an increasing one. Arguably, the simplest way to do this is by adding a power

a^m with $m > 0$; in fact, let us make $m = 2 - 3n$. We end up with

$$\dot{a}^2 = \frac{8\pi G}{3} a^{-1} + B a^{2-3n}\,, \tag{1.67}$$

which is the same as having

$$H^2 = \frac{8\pi G}{3} \bar{\varepsilon}_m + B' \bar{\varepsilon}_m^n, \tag{1.68}$$

where $B' = B/\bar{\varepsilon}_{m0}^n$ is a constant. But this is precisely the cardassian model; it consists of a very simple ad hoc modification to the Friedmann equation. These power-law corrections appear frequently in many modified gravity theories that involve extra-dimensions.

However, since $\bar{\varepsilon}_m/\bar{\varepsilon}_{m0} = a^{-3}$, we can trivially rewrite (1.68) in the form

$$\mathrm{H}^2 = \Omega_{m0} a^{-3} + \Omega_{\phi 0} a^{-3n}\,, \tag{1.69}$$

where $B' = \Omega_\phi/\bar{\varepsilon}_{m0}^n$, $n = 1 + \omega$ and also $\Omega_m + \Omega_\phi = 1$. It follows that we can readily interpret the cardassian expansion as if being caused by a constant-$\bar{\omega}$ quintessence field with $\bar{\omega} = n - 1$. (Note that since $m > 0$ then $n < 2/3$, implying $\bar{\omega} < -1/3$.) Therefore, for most practical purposes, the cardassian model is indistinguishable from a two-component model of dust and quintessence. We say 'for most practical purposes' since the cardassian model does not specify the behavior of cosmological density fluctuations on scales larger than the horizon (although it is assumed that Newtonian gravity holds on small scales). Cardassian models are thus incomplete, effective toy-models describing the *average* universe, which is what one really needs for most practical purposes. On the other hand, if we wanted to calculate the CMB anisotropy on large angular scales, we would need to go beyond this simplified procedure. It is also worth emphasizing that one can't meaningfully claim that one interpretation is much better justified than the other, at least for now, since both are nothing more than toy-models.

Chapter 2

Background UDE Cosmology

In this chapter, we will explore in detail the homogeneous properties of UDE models, mainly through a generalized Chaplygin gas [29] (henceforth abbreviated gCg). Everywhere in $\mathcal{M}$, this exotic fluid bears the following equation of state

$$p = -A/\varepsilon^{\alpha}\,, \tag{2.1}$$

where $A > 0$ and $|\alpha| \leqslant 1$ are constants (the condition on α insures that the gCg obeys the dominant condition $|\bar{\omega}| \leqslant 1$). For reasons that will become apparent in the following chapter, we will be mostly concerned with cases where $\alpha \geqslant 0$, essentially to avoid instabilities associated with imaginary sound speeds. This isentropic fluid has many remarkable properties that make it cosmologically interesting. Let us start our exploration by inquiring how the energy density of a homogeneous gCg evolves with 'time' (the scale factor, really). To find out, we must solve (1.52) which is not too difficult to do; the trick is to multiply everything by $(\bar{\varepsilon}a^3)^{1+\alpha}$ and rewrite the resulting expression as

$$d[a^{3(1+\alpha)}(\bar{\varepsilon}^{1+\alpha} - A)] = 0\,. \tag{2.2}$$

It follows that

$$\bar{\varepsilon} = \bar{\varepsilon}_0 \left[\mathcal{A} + (1-\mathcal{A})\left(\frac{a_0}{a}\right)^{3(1+\alpha)}\right]^{1/1+\alpha}, \tag{2.3}$$

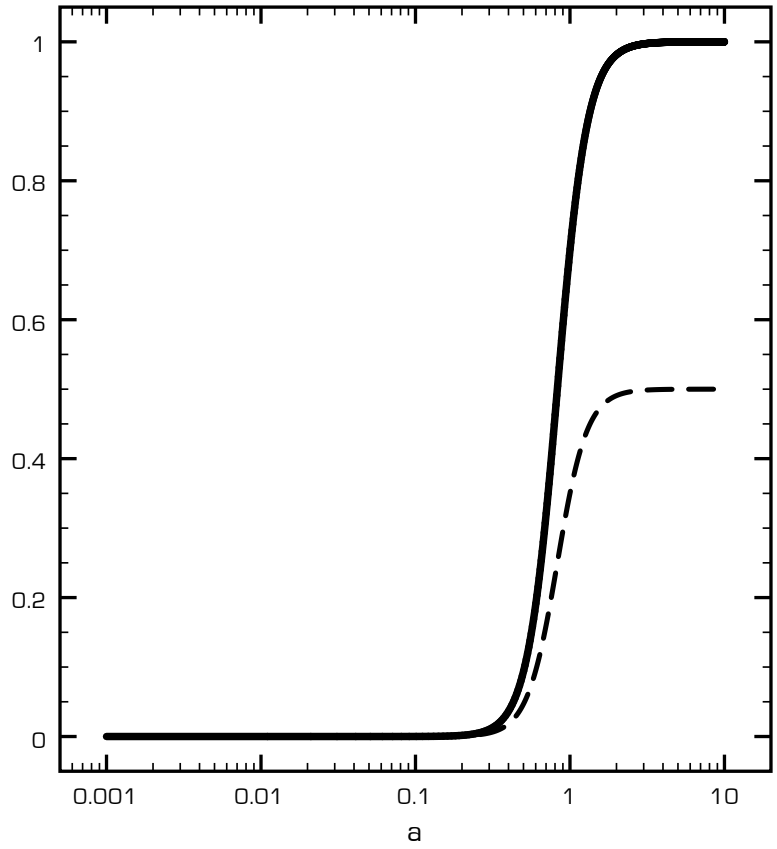


Figure 2.1: The gCg 'state parameter' $|\bar{\omega}| = |\bar{p}/\bar{\varepsilon}|$ (solid line) and its square sound speed $\bar{c}_s^2$ (dashed line; see Chapter 3, for a discussion), as a function of the scale factor. Notice the 'phase' transition from dust to a cosmological constant. Also note how late in the matter era this transitions occurs.

where $\mathcal{A} = A/\bar{\varepsilon}_0^{1+\alpha}$ is a constant. (This expression is only strictly valid for $\alpha > -1$, though.) Some unusual things are immediately noticeable, namely the fact that when a is small $\bar{\varepsilon} \propto a^{-3}$ and $\bar{p} \approx 0$ (or $\bar{\omega} \approx 0$) and that when a is big, $\bar{p} \approx \bar{\varepsilon} \approx const$ (or $\bar{\omega} \approx -1$). This suggests that the homogeneous gCg smoothly transitions from a 'matter state' to a 'vacuum energy state' as the Universe expands (see Fig. 2.1). This suggestion is at the heart of *quartessence* and the idea of unified dark energy. Another interesting property is obtained by making $\alpha = 0$ in (2.3); the gCg then reduces to a ΛCDM model with an equivalent $\Omega_\Lambda = 1 - \mathcal{A}$. (Of course, this argument only establishes a background equivalence, valid in the absence of perturbations; it

turns out, however that this equivalence actually extends to any order, as we will demonstrate in the following sections.) Incidentally, when $\mathcal{A} = 1$, the gCg reduces to a Λ cosmological constant. To wrap up the basic characteristics of this gas, we highlight the following counter-intuitive properties: no matter how much we expand the gCg, its density never drops below a certain value $\bar{\varepsilon} \geqslant \bar{\varepsilon}_0 \mathcal{A}^{1/1+\alpha}$ (this minimum density corresponds to the 'vacuum energy' the gCg tends to). Also, the lower its density, the higher its pressure (in modulus, that is, if $\alpha > 0$). Finally, we can infer from the Raychaudhuri equation (1.55) that a gCg background will start accelerating when

$$a > a_\star = [1 - \mathcal{A})/2\mathcal{A}]^{1/3(1+\alpha)} \,. \tag{2.4}$$

Thus, if $\mathcal{A} > 1/3$, this acceleration will occur prior to the present day. On the other hand, if $|\mathcal{A} - 1| \gtrsim 10^{-9}$ for any α, the acceleration will occur after recombination. This means that for most $(\mathcal{A}, \alpha)$ values, the background transition of the homogeneous gCg from CDM to Λ, will occur in the matter era.

2.1 The gCg as a Scalar Field

Let us now address the question of implementing (in a low-level way) the gCg as a scalar field. This is a fundamental step to further our knowledge of UDE. Recall that not all perfect fluids are isentropic, of course; quintessence is a prominent example. If we go back to (1.46), we can easily see that the pressure $p = 2X - \varepsilon$ depends not only on the proper energy but also on the kinetic term. If we plug these expressions into (1.51), we obtain

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \frac{dV}{d\phi} = 0 \tag{2.5}$$

the same had we used (1.32) directly. This differential equation closely resembles a damped oscillator. Thus, in general, the field will 'roll down the potential' and the 'friction term' will dampen the motion. Consequently, a canonical scalar field in a sufficiently shallow potential will roll very slowly, leading to a kinetic term $X \ll V$

and to $\omega \approx -1$. This means that some scalar fields can act as vacuum energy (and those that do are properly called quintessence) but unlike their cosmological constant counterpart, their proper energy density is allowed to vary. Note that in this framework, Λ is merely a quintessence field 'frozen' everywhere in the same stable potential minimum.

Now, can we describe the isentropic gCg using a single (real) scalar field? Yes, albeit not with quintessence; a new kind of scalar is required. To show this, let us first generalize the canonical Lagrangian from $\widehat{\mathcal{L}} = X - V = p$ to an arbitrary function $p(X, \phi)$ of the kinetic term and the field [33]. This defines a new class of scalars broad enough for our purposes. (Note that we are still calling the new Lagrangian p; the reason for this is that it still plays the role of pressure as before, as we will shortly see.) It follows, by varying the action of this scalar that

$$\begin{aligned} \delta S_\phi &= \int d^4x \, \left(\delta p \sqrt{-g} + p \, \delta\sqrt{-g}\right) = \int d^4x \, \left(\frac{\partial p}{\partial X}\delta X \sqrt{-g} + p \, \delta\sqrt{-g}\right) \\ &= \int d^4x \sqrt{-g}\delta g^{\mu\nu} \left[-\frac{1}{2}\nabla_\mu\phi \, \nabla_\nu\phi \, p_{,X} - \frac{1}{2}g_{\mu\nu}p\right] , \end{aligned} \tag{2.6}$$

where $p_{,X} \equiv \partial p/\partial X$, and so

$$T_{\mu\nu} = p g_{\mu\nu} + p_{,X}\nabla_\mu\phi\nabla_\nu\phi \, . \tag{2.7}$$

Thus, if $p = X - V$, (2.7) reduces to (1.42), as it should; and also just as with quintessence, we can still explicitly rewrite this energy-momentum tensor in a perfect fluid form, by making the following identifications

$$u_\mu = \frac{\nabla_\mu\phi}{\sqrt{2X}} \, , \qquad \varepsilon = 2Xp_{,X} - p \, . \tag{2.8}$$

From this we conclude that if $p = p(X)$, then $\varepsilon = \varepsilon(X)$. Unfortunately, it's not always possible to invert $\varepsilon(X)$, and obtain $X(\varepsilon)$ but when it is, the fluid has an *explicit* isentropic equation of state $p = p(\varepsilon)$. A useful example is $p \propto X^n$ which, as one can easily check, describes a constant $\omega = 1/(2n-1)$ fluid; in particular, when $n = 0$ the scalar corresponds to a Λ cosmological constant, $n = 1$, to a massless scalar field, $n = 2$, to background radiation, and so on. In the limit of large n, the scalar can be interpreted as dust (pressureless non-relativistic matter).

How about the gCg? We can try retro-engineer (2.8) to ascertain the necessary Lagrangian by solving

$$\varepsilon = 2X\alpha \frac{A}{\varepsilon^\alpha} \frac{\varepsilon_{,X}}{\varepsilon} + \frac{A}{\varepsilon^\alpha}\,, \tag{2.9}$$

a non-linear differential equation. Before you pull some hairs trying to solve this one (we certainly did), we'll tell you the trick to do it. It's actually quite simple: divide everything by ε and rewrite the expression in terms of a new function $\xi = A/\varepsilon^{1+\alpha}$. We end up with a much nicer linear version

$$1 = -\frac{\alpha}{1+\alpha} 2X\xi_{,X} + \xi\,, \tag{2.10}$$

that is simple to solve. The solution is $\xi = 1 - (2X)^{\frac{1+\alpha}{2\alpha}}$ and thus the necessary Lagrangian for reproducing the gCg becomes

$$p(X) = -\frac{A}{\varepsilon^\alpha} = -\left(A\xi(X)^\alpha\right)^{\frac{1}{1+\alpha}}\,. \tag{2.11}$$

Note that to insure that $p(X)$ is a non-null real value we must have $0 < 2X < 1$: The lower limit comes from the dominant condition while the upper limit corresponds to the null pressure case.

2.2 The gCg $\alpha \to 0$ Limit

Implementing the gCg as a scalar field obeying a particular Lagrangian is useful on a number of levels. For one, it allows us to prove our earlier assertion that in the limit $\alpha \to 0$, the gCg is totally equivalent to an ordinary ΛCDM model. This fact plays an important role when comparing UDE models with observations. While obviously true in the absence of perturbations, the need to explicitly demonstrate this beyond a background equivalence only became apparent after [61] appeared. In it, Fabris, Gonçalves & Ribeiro made the surprising claim that the linear evolution of perturbations actually differed in each case. Shortly after, we showed in [62] that this was not true; their equivalence to first order was established (in Chapter 3, we give a formal demonstration of this) and it was equally argued that the correspondence

went well beyond linear order. Here, we give a much simpler demonstration of this: Expand (2.11) in a power series and take the limiting case $\alpha \to 0$

$$\begin{aligned} p(X) &= \lim_{\alpha\to 0} -A^{1/1+\alpha}\left[1 - \frac{\alpha}{1+\alpha}(2X)^{\frac{1+\alpha}{2\alpha}} - \frac{1}{2}\frac{\alpha}{1+\alpha}\frac{1}{1+\alpha}(2X)^{\frac{1+\alpha}{\alpha}} + \cdots\right], \\ &= -A + 0\cdot(2X)^{\infty} = -A\,. \end{aligned} \tag{2.12}$$

We see that everywhere in $\mathcal{M}$, the Lagrangian density decomposes nicely into a cosmological constant plus 'matter', thus demonstrating the equivalence to any order between ΛCDM and the $\alpha = 0$ limit of the gCg; gravity alone does not distinguish the two. Moreover, since both are to a certain extent simply toy-models with somewhat nebulous fundamental motivations, this is probably as far as they can be meaningfully compared.

2.3 A Truly 'Atomic' Fluid?

We have just established the equivalence between the gCg and ΛCDM in the limit $\alpha \to 0$. An obvious follow-up question is wether this equivalence between a unified model and some family of minimally coupled components is valid in general. Recently, Kunz has argued in [63] that it is always possible to split a single unified dark energy fluid into several minimally coupled components or, conversely, to combine several fluids into a single fluid that behaves in exactly the same way as the original mixture (from a cosmological point of view, that is). Although this is a rather obvious statement in the absence of perturbations, Kunz also argued that this degeneracy went beyond the background level.

Before trying to answer this question, we need to clarify what we mean by a 'single fluid'. It is fairly obvious that by allowing the complexities of the fluid to be arbitrarily large (for example, by considering very high-order tensor fields), we may get disproportionately non-trivial fluid dynamics. In such cases, are we in the presence of a *single* fluid or *several* interacting fluids? In general, even something as trivial as a complex scalar field is prone to a multi-fluid interpretation. These

considerations highlight the fact that we must be very careful about what we call a 'single' fluid. In this thesis, we will refer to a single indivisible ('atomic') fluid as one whose Lagrangian has the form $\widehat{\mathcal{L}}(X, \phi)$, where ϕ is a real scalar. This definition covers, for instance, quintessence and isentropic fluids, which are hard to imagine as anything but single fluids. Treating UDE as a *single* fluid (in the sense just described) is arguably the best possible way of implementing quartessence; we will call this implementation, *canonic* quartessence, thus distinguishing it from other more elaborate and complex possibilities that lack the simplicity suggested by the isentropic gCg.

It goes without saying, of course, that we can always split the energy-momentum tensor of any fluid into several components. There are infinite ways to go about doing this. In the case of quartessence, the critical question is how to interpret any such decomposition: Are the resulting components real *physical* fluids? Do they exist independently from each other? We have already seen that the answer is *no* if the gCg does originate from a single scalar field governed by (2.11). Each piece is then a *virtual* component without independent existence (except in the special case $\alpha = 0$). In fact, the evolution of the individual virtual components is not, in general, constrained by causality. (From this point of view, the perturbative treatment of the gCg in [64], by Bento, Bertolami & Sen, is inconsistent.)

Conversely, if we are in the presence of various fluids, we can also add their energy-momentum tensors. However, the dynamics of the resulting fluid is potentially very complex and, in general, will not be describable by a real scalar field subject to a Lagrangian of the form $\widehat{\mathcal{L}}(X, \phi)$. From a cosmological point of view, all the relevant information is contained in the energy-momentum tensor acting as the source of the gravitational field. Pending laboratory evidence (which, in principle, can detect not only the fields themselves but even the couplings between them), we are only sensitive to the total (effective) energy-momentum tensor $T_{\mu\nu}$. Consequently, Kunz argues that cosmology alone does not provide useful information on wether a single unified dark energy fluid or a family of minimally coupled interacting fluids is responsible for the observations.

This isn't exactly true, however. Let us consider the evolution of very large wavelength perturbations in $\mathcal{M}$. These perturbations essentially amount to homogeneous and isotropic patches of the Universe. Now, for the sake of argument, let us pick one of these patches and decompose its total energy density ε (and pressure p) in two different ways:

I. $\omega_1 = -1 = -\omega_3$ and $\omega_2 = 0$ with $\varepsilon_i/\varepsilon = 1/3$ for $i = 1, 2, 3$

II. $\omega_1 = -\sqrt{6}/3 = -\omega_2$ with $\varepsilon_i/\varepsilon = 1/2$ for $i = 1, 2$

where the components are minimally coupled. It is easy to check that both have the same $\dot{\varepsilon}$ and $\dot{p}$ but clearly their evolutions will differ. This means that even if we have the same initial conditions for the effective fluid $\varepsilon, \dot{\varepsilon}, p, \dot{p}$ (and H) in the patch, the subsequent evolution depends on the details of the composition. (Moreover, we are even allowed to set up adiabatic or iso-curvature fluctuations or indeed a combination of them.) This degeneracy (we have given two explicit examples, but there are infinite more), on the other hand, is not present in single fluid models: for a given equation of state $p = p(\varepsilon)$, the evolution is univocally determined by the initial density and pressure conditions on the patch. To a certain extent, this can be thought as saying that a single fluid only carries one degree of freedom. This 'lack of freedom', therefore, significantly constrains what single fluids can do.

2.4 UDE Background Tests

We will now embark on the task of testing UDE background dynamics in the simplest possible manner, using the homogeneous gCg as a prototype for quartessence. This effort is mostly based on our previous work [65]; there, a sample of 92 supernovae was used to constrain a homogeneous flat model involving matter and a generalized Chaplygin gas. Since that time, the number and quality of supernovae observations have increased somewhat and here we use the latest dataset to update our previous work. The differences won't be dramatic but are nonetheless important. Before we

start, however, we would like to stress the fact that our model includes a separate matter component; it is not a single gCg model, as it is commonly done [66, 67, 68, 69, 70, 71, 72, 73, 74, 75]. Radiation will be ignored and we concentrate on a flat geometry, as this is currently the preferred geometry.

2.4.1 Background Model

The Friedmann equation for our model is given by

$$H^2/H_0^2 = \Omega_m^0 \left(\frac{a_0}{a}\right)^3 + \Omega_{\rm gCg}^0 \left[\mathcal{A} + (1-\mathcal{A})\left(\frac{a_0}{a}\right)^{3(1+\alpha)}\right]^{1/(1+\alpha)} , \tag{2.13}$$

where $\Omega_m^0 + \Omega_{gCg}^0 = 1$ insures flatness and $0 \leqslant \alpha \leqslant 1$ avoids possibly 'nasty' instabilities (see Chapter 3, for a discussion). The luminosity distance, on the other hand, is simply

$$d_L = d_H(1+z)\int_0^z \frac{dz'}{H(z')} , \tag{2.14}$$

where d_H is the Hubble scale today and the distance modulus is given by $\mu \equiv m - M = 5\log d_L + 25$. Now, it is a well know fact that the luminosity distance is not a very sensitive quatity to time variations of the background state parameter $\bar{\omega}(z)$ [44, 43]. Essentially, this happens because the integral in (2.14) smoothes these variations out, in other words, the luminosity distance is mostly sensitive to an effective $\bar{\omega}$. In the case of interest to us, the time variation in $\bar{\omega}$ comes from the gCg component, which is controlled by the parameters $\mathcal{A}$ and α. Of these, $\mathcal{A}$ has the strongest influence on the evolution of the background, as can be easily inferred. This means that α will be poorly constrained by supernovae luminosity distances in comparison to $\mathcal{A}$. Indeed, the currently available data do not place any significant constraint on α (see Fig. 2.4). Several supernovae observations at much higher redshifts are still necessary to improve our knowledge of this parameter. This, however, can already be achieved by using other (higher-order) methods to be discussed in the following chapter.

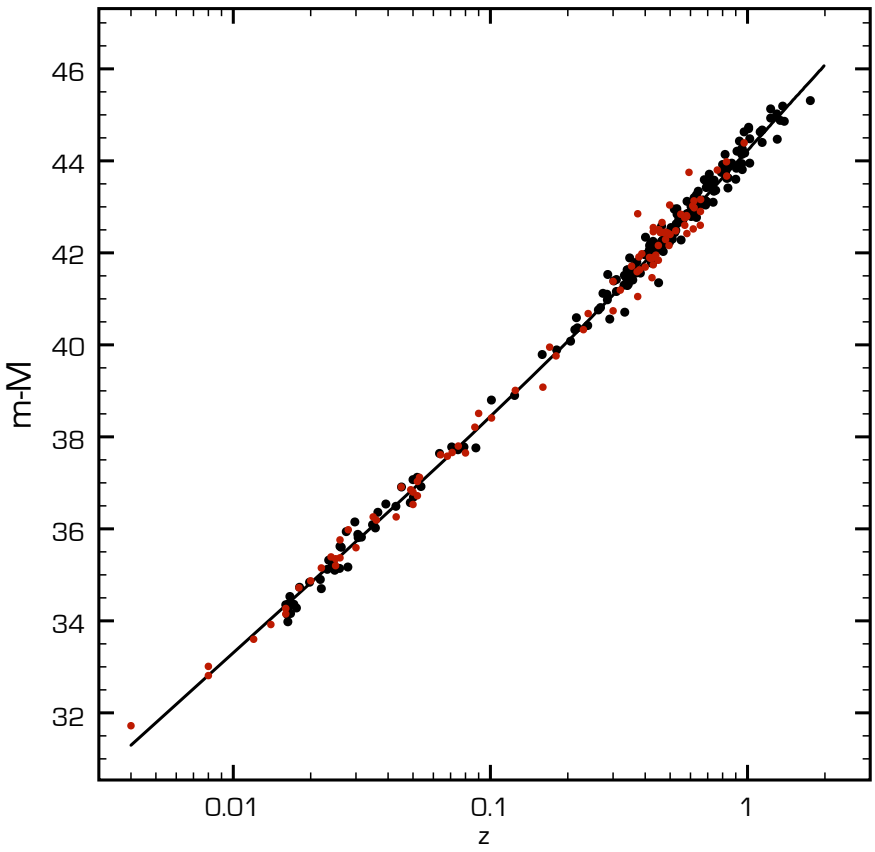


Figure 2.2: Modulus distance vs. redshift for the old (red) and new (black) supernovae samples; error bars not shown. The solid line represent a ΛCDM fit with 30% cold dark matter.

2.4.2 Old vs. New Samples

The supernovae sample used in our original work [65] was assembled from the initial release of the Supernova Cosmology Project (SCP) [1] and from the High-z Supernova Search Team (HzST)[2]: It included 60 Type Ia supernovae from the first and 50 from the second, 18 of which where common, for a grand total of 92 supernovae. The data from the different groups was 'sewn' together following a procedure first described in Wang's [76] (we refer the reader to our previous work for details). Here, we will use a new dataset described in [77]: It includes 60 Type Ia supernovae from

the ESSENCE (Equation of State SupErNovae trace Cosmic Expansion) survey [78], 57 from the SNLS (SuperNovae Legacy Survey) [79], 45 nearby supernovae and 30 more at $0.216 \leqslant z \leqslant 1.755$ observed by the Hubble Space Telescope and classified as 'gold' by Riess et al [80]. (Again, we defer the reader to the original paper by Davis et al for the details on how this dataset was assembled.) Both samples can be seen depicted in Fig. 2.2 along with a standard ΛCDM fit. As is immediately apparent, both samples have more or less the same scatter but the new one includes considerably more observations at higher redshifts. These high-z supernovae are the main reason behind the improvements we'll get.

2.4.3 Supernovae Statistics

In this section, we briefly discuss the statistics involved in constraining our model using supernovae distance modulus (see, for instance, [81, 82] for a more detailed introduction). We start by denoting our background model $M(\mathbf{p})$, where $\mathbf{p}$ are the parameters like $\mathcal{A}$, α, etc. Given several $\{\mu_i\}$ observations, we want to obtain the probability of $M(\mathbf{p})$, in other words, $p(M(\mathbf{p}) \,|\, \mu_1 \dots \mu_n)$. Here, it is safe to assume that the several supernovae observations are independent from each other; hence, it follows that

$$p(M(\mathbf{p}) \,|\, \mu_1 \dots \mu_n) = p(M(\mathbf{p}) \,|\, \mu_1) \dots p(M(\mathbf{p}) \,|\, \mu_n)\,. \tag{2.15}$$

Now, let us focus on the probability of $M(\mathbf{p})$ given a single observation. From Bayes theorem, we know that

$$p(M(\mathbf{p}) \,|\, \mu_i) = \frac{p(M(\mathbf{p}))}{p(\mu_i)}\, p(\mu_i \,|\, M(\mathbf{p}))\,. \tag{2.16}$$

Here, $p(\mu_i)$ can be seen as a constant independent of the model. On the other hand, $p(M(\mathbf{p}))$, called the model's prior, is not known a priori; the common bayesian lore is to assume a uniform prior and we will do so in this analysis. As to $p(\mu_i \,|\, M(\mathbf{p}))$, we will assume that supernovae observations are normally distributed around the models prediction $\mu_i(\mathbf{p})$ with a σ_i dispersion.

Thus, we conclude that

$$p(M(\mathbf{p})\,|\,\mu_1 \ldots \mu_n) \propto \exp\left(-\frac{1}{2}\sum_i \frac{(\mu_i - \mu_i(\mathbf{p}))^2}{\sigma_i^2}\right) . \tag{2.17}$$

The summation terms in (2.17) are usually called χ_i^2 and each σ_i can be approximated by the measuring error associated with the i-th observation. The normalized probability distribution is then obtained by dividing the above expression by the sum over 'all' possible values of the parameters. From this, it is a simple matter to obtain the confidence regions for the parameters: for every grid point $\mathbf{p}$, we accumulate the probabilities larger than the probability at that point. The 68%, 95% and 99% contour curves of this accumulated distribution determine the usual 1σ, 2σ and 3σ confidence regions. We should point out, however, that there are a few potential *caveats* to this χ^2 analysis that result from assuming Gaussian errors in the supernovae measurements: see, for example, [83] where a modified median statistics was used instead.

2.5 Results and Comments

In Fig. 2.3, we have plotted the confidence regions for the $(\Omega_{\mathrm{gCg}}, \mathcal{A})$ parameters that resulted from χ^2 fitting our model to the luminosity distances of the old and new supernovae samples. The uncertainty in the Hubble constant was eliminated by summing over several h. α was also summed over. While browsing these figures, keep in mind that the closer $\mathcal{A}$ is to unity, the closer the gCg will be to a Λ cosmological constant. We see that the new sample has led to tighter constraints: The $\mathcal{A}$ parameter is restricted with a 95% confidence level to be in the region $0.94 < \mathcal{A} < 1$ if $\Omega_m^0 \sim 0.3$ while if $\Omega_m^0 \simeq 0.04$, $0.7 < \mathcal{A} < 0.92$. In other words, if the matter component in our model represents non-barionic CDM, then the gCg is forced to behave as a cosmological constant (which is hardly surprising). We should point out, however, that since one of the strongest claims of the Chaplygin gas is that of a unified explanation for dark matter and dark energy, one might expect that the only components of the universe would be a Chaplygin gas and a tiny amount of

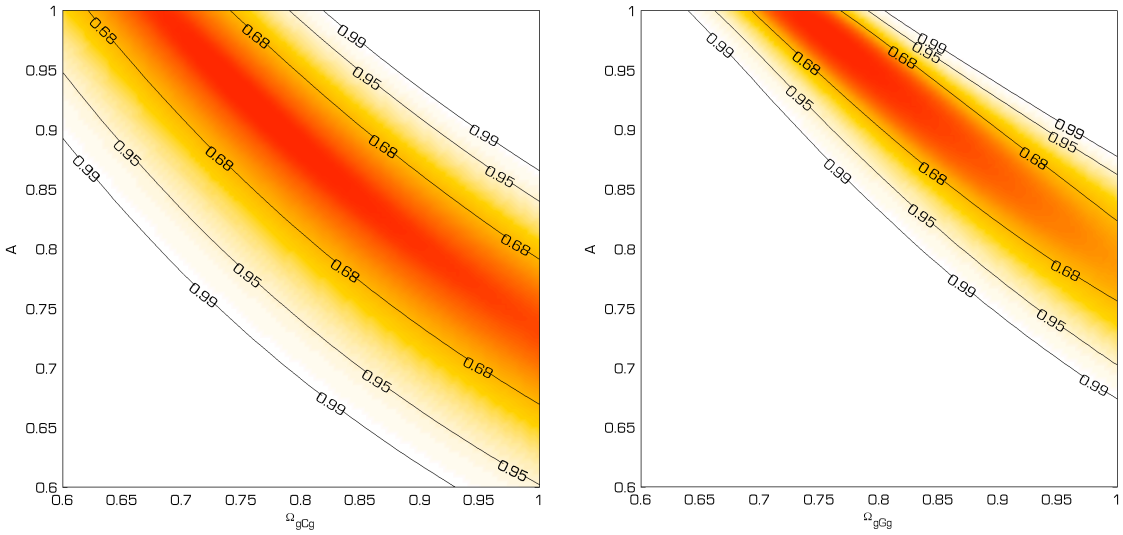


Figure 2.3: Confidence regions for the parameters $(\Omega^0_{\rm gCg}, \mathcal{A})$ resulting from χ^2 fitting the old (left) and new (right) supernovae samples. The parameters h and α have been summed over. Note that the gCg is not alone here; it coexists with a minimally coupled matter component such that $\Omega^0_{\rm gCg} + \Omega^0_m = 1$.

baryonic matter. In this case, we see that the supernovae data strongly exclude a Λ-like behavior.

In Fig. 2.4, we depict the case where only baryons (with a present day density of of $\Omega^0_b = 0.044$ [84]) coexist with the generalized Chaplygin gas. We clearly see that the parameter α is not constrained by the available supernovae samples, while $\mathcal{A}$ is around 0.8. To investigate this case a little further, we lift the flatness restriction but fix, for simplicity, $\alpha = 1$. The results are depicted in Fig. 2.5. A large degeneracy is clearly evident. However, we can say with a 95% confidence level that $\mathcal{A} > 0.75$. Note that for lower values of the Chaplygin gas density, we're forced to approach the cosmological constant limit. Also note that for the case of a plane geometry the acceptable values for $\mathcal{A}$ already exclude a Λ scenario.

In summary, if we assume that the present day matter density is around 30%, the gCg is forced to behave very closely to a cosmological constant. (Of course,

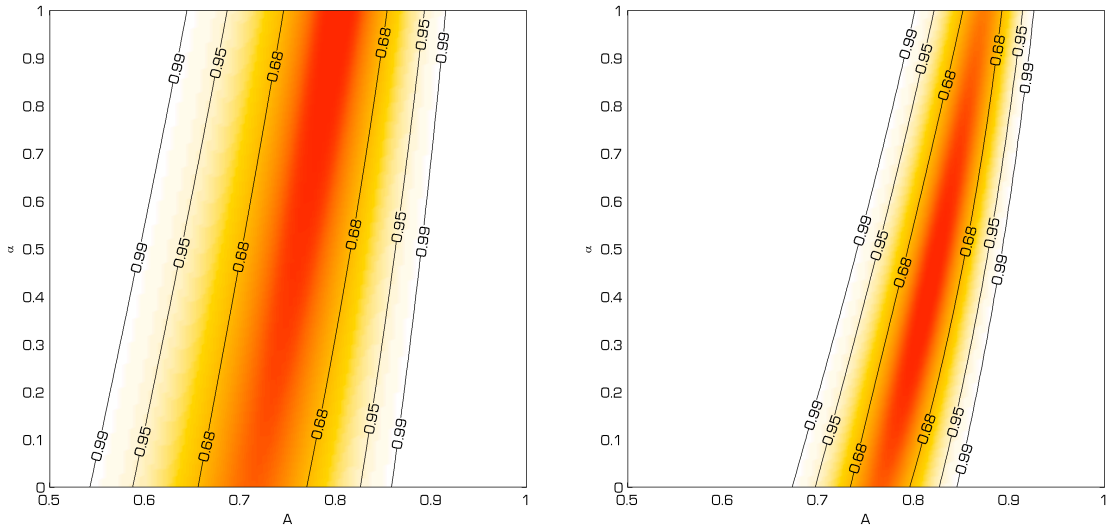


Figure 2.4: Confidence regions for the gCg parameters $(\alpha, \mathcal{A})$ resulting from χ^2 fitting the old (left) and new (right) supernovae samples. Here, only baryons with a present day density of of $\Omega_b = 0.044$ coexist with the gCg. The Hubble constant was summed over.

this result is also known to apply for standard quintessence models [85].) Indeed, if by an independent method we were able to determine the total matter content of the Universe (including dark matter) to be around ~ 0.3, then in the context of this model we would, in fact, *require* a cosmological constant so as to account for the current observational results. Conversely, the case where the matter content is entirely baryonic (arguably the best-motivated one), is the case where the differences with respect to the standard model should be maximal. In this case, the Λ-like limit is already strongly disfavored by observations.

Finally, we would like to stress the fact that these results are only strictly valid *in the absence of perturbations*. As we will later discuss in Chapter 4, non-linear small scale clustering in the quartessence component may critically affect the large scale behavior of the Universe and render these results invalid.

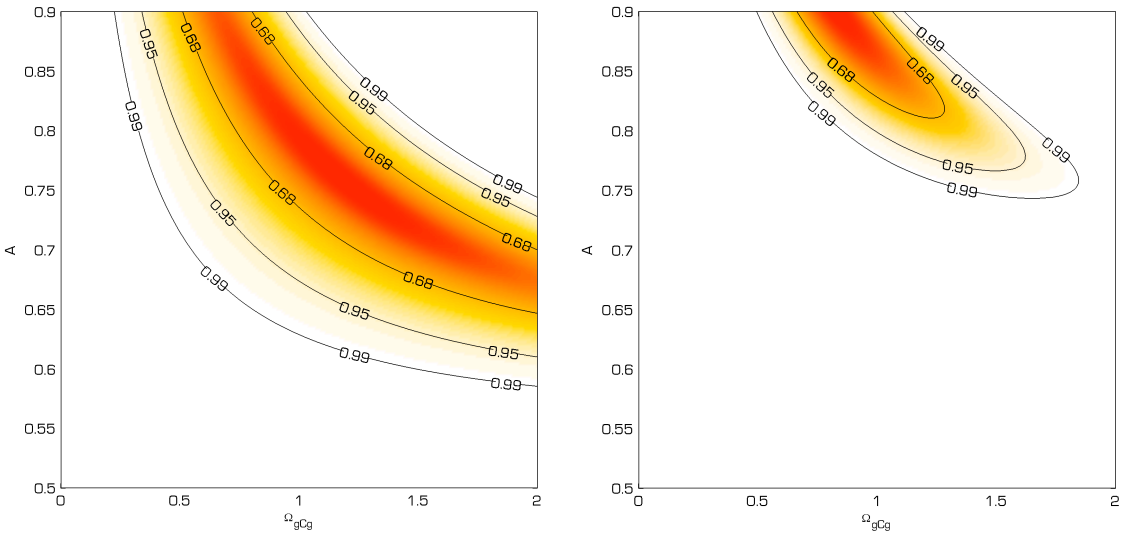


Figure 2.5: Confidence regions for the parameters $(\Omega^0_{\mathrm{gCg}}, \mathcal{A})$ resulting from χ^2 fitting the old (left) and new (right) supernovae samples. Here, baryons with a present day density of of $\Omega_b = 0.044$ coexist with a pure Chaplygin gas ($\alpha = 1$), but the flatness restriction as been lifted. The Hubble constant was also summed over.

Chapter 3

UDE — Linear Evolution

In the previous chapter, we discussed background constrains to the UDE hypothesis. The main message that emerges is that such constraints critically depend on whether one treats the gCg as *true* quartessence (replacing both dark matter and dark energy) or if one allows it to coexist with a 'normal' dark matter component (which could be called the 'Chaplygin quintessence' scenario). Going beyond this zero-order analysis and studying linear perturbation theory, seemingly dealt the first blow to the UDE hypothesis. In a provocatively titled paper, 'The End of Unified Dark Matter?' [86], Sandvik, Tegmark, Zaldarriaga & Waga constrained the mass power spectrum (the linear part) of a single gCg model using 2dF [84] data and obtained a spectacular result: $|\alpha| < 10^{-5}$. In other words, they found that the gCg was forced to behave very close to a ΛCDM model in order to reproduce the observed large scale structure (LSS) of the Universe. This was subsequently presented as a clear sign for the failure of UDE. (Strictly speaking, however, this is not a *fundamental* failure; dark matter and dark energy are still allowed to have a common origin. After all, the $\alpha \to 0$ limit of the gCg is totally equivalent to ΛCDM.) This dramatic result, however, is partially due to the fact that *baryons* have been neglected in the analysis carried out by Sandvik and his collaborators. (This concern was also voiced, but not exploited, by Colistete, Fabris & Gonçalves in [70]). The reason for this is simple to understand: as the quartessence background

transitions from 'CDM' to a cosmological constant, perturbations become heavily damped. This happens (as we will shortly show) because during this period the gCg attains very large sound speeds (unless $\alpha \approx 0$, in which case, the sound speed is very small). Thus, by adding an independent component with a low sound speed (as is the case of baryons), the normal growth of perturbations can still continue, even when the gCg starts behaving differently from CDM. By overlooking baryons, the authors of [86] have artificially constrained the possible spectra that UDE scenarios can cover. Although baryons are not that important for background studies, they are, nonetheless, quite important in the context of large scale structure. We mention, in passing, that [87, 71] also accurately studied perturbation growth in these models (including CMB). Although of a broader scope than ours (they considered a *baryon* + CDM + Chaplygin model), we do not quite agree on their interpretation of UDE. While we agree that this 'Chaplygin quintessence' scenario is all but ruled out (or at least strongly disfavored relative to ΛCDM), it seems to miss the point that these scenarios came into existence as an attempt to unify dark energy and dark matter, and that in this context their behavior can be potentially quite different from ΛCDM. Yet, this quartessence scenario is all but ignored in their discussion. Therefore, we re-derived the analysis of [86] (in order to accommodate for baryons) and part of the analysis of [71] to explore more fully the viability of the 'quartessence' scenario.

3.1 Quartessence Sound Speed

Before we discuss some of the nitty-gritty details of (linear) perturbation theory, recall that for any isentropic fluids, the 'sound speed' is a well defined concept given by $c_s^2 \equiv \delta p/\delta\varepsilon \simeq \bar{c}_s^2 = d\bar{p}/d\bar{\varepsilon}$ [33, 34, 36]. This quantity roughly corresponds to the speed with which perturbations 'spread' across the manifold $\mathcal{M}$. In the case of the gCg, this means that the sound speed is $\alpha A/\bar{\varepsilon}^{1+\alpha}$. It's a simple matter to show that this velocity is bounded by α and that today the sound speed equals $\alpha\mathcal{A}$. Since local speeds cannot exceed the speed of light, it follows that $\alpha\mathcal{A} \leqslant \alpha \leqslant 1$, implying

$\mathcal{A} \leqslant 1$ and $\alpha \leqslant 1$. There is nothing stopping α from being negative, though (in which case, the sound speed is an imaginary quantity). However, the dominant condition ($|\bar{\omega}| \leqslant 1$) does at least impose $\alpha \geqslant -1$. (At a toy-model level, we could still consider models having $\alpha < -1$ but these are all dark energy dominated at early times and we are not interested in this.) Also recall that the sound speed for any scalar field (canonic or not) is a well defined quantity given in linear theory by $c_s^2 \equiv p_{,X}/\varepsilon_{,X}$ [33]; contrast this to the isentropic result $\delta p/\delta\varepsilon$. It follows that quintessence has a constant sound speed of one; another reason why quintessence cannot reproduce the gCg. On the other hand, if we apply this formula to the scalar field governed by (2.11), the sound speed does come out the same as the isentropic one and all is well.

Thus, quartessence in general (unless $\alpha = 0$) has a non-null sound speed (see Fig. 2.1). In fact, it is a simple matter to show, starting with the Friedmann equations (1.55) and (1.56) that for any homogeneous UDE fluid

$$\frac{d\bar{\varepsilon}}{dt} = \frac{3}{4\pi G} H \frac{dH}{dt}\,, \tag{3.1}$$

$$\frac{d\bar{p}}{dt} = -\frac{1}{4\pi G}\frac{d}{dt}\left[\left(\frac{\ddot{a}}{a} + \frac{1}{2}H^2\right)\right]\,, \tag{3.2}$$

which together imply

$$\bar{c}_s^2 \equiv \frac{d\bar{p}}{d\bar{\varepsilon}} = \frac{1}{3H}\frac{d}{dH}\left[H^2\left(q - \frac{1}{2}\right)\right]\,, \tag{3.3}$$

where $q \equiv -\ddot{a}/(aH^2)$ is the 'deceleration' parameter. From this we conclude that the sign of the square sound speed $\bar{c}_s^2$ is determined by how q is evolving: If it is evolving sufficiently fast (towards negative values), then $\bar{c}_s^2 > 0$, otherwise $\bar{c}_s^2 < 0$. On the other hand, the evolution of q is linked to how fast the transition from dark matter to dark energy occurs in the background: If it is steep enough (faster than ΛCDM, that is), $\bar{c}_s^2$ will be positive (negative otherwise). Therefore, the sign of the sound speed square is directly linked to the dynamics of the background and thus carries important information.

You may not recall exactly what an imaginary sound speed entails, so we briefly

discuss it here. Well inside the 'horizon', linear theory describes the evolution of a perturbation as a wave [33, 36, 45]

$$\ddot{\delta} - \bar{c}_s^2 \, \nabla^2 \delta \simeq 0 \,, \tag{3.4}$$

where $\delta = \delta\varepsilon/\bar{\varepsilon}$ is the so-called density contrast. The general solution is thus a superposition of plane waves $\exp i(wt - \mathbf{k} \cdot \mathbf{r})$ where $w^2 = \bar{c}_s^2 \, k^2$. This makes the interpretation of the square sound speed $\bar{c}_s^2$ sign very straightforward: If $\bar{c}_s^2 > 0$, the contrast δ will oscillate as an acoustical wave, acting against the formation of voids and dense regions (pressure support). On the other hand, if $\bar{c}_s^2 < 0$, the opposite will happen: collapsing regions and voids get amplified by pressure. (For a more careful discussion see Wu's [88].) These simple considerations, coupled with the fact that the gCg sound speed grows very large during the background transition, are enough to qualitatively understand the results obtained by Sandvik et al in [86] (see Fig. 3.1): the matter power spectrum emerging from equality will be exponentially blown up for models with $\alpha < 0$, while for $\alpha > 0$, LSS will be significantly 'erased'. Thus, only models with a small α stand a chance of reproducing the observed power spectrum; what Sandvik et al found was that α had to be *very* small indeed. (CMB constraints on α have also been obtained, but since CMB formation takes place at a time where the gCg sound speed is still relatively small, the constrains on α are much less spectacular; $\alpha < 0.2$ is the latest result [77].)

3.2 Cosmological Perturbations

In this section, we briefly summarize the theory behind first order linear perturbations (for a more detailed account see [89, 90, 91, 92, 45, 93, 33]). We begin by clearly separating the real Universe $\mathcal{M}$ from its smooth idealized background $\overline{\mathcal{M}}$. Also, we will adhere to the convention of denoting quantities from $\overline{\mathcal{M}}$ with a bar on top like so $\bar{Q}$ (unless they obviously belong to the background like a and $\mathcal{H}$, for example) and no bar for quantities in $\mathcal{M}$. The only exception to this are the background coordinates that will still be denoted by the usual symbol x^μ, not $\bar{x}^\mu$. Our goal is to treat $\mathcal{M}$ as a perturbed $\overline{\mathcal{M}}$; in simple terms this means we want to

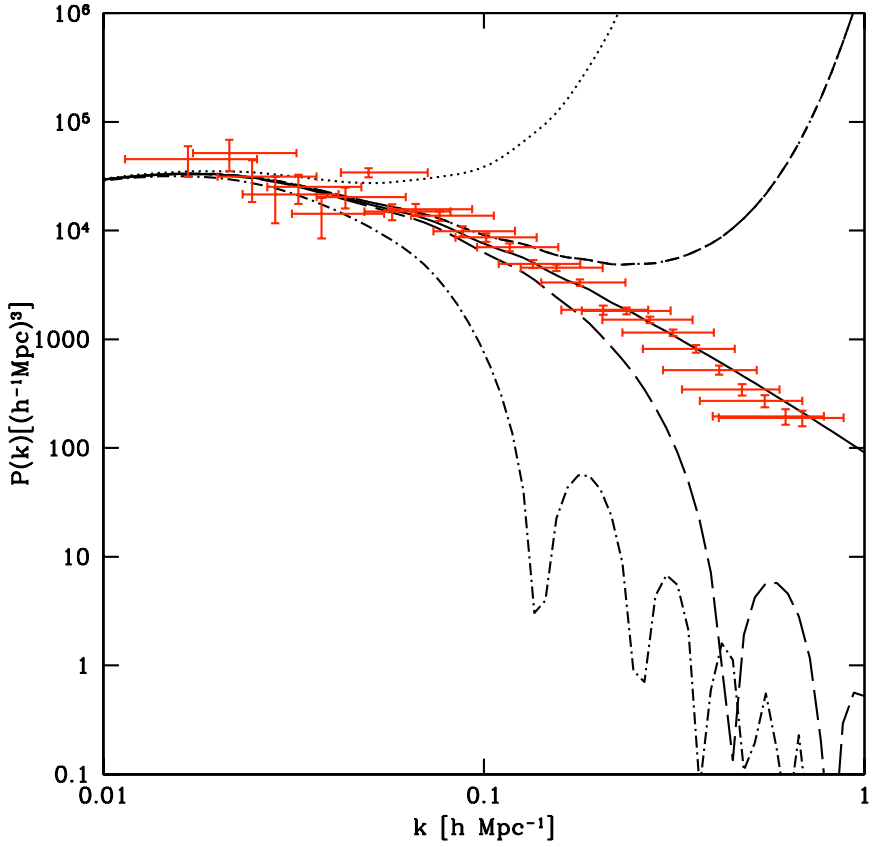


Figure 3.1: Power spectra obtained by Sandick et al [86] in the context of a single gCg model having (top to bottom) $\alpha = -10^{-4}$, -10^{-5}, 0, 10^{-5} and 10^{-4}, respectively. The red points represent the 2dF power spectrum.

decompose any quantity Q into $\bar{Q} + \delta Q$ and treat δQ as a small perturbation. With this in mind, let us set in the background the metric

$$\begin{aligned} d\bar{s}^2 &= \bar{g}_{\mu\nu}\, dx^\mu dx^\nu \,, \\ &= a^2(\eta)\left[-d\eta^2 + \bar{\sigma}_{ij}(x^k)\, dx^i dx^j\right] \,, \end{aligned} \tag{3.5}$$

where η is the conformal time and $\bar{\sigma}_{ij}$ is the 3-metric of the constant curvature slice $\overline{\Sigma}$. Since we are only interested in a *flat* $\overline{\Sigma}$, we'll simplify things from the start by setting a cartesian metric $\bar{\sigma}_{ij} = \delta_{ij}$ (a general treatment would be overkill for our purposes). In these slices the 'background fluid' is homogeneous and does not move.

Once set, we can 'export' these coordinates to $\mathcal{M}$ by simply dragging them over using any diffeomorphism $\varphi : \overline{\mathcal{M}} \to \mathcal{M}$ (1-1 smooth map) linking the two. We'll say the coordinates are the 'same' in both manifolds (and this is why we use the same symbol for the coordinates) but obviously they will have different metrics.

In $\mathcal{M}$, we'll write the metric as a perturbed $\bar{g}_{\mu\nu}$ like so

$$\begin{aligned} ds^2 &= g_{\mu\nu}\, dx^\mu dx^\nu = (\bar{g}_{\mu\nu} + \delta g_{\mu\nu}) dx^\mu dx^\nu \,, \\ &= a^2(\eta)\{-(1+2\psi)d\eta^2 + 2\omega_i d\tau dx^i + [(1-2\phi)\delta_{ij} + 2h_{ij}]dx^i dx^j\} \,, \end{aligned} \quad (3.6)$$

and assume $\delta g_{\mu\nu} \ll 1$ (since we are only interested in the linear regime). Given that it is always possible to incorporate the trace of h_{ij} into ϕ, we will assume it traceless. Now, the fluid in the η-slices (denoted by Σ) is in a perturbed state, no longer still and homogeneous. However, before we proceed, we would like to stress the fact that there is a notational sleight of hand at work in (3.6). This is related to how we *operationally* define $\delta g_{\mu\nu}$; if you find this strange consider the fact that $g_{\mu\nu}$ and $\bar{g}_{\mu\nu}$ are tensor objects that belong to different manifolds, so how do we even begin to compare them, let alone subtract them? The answer is that there are, in fact, two equivalent ways of doing this, and the above notation is designed to mask out these details. Unfortunately, this is also invariably the cause of great confusion. One way is to use *pushforwards* and *pullbacks* [31, 32, 34] to bring the objects in question to a common manifold, usually $\mathcal{M}$; then we simply subtract them and define $\delta g_{\mu\nu}$. This is the cleanest way to define perturbations but it is also the most technically demanding. A much less fancier way consists of literally 'dragging' the components $\bar{g}_{\mu\nu}(x)$ from $\overline{\mathcal{M}}$ to $\mathcal{M}$ along with the coordinates. (Note that this is not the same as pulling or pushing a tensor object; we are dragging a function from one place to another.) Then, we simply subtract the dragged components from $g_{\mu\nu}(x)$ and define $\delta g_{\mu\nu}(x)$. Both ways are physically equivalent but since dragging functions is much easier than dragging tensors, the second method is the most frequently used. Hence, we see that '$\bar{g}_{\mu\nu}$' in (3.6) doesn't literally mean the background metric but instead stands for a dragged tensor/component from $\overline{\mathcal{M}}$ to $\mathcal{M}$. (As a matter of taste, we could have dragged the metric from $\mathcal{M}$ to $\overline{\mathcal{M}}$ instead, and effectively imagine that

perturbations live as fields in unperturbed space as in [45].)

3.2.1 Gauge Ambiguities

This construction also clearly illustrates how $\delta g_{\mu\nu}$ is *operationally* tied to φ and the induced gauge (normally the word 'gauge' refers to the threading and slicing of a manifold but here we'll call φ a gauge too). Change the gauge (either by using a different diffeomorphism or by making a coordinate transformation in $\mathcal{M}$) and $\delta g_{\mu\nu}$ will change. This highlights the somewhat 'ambiguous' nature of perturbations in that they are only well defined up to a gauge transformation.

Unfortunately, while General Relativity doesn't care about gauges (they all work), linear theory is very fussy about them. This is because even a simple coordinate transformation will generally modify the order of $\delta g_{\mu\nu}$ (by making them big, for instance) thus ruining the linear approximation. Linear theory doesn't inherit the full blown gauge-freedom of General Relativity; instead it is restricted to *infinitesimal* coordinate transformations as these are the only transformations that preserve the order of $\delta g_{\mu\nu}$. If we write the infinitesimal coordinate transformation as $x^\mu \to x^\mu + \xi^\mu$, it turns out that the new perturbations

$$\widetilde{\delta Q} = \delta Q - \mathcal{L}_\xi \bar{Q} \qquad \widetilde{\delta Q} \equiv \delta Q \,, \tag{3.7}$$

where $\mathcal{L}_\xi$ is the Lie derivative along the congruence ξ, are *physically* equivalent to δQ, as far as linear theory is concerned. (Recall that gauge transformations induce in a given manifold the notion of physical equivalence between different fields and this is why the symbol $\equiv$ is used as opposed to the equal sign.) By the way, this type of transformation is nowadays called an *external* gauge transformation. It is called external in the sense that it acts on $\mathcal{M}$, as opposed to a transformation that would act *internally* on a tangent space to $\mathcal{M}$. Most 'gauge theories' [94, 95] are in fact of the internal kind; electromagnetism is a well known example (the gauge transformation modifies the A_μ field without ever touching the Minkowski coordinates) and so are the weak and strong forces. General Relativity, on the

other hand, is a more complicated gauge theory because the interval ds^2 has to be preserved in the external spacetime.

3.2.2 Gauge Freedom

This gauge freedom also demonstrates that the 10 degrees of freedom in the metric perturbations are not all independent from each other; only 6 of them correspond to actual 'physical' observables. The other 4 are called 'gauge degrees of freedom' and they represent *fictitious* perturbations in the sense that they don't correspond to any real perturbation in the curvature of the manifold; in other words, the Riemann tensor is not affected by them. They are simply coordinate artifacts that do not warp or bend spacetime. Thus, we must somehow eliminate this gauge-freedom before we can meaningfully talk about perturbations. There are two ways to achieve this: The first and most obvious is to simply define a coordinate system (a gauge) and stick to it all the time. The other route is to define gauge-invariant quantities and always work with them. (Gauge-invariant quantities are real physical quantities that can be measured experimentally with tools like measuring rods, clocks, counting devices etc., like the E and B fields of electromagnetism; contrast this to the A and ϕ fields which cannot be measured uniquely.) Then the gauge won't matter. While such gauge-invariant formalisms do exist [90, 96], for our purposes, working in a specific gauge is more than sufficient, and we chose to work in the popular synchronous gauge (a particular one, that is).

3.2.3 (Initially Unperturbed) Synchronous Gauge

First introduced by Lifshitz in 1946, the synchronous gauge is defined by the conditions $\psi = \omega_i = 0$. This gauge is physically simple to understand: $\psi = 0$ makes η coincide with the proper time along the threads and $\omega_i = 0$ makes the threading orthogonal to the slices. Furthermore, the threads are *geodesics* as one can easily verify by checking that $u^\mu \nabla_\mu u^\nu = 0$ along them. This means that 'fundamental'

observers (i.e., free falling inertial observers) move along η-threads and thus do not change their spacial coordinates.

Pausing for a moment, this is actually quite strange; in most cases, if an observer is 'parked' at a particular position, it is because some external force is keeping it there. It follows that such an observer should not be an inertial observer to begin with (this is what happens to us at the surface of the Earth, by the way). Nevertheless, the local observers that define this gauge *are* inertial despite having constant x^i coordinates. Having said this, it's actually quite useful to imagine $\mathcal{M}$ as being densely populated by these fundamental local observers, each carrying a conformal clock and a fixed spatial coordinate label x^i. Thus, the spatial coordinates in the synchronous gauge act as Lagrangian coordinates by comoving with the natural flow of the fluid (the resulting fluid, that is, as opposed to comoving with a particular component). Because of this, unlike what happens in $\overline{\mathcal{M}}$, the comoving coordinate lines become highly deformed as the fluid evolves. Given enough time and the threads will intersect with each other and form inevitable coordinate singularities, a process known as caustic formation. However, this flaw of the synchronous gauge is only noticeable when perturbations grow large enough. Hence, the synchronous gauge is perfectly adequate for linear theory where perturbations are always weak enough and caustics never form.

Unfortunately, the synchronous gauge conditions do not fix the gauge entirely. There is still some residual freedom that can obscure the physical interpretation of perturbations, especially on scales larger than the Hubble scale. (On the other hand, well inside the 'horizon', all gauges are virtually the same [45] given that things are mostly flat.) This is simple to understand if you realize that we're still free to adjust the initial settings of the comoving clocks and the x^i labels carried by the fundamental observers. We can show this explicitly by modifying the initial gauge in the following way [91]

$$\tilde{\eta} = \eta + \frac{\alpha}{a}\,, \qquad \tilde{x}^i = x^i + \partial^i\beta \int \frac{d\eta}{a} + \epsilon^i\,, \tag{3.8}$$

where the scalars α and β and the vector ϵ^i all depend solely on space coordinates (in other words, 3-fields in Σ) and $\nabla \cdot \boldsymbol{\epsilon} = 0$; the new gauge will still be a synchronous

gauge (i.e. $\tilde{\psi} = \tilde{\omega}_i = 0$) but now with different perturbations. This means that solutions to the linearized Einstein equations are still contaminated by physically irrelevant gauge modes, an annoying fact that substantially complicates the process of setting initial conditions. Not only this, but numerically we must also ensure that the gauge modes do not swamp the physical ones, thereby causing significant roundoff error in the solutions we care. The remedy is to place additional conditions on the gauge so as to fix it completely, a process Lyth & Stewart call 'dropping the gauge modes' [97]. One simple way to achieve this is to force the synchronous gauge to 'coincide' with the background FRW gauge far outside the Hubble horizon. Conceivably, this can be done very early in the history of the Universe when $\overline{\mathcal{M}}$ and $\mathcal{M}$ are virtually the same manifold given the 'absence' of perturbations. For this reason, Veeraraghavan & Stebbins call the resulting gauge the 'initially unperturbed synchronous gauge' [89].

3.2.4 Perturbation Types

It is useful to classify perturbations *algebraically* according to their invariance proprieties in the η-slices of $\mathcal{M}$. Note that if we make a *spatial* only coordinate transformation then, in order to preserve ds^2, $\psi(\eta, \mathbf{x})$ and $\phi(\eta, \mathbf{x})$ have to transform as 3-scalars, $\omega_i(\eta, \mathbf{x})$ as a 3-vector and $h_{ij}(\eta, \mathbf{x})$ as a 3-tensor in Σ. Also note that because they are still 4-tensors in $\mathcal{M}$, the perturbation components have to be raised or lowered using the conformal spacial part of the 4-metric, i.e. $(1 - 2\phi)\delta_{ij} + 2h_{ij}$; however, given that we are only interested in the linear regime (meaning that the perturbations are small and that quadratic and higher order terms are to be neglected), the *effective* metric in the Σ slice reduces to the background metric $\bar{\sigma}_{ij} = \delta_{ij}$ in $\overline{\Sigma}$. Nevertheless, we must still use the full metric to raise and lower time indices.

It is equally useful to look at perturbations as 3-*fields* in Σ and decompose them into fundamental scalar, vector and tensor 'building block' modes. Recall that a vector field like ω^i can always be decomposed into longitudinal and transverse parts

(the so-called Helmholtz's Theorem [98])

$$\omega^i = \omega^i_{\parallel} + \omega^i_{\perp}\,, \tag{3.9}$$

where the longitudinal vector is curl-free ($\nabla \times \boldsymbol{\omega}_{\parallel} = 0$) and the transverse vector is divergence-free ($\nabla \cdot \boldsymbol{\omega}_{\perp} = 0$). Note that in a flat geometry such as ours, we'll have that $\nabla \cdot \boldsymbol{v} = \partial_i v^i$, $\nabla \times \boldsymbol{v} = \epsilon^{ijk}\partial_j v_k\, \boldsymbol{e}_i$ where ϵ^{ijk} is the Levi-Cevita tensor and $\nabla^2\zeta = \partial_i\partial^i\,\zeta$. Given that the curl of a gradient is always zero, we can write the longitudinal vector as the gradient of a scalar $\omega_{i\parallel} = \partial_i\lambda$ which for obvious reasons is called the *scalar* part of ω_i. The transverse part, on the other hand, can be written as the curl of some other vector $\omega^i_{\perp} = \epsilon^{ijk}\,\partial_j\xi_k$ (given that the divergence of a curl is always zero) and is called the *vector* part of ω^i. Clearly, the scalar λ represents one degree of freedom contained in ω^i meaning that ξ_i can only carry the remaining two (not three, as one might first think). Note that ξ_i is actually only well defined up to a gradient, i.e., we can always add the gradient of a scalar to ξ^i and the curl of the new vector field still gives $\omega^i_{\perp}$. This gradient can be interpreted as a physically irrelevant gauge degree of freedom and this is why ξ^i only carries two. If we really care, we can 'fix' the gauge by imposing an additional condition like $\partial_i\xi^i = 0$ which makes it transverse, for example. Incidentally, λ as a field is also only well defined up to a constant but constants obviously do not carry degrees of freedom so this is not a problem.

Similarly, we can also decompose the traceless symmetric tensor perturbation h_{ij} into scalar, vector and *tensor* modes (i.e., the one that cannot be assembled from scalars and vectors) also called, respectively, longitudinal, *solenoidal* and transverse parts

$$h^{ij} = h^{ij}_{\parallel} + h^{ij}_{\circlearrowleft} + h^{ij}_{\perp}\,, \tag{3.10}$$

where the transverse part is divergence-free $\partial_i h^{ij}_{\perp} = 0$, while the divergence of the solenoidal part is a transverse vector (and therefore divergence-free) $\partial_i\partial_j h^{ij}_{\circlearrowleft} = 0$ and the divergence of the longitudinal part is a longitudinal vector (and therefore curl-free) $\epsilon^{jkl}\,\partial_k\partial_i {h_{\parallel}}^i{}_j = 0$. (By the way, not everybody uses the labels solenoidal and transverse in the same manner as we did here; we are following the same nomenclature Carroll does [34], but Bertschinger [91], for example, uses them in a slightly

different manner.) This means that we can write the longitudinal and transverse parts at the expense of a scalar field θ and a transverse vector $\zeta_\perp^i$ like so

$$h_{\parallel ij} = \left[\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right]\theta\,, \tag{3.11}$$

$$h_{\circlearrowleft ij} = \partial_{(i}\zeta_{\perp\, j)}\,, \tag{3.12}$$

where the parenthesis denote symmetrization. Thus, the longitudinal part carries one degree of freedom while the solenoidal part carries two and the tensor part the remaining two, adding up to the five in h_{ij}. Note that the tensor part cannot be decomposed further and is actually gauge-invariant [33].

Why is this field decomposition in Σ useful at all? First and foremost, scalar, vector and tensor modes represent distinct physical phenomena: Scalar perturbations, characterized by $(\psi, \phi, \lambda, \theta)$, describe spatial density fluctuations and because they exhibit gravitational instability, they may lead to the formation of structure. On the other hand, vector perturbations, represented by $(\boldsymbol{\xi}_\perp, \boldsymbol{\zeta}_\perp)$, are related to rotational motions of the fluid. However, in the absence of source terms, they have the tendency to decay rather quickly and therefore are not normally very interesting for Cosmology. As for tensor modes $(h_\perp^{ij})$, they represent gravitational waves (which can be interpreted as the *physical* gauge-invariant degrees of freedom of the gravitational field itself). For our part, we'll only be interested in scalar perturbations. Second, this spatial decomposition is hardly confined to perturbations alone; we can easily extend it to the Einstein and energy-momentum tensors, for example. It turns out that in the linear regime, the scalar, vector and tensor modes all decouple from each other and therefore have separate evolutions. Finally, this classification is particularly useful for identifying and dealing with the four unphysical gauge modes (in the form of two scalars and one transverse vector) that can plague the field equations. For example, we can see why the synchronous gauge conditions $\psi = \omega_i = 0$ don't fix the gauge entirely; the condition $\omega_i = 0$ does not completely specify λ or ξ^k which, in turn, are related to the β and ϵ^i in the coordinate transformation (3.8).

3.2.5 Perturbed Stress-Energy Tensor

We are going to assume that the perturbed fluid in $\mathcal{M}$ can be treated as a perfect fluid also. Obviously, there are periods in the evolution of the Universe where this approximation is not valid and viscosity, thermal conductivity and other such physical processes have to be included (like at the time of CMB formation), but they are unnecessary for our goals. Thus, we will take the energy-momentum tensor in $\mathcal{M}$ to be

$$T^{\mu}_{\ \nu} = (\varepsilon + p)u^{\mu}u_{\nu} + p\delta^{i}_{j} \tag{3.13}$$

(no anisotropic stresses) where $u^{\mu} = dx^{\mu}/d\tau$ is the 4-velocity, τ is the proper time and $g_{\mu\nu}u^{\mu}u^{\mu} = -1$. Let us first consider a local frame comoving with the real fluid; this frame is simply defined by the condition $u^{i} = 0$. Then, the normalization of the 4-velocity implies that $u^{0} = a^{-1}(1-\psi)$ to first order in ψ and from $u_{\mu} = g_{\mu\nu}u^{\nu}$ we obtain, also to first order, that $u_{0} = -a(1+\psi)$ and $u_{i} = a\omega_{i}$. In the synchronous gauge, this comoving frame is actually orthogonal and *inertial*. Also carefully note that by construction, it is only the resulting fluid (as a whole) that does not move in these frames; the individual components (if any) *do*. This is because the local synchronous gauge observers follow the local 'center-of-mass' of the fluid, not the individual components. Thus, in general, we'll need to include the motion of each fluid component, even in the (comoving) synchronous gauge.

It is actually easier to work out the general case for a single fluid first, and then adapt to our case. Hence, in an arbitrary gauge, the fluid velocity field can be written as $u^{\mu} = u^{0}(1, v^{i})$ where $v^{i} = dx^{i}/d\eta$ is the *coordinate* (not proper) 3-velocity of the fluid in the conformal Σ and $u^{0} = d\eta/d\tau$. Again, from the normalization condition we obtain that

$$u^{0} = \frac{1}{a\sqrt{1-v^{2}}}\left[1 - \frac{\psi - \omega_{i}v^{i} + \phi v^{2} - h_{ij}v^{i}v^{j}}{1-v^{2}}\right], \tag{3.14}$$

where $v^{2} = \delta_{ij}v^{i}v^{j}$. If we assume that quadratic terms in v can be neglected (this basically means that the fluid elements are non-relativistic), a distinct hypothesis from the weak field approximation, then we find that $u^{0} = a^{-1}(1-\psi)$, $u^{i} = a^{-1}v^{i}$,

and again from $u_\mu = g_{\mu\nu}u^\nu$ that $u_0 = -a(1+\psi)$ and $u_i = a(v_i + \omega_i)$. If we now write $u^\mu = \bar{u}^\mu + \delta u^\mu$ and remember that in the background $\bar{u}^\mu = (a^{-1}, 0, 0, 0)$, then the velocity perturbations will be $\delta u^0 = -\psi$, $\delta u^i = u^i$, $\delta u_0 = -a\psi$ and $\delta u_i = u_i$. If we now expand the stress-energy tensor (3.13) into a background part plus weak perturbations

$$\begin{aligned} T^\mu_{\ \nu} &= \bar{T}^\mu_{\ \nu} + \delta T^\mu_{\ \nu}\,, \\ &= \bar{T}^\mu_{\ \nu} + (\bar{\varepsilon} + \bar{p})\big(\bar{u}^\mu \delta u_\nu + \bar{u}_\nu \delta u^\mu\big) + (\delta\varepsilon + \delta p)\,\bar{u}^\mu \bar{u}_\nu + \delta P\,\delta^\mu_\nu \end{aligned} \tag{3.15}$$

we find that

$$\begin{aligned} T^0_{\ 0} &= -(\bar{\varepsilon} + \delta\varepsilon)\,, \quad T^i_{\ 0} = -(\bar{\varepsilon} + \bar{p})v^i\,, \\ T^0_{\ i} &= (\bar{\varepsilon} + \bar{p})(v_i + \omega_i)\,, \qquad T^i_{\ j} = (\bar{p} + \delta p)\delta^i_j\,. \end{aligned} \tag{3.16}$$

These expressions apply individually to each fluid component *if* they are only minimally coupled. This is precisely our case; in our model, the baryonic component and the gCg are assumed to exchange energy and momentum solely through gravity. Note that because we're only interested in scalar perturbations, v_i for us will be a scalar vector and $\omega_i = 0$ because of the gauge we chose. For later convenience, we introduce here the following notations: $\delta = \delta\varepsilon/\bar{\varepsilon}$ will denote the density contrast, $\bar{\omega} = \bar{p}/\bar{\varepsilon}$, the zero-order equation of state in $\overline{\mathcal{M}}$ and $c_s^2 = \delta p/\delta\varepsilon$, the sound speed in Σ. Note that in the linear regime c_s^2 in $\mathcal{M}$ is approximately the same as $\bar{c}_s^2 = \dot{\bar{p}}/\dot{\bar{\varepsilon}}$, the zero-order sound speed in $\overline{\mathcal{M}}$.

Let us now discuss energy-momentum conservation $\nabla_\mu T^\mu_{\ \nu} = 0$ in terms of metric perturbations and fluid variables. Calculating the covariant derivatives is a straightforward process, albeit a tedious one and we simply present the result. Using the weak field approximation and assuming a non-relativistic fluid, one finds in an arbitrary gauge that [91]

$$\dot{\varepsilon} + 3(\mathcal{H} - \dot{\phi})(\varepsilon + p) + \boldsymbol{\nabla}\cdot[(\varepsilon + p)\boldsymbol{v}] = 0\,, \tag{3.17}$$

and

$$\frac{\partial}{\partial\eta}\left[(\varepsilon+p)(\boldsymbol{v}+\boldsymbol{\omega})\right]+4\mathcal{H}(\varepsilon+p)(\boldsymbol{v}+\boldsymbol{\omega})+\boldsymbol{\nabla}p+(\varepsilon+p)\boldsymbol{\nabla}\psi=0 \tag{3.18}$$

which can be interpreted as the relativistic analogues of the continuity and Navies-Stokes equations of Newtonian fluids. Once again, these expressions apply individually if the various components are only minimally coupled; also, in our case $\omega_i = 0$ and v^i is a longitudinal vector. We can easily separate (3.17) and (3.18) into unperturbed and perturbed parts. The unperturbed parts, of course, 'live' in $\overline{\mathcal{M}}$ and are sometimes called Arnowitt, Deser & Misner (ADM) energy-momentum constrains [40]; we calculated them in Chapter 1. Then, in the synchronous gauge, the scalar modes turn out to be

$$\dot{\delta}+3\mathcal{H}(\bar{c}_s^2-\bar{\omega})\delta+(1+\bar{\omega})(\theta-3\dot{\phi})=0\,, \tag{3.19}$$

and

$$\dot{\theta}+\mathcal{H}(1-3\bar{c}_s^2)\theta+\frac{\bar{c}_s^2}{1+\bar{\omega}}\nabla^2\delta=0\,, \tag{3.20}$$

where $\theta = \boldsymbol{\nabla}\cdot\boldsymbol{v}$; these are the same expressions as in Veeraraghavan & Stebbins [89], save for the fact that their h_{ij} is not trace-free as ours is and so to compare we need to use $h = -6\phi$. Alas, we have three unknowns (δ, θ, ϕ) but only two equations for each fluid (obviously, the background quantities do not count as unknowns as they have to be specified from the outset).

3.2.6 Perturbed Field Equations

To close (3.19) and (3.20), we must bring the scalar modes from the Einstein field equations in $\mathcal{M}$, i.e., $G^{\mu}{}_{\mu} = 8\pi G T^{\mu}{}_{\mu}$, which is even a more tedious job than before so again we merely write down the result. As always, using the background ADM constraints in the unperturbed parts, one finds that

$$\ddot{\phi}+\mathcal{H}\dot{\phi}-\frac{1}{2}\mathcal{H}^2\sum\nolimits_i(1+3\bar{c}_{s,i}^2)\Omega_i\,\delta_i=0\,, \tag{3.21}$$

where the sum goes over all fluid components in the model. Again, to compare with [89] just use $h = -6\phi$. With the above equation, we are finally in a position, to

study the evolution of scalar perturbations in a model of baryons plus a gCg. But before we do so, we briefly revisit the problem of the equivalence between ΛCDM and the $\alpha = 0$ generalized Chaplygin gas.

3.2.7 ΛCDM linear equivalence to the $\alpha = 0$ gCg

In §2.2, we already succeeded in showing that in the limit $\alpha \to 0$, the gCg is totally equivalent (as far as gravity is concerned) to ΛCDM. In the past, this assertion was mostly taken for granted without any formal demonstration. Such need only arose after [61] appeared. In it, Fabris, Gonçalves & de Sá Ribeiro made the surprising claim that the linear evolution of density perturbations was actually different in each case. This section is a simple rehash of [62] where we proved that the gCg is indeed equivalent to ΛCDM. Let us start by calling the solo gCg with $\alpha = 0$, model I, and ΛCDM, model II. Model I has a zero background sound speed as a result of its constant pressure, and so (3.20) reduces to $\theta = 0$ all the times. This was to be expected since the synchronous gauge is a comoving gauge and this model only has one fluid. Then, from (3.19) we obtain

$$3\dot{\phi} = \frac{\dot{\delta}_{\rm I} - 3\,\bar{\omega}_{\rm I}\mathcal{H}\,\delta_{\rm I}}{1+\bar{\omega}_{\rm I}}\,, \tag{3.22}$$

which combined with the fact that $\dot{\bar{\omega}}_{\rm I}/\bar{\omega}_{\rm I} = 3\mathcal{H}(1+\bar{\omega}_{\rm I})$ can be recast as $3\dot{\phi} = \dot{\delta}_\star$ where $\delta_\star = \delta_{\rm I}/(1+\bar{\omega}_{\rm I})$. It is now straightforward to show from (3.21) that the density perturbations are given by

$$\ddot{\delta}_\star + \mathcal{H}\dot{\delta}_\star - \frac{3}{2}\mathcal{H}^2(1+\bar{\omega}_{\rm I})\delta_* = 0\,. \tag{3.23}$$

On the other hand, model II has two components and both have zero background sound speeds. Hence, (3.20) also implies $\theta = 0$ for each. This means that the synchronous gauge is comoving at the same time with Λ *and* CDM. How can this be? This is actually quite simple to understand when one realize that because Λ is always constant, it also looks stationary in the inertial gauge comoving with the CDM component. Let us now turn to the evolution of perturbations in the dark matter component (by definition Λ cannot be perturbed): Using (3.21) plus the fact

that $3\dot{\phi} = \dot{\delta}_{\rm CDM}$ and also the relation $\Omega_{\rm CDM} = \bar{\varepsilon}_{\rm CDM}/(\bar{\varepsilon}_{\rm CDM} + \bar{\varepsilon}_\Lambda) = 1 + \bar{\omega}_{\rm II}$ it is easy to find that

$$\ddot{\delta}_{\rm CDM} + \mathcal{H}\,\dot{\delta}_{\rm CDM} - \frac{3}{2}\mathcal{H}^2(1+\bar{\omega}_{\rm II})\delta_{\rm CDM} = 0. \tag{3.24}$$

It follows that $\delta_{\rm I}$ and

$$\delta_{\rm II} \equiv \frac{\delta\varepsilon_{\rm II}}{\bar{\varepsilon}_{\rm II}} = \frac{\delta_{\rm CDM}}{1 + \bar{\varepsilon}_\Lambda/\bar{\varepsilon}_{\rm CDM}} = (1+\bar{\omega}_{\rm II})\delta_{\rm CDM}\,, \tag{3.25}$$

will have the same evolution if we identify $\bar{\omega}_{\rm II}$ with $\bar{\omega}_{\rm I}$. Fabris, Gonçalves & de Sá Ribeiro based their claim on the different evolutions of $\delta_{\rm CDM}$ and $\delta_{\rm I}$ but as we've just seen, they are not the correct variables to compare. As a matter of fact, if we wrote the evolution of perturbations in Fourier space for models I and II, we would find them to be independent of the wave number k, meaning that all scales evolve the same way. Considering that a density perturbation with infinite wavelength (that is, a uniform perturbation) evolves as a uniform $\overline{\mathcal{M}}$ does, then it is not at all surprising that ΛCDM and the gCg with $\alpha = 0$ are linearly equivalent given that both have the same zero-order background and all scales evolve the same way.

3.3 Baryons + gCg — The Model

The model we want to study includes baryons on top of a gCg. The hope is that the baryonic component may carry the structure that gets wiped (or exponentially enhanced) in the single gCg model when $|\alpha| > 10^{-5}$ [86]. Here, we'll treat baryons as if ordinary CDM, in other words, as a pressureless non-relativistic fluid with null background sound speed; this has the immediate consequence that $\theta_b = 0$ at all times, meaning that the synchronous gauge will comove with them. Also, from (3.19), we find that $3\dot{\phi} = \dot{\delta}_b$. It follows that the linear evolution of scalar perturbations for our flat model is given by

$$\begin{aligned}
\ddot{\delta}_b + \mathcal{H}\dot{\delta}_b - \frac{3}{2}\mathcal{H}^2\left[\Omega_b\delta_b + (1-3\alpha\,\bar{\omega})\Omega_{\rm gCg}\,\delta\right] &= 0\,,\\
\dot{\delta} - 3\mathcal{H}(1+\alpha)\bar{\omega}\,\delta + (1+\bar{\omega})(\theta - \dot{\delta}_b) &= 0\,,\\
\dot{\theta} + \mathcal{H}(1+3\alpha\,\bar{\omega})\theta - \frac{\alpha\,\bar{\omega}}{1+\bar{\omega}}\nabla^2\delta &= 0\,.
\end{aligned} \tag{3.26}$$

where any perturbation variable without an index refers to the gCg. We still need to add the background Friedmann equation, of course. However, we only need to integrate it along with the others, if we insist on determining the evolution of perturbations as a function of the conformal time η. Hence, we can substantially decrease the numerical complexity of the problem if we only require the evolution of perturbations as a function of the scale factor a; actually $x = \ln a$ is better. Thus, we rewrite (3.26) in terms of x

$$\begin{aligned} \delta_b'' + (1+\xi)\delta_b' - \frac{3}{2}\left[\Omega_b\delta_b + (1-3\alpha\,\bar{\omega})\Omega_{\rm gCg}\,\delta\right] &= 0\,, \\ \delta' - 3(1+\alpha)\bar{\omega}\delta + (1+\bar{\omega})(\theta/\mathcal{H} - \delta_b') &= 0\,, \\ \theta' + (1+3\alpha\,\bar{\omega})\theta - \frac{\alpha\bar{\omega}}{(1+\bar{\omega})\mathcal{H}}\,\nabla^2\delta &= 0\,, \end{aligned} \tag{3.27}$$

where $\xi = \mathcal{H}'/\mathcal{H}$ and the prime stands for $' \equiv d/dx$. Confront this to result obtained by Sandvik et al [86] (in the same gauge as ours)

$$\delta'' + [2 + \xi - 3(2\bar{\omega} - \bar{c}_s^2)]\delta' + \frac{3}{2}[1 - 6\bar{c}_s^2 + 8\bar{\omega} - 3\bar{\omega}^2]\delta = \frac{\bar{c}_s^2}{\mathcal{H}^2}\nabla^2\delta\,, \tag{3.28}$$

for a single gCg model. There are a few differences worth emphasizing: First, (3.28) is only valid for subhorizon scales ($|\mathbf{k}| \gg H$) while ours is valid for any scale. Second, the derivation leading up to (3.28) assumes from the start an isentropic fluid [99], ours doesn't.

It is quite useful to separate each perturbation variable into a time part and a space part, e.g. $\delta_b = \delta_b(\eta, \mathbf{k})f(\mathbf{k}, \mathbf{x})$ where $\mathbf{k}$ is a called a separation constant. (This procedure is sometimes justified on the basis that, to first-order, perturbations see Σ as a homogeneous and isotropic $\overline{\Sigma}$. However, this separation *is* always possible, even in the non-linear regime, which is not to say that it is always useful.) As a linear system, the general solution will be a superposition of these modes. For f, it is customary to use the harmonic solutions of the scalar Helmholtz equation $\nabla^2 f - k^2 f = 0$, since they form a complete spatial basis [98]; in the case of a flat cartesian $\overline{\Sigma}$, they are simply ordinary plane waves $\exp(i\mathbf{k}\cdot\mathbf{x})$. Thus, in (3.27), every perturbation variable reduces to a Fourier amplitude (the spacial parts cancels out) and $\nabla^2 \equiv -k^2$. We end up with three equations for three unknowns: δ_k^b, δ_k and

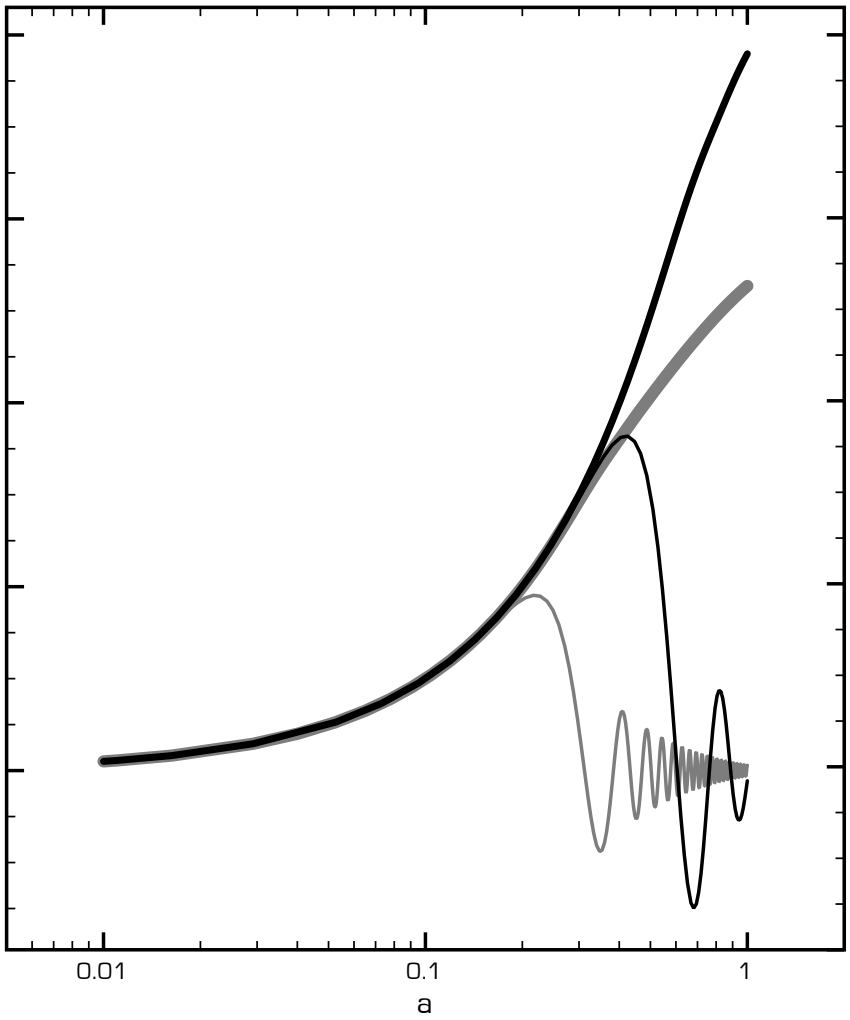


Figure 3.2: The evolution of perturbations in a model of baryons (strong lines) plus a gCg (lighter lines) for two scales: $k = 0.01\,h\,\mathrm{Mpc}^{-1}$ (black) and $k = 0.1\,h\,\mathrm{Mpc}^{-1}$ (gray). The baryons and the gCg grow in unison for a while but 'decouple' as soon as the gCg background starts transitioning to a cosmological constant.

θ_k. Given $\bar{\omega}$ and $\mathcal{H}$ (and ξ, Ω_{gCg}) as functions of x, we can easily transform this system into four first order differential equations and integrate it using any standard Runge-Kutta method. In particular, we have used the GNU Scientific Library (GSL 1.10) for this purpose (see the Appendix for the code). Fig. 3.3 illustrates typical solutions. (We have checked numerically that the solutions of (3.28) coincide with ours for the gCg component inside the Hubble scale.) As with any other differential

system, however, we still need to specify the initial conditions to evolve but before we engage this discussion, we briefly review the statistics of structure formation necessary to bridge theory with observations (for a more comprehensive account see Joyce's et al [100], for example).

3.3.1 Statistics of Scalar Perturbations

Let us consider a cubic box of size $L \gg l_s$, where l_s represents the maximum scale where significant structure still exists. This volume can be thought as a fair sample of the Universe. It is customary to decompose the energy distribution (the contrast, really) inside this box in a Fourier series

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{x}) = \sum_{\mathbf{k}} \delta^*_{\mathbf{k}} \exp(-i\mathbf{k}\cdot\mathbf{x}) \,, \tag{3.29}$$

for each component, where $\mathbf{k} = 2\pi\mathbf{n}/L$ is a discrete wavenumber and $\mathbf{n} \in \mathbb{Z}^3$. (Decomposing the energy distribution this way, however, automatically implies that $\delta(\mathbf{x} + L\mathbf{n}) = \delta(\mathbf{x})$ for any $\mathbf{n}$, i.e. it imposes an artificial periodicity outside the box; this periodicity, on the other hand, can be safely ignored if the relevant scales are already encompassed by the box.) It is straightforward to show that the Fourier amplitudes of this expansion are given by

$$\delta_{\mathbf{k}} = \frac{1}{V} \int_V \delta(\mathbf{x}) \exp(-i\mathbf{k}\cdot\mathbf{x})\, d\mathbf{x} \tag{3.30}$$

where $V = L^3$. It follows that $\delta^*_{\mathbf{k}} = \delta_{-\mathbf{k}}$ since $\delta(\mathbf{x})$ is a real function and also $\delta_{\mathbf{0}} = 0$. (Note, however, that in general we never actually measure the energy distribution with infinite resolution; thus, in practice, the sum in (3.29) is really a finite sum for scales above the sampling scale. This finite resolution inevitably introduces some aliasing, but most of the time we can get away with it, if we only care about scales much bigger that the resolution scale.) It is many times useful to take the limit of a large box

$$\left(\frac{2\pi}{L}\right)^3 \sum_{\mathbf{k}} \to \int d\mathbf{k} \,, \tag{3.31}$$

and convert the Fourier series into a Fourier integral

$$\delta(\mathbf{x}) = \frac{V}{(2\pi)^3} \int \delta_{\mathbf{k}} \, \exp(i\mathbf{k} \cdot \mathbf{x}) \, d\mathbf{k} \,. \tag{3.32}$$

Note that by a 'large box', people usually mean 'in the limit $L \to \infty$'. Nevertheless, an infinite sized box is not really necessary to justify (3.31); we can jump to the continuum with any finite sized box, as long as its size is much bigger than the sampling scale $L \gg \Delta$. (Be warned that the numerical prefactors in the Fourier pair (3.30) and (3.32) vary substantially in the literature as many authors rescale the amplitudes as they see fit. This can be quite annoying but is physically irrelevant. A popular rescaling is $\delta_{\mathbf{k}} \to (2\pi)^{3/2}\delta_{\mathbf{k}}/V$, which makes the Fourier basis orthonormal instead of just orthogonal.)

We can, of course, decompose any number of boxes 'scattered' all over the Universe this way. Obviously, each will yield a different set of amplitudes $\{\delta_{\mathbf{k}}\}$. In fact, it is very useful to interpret a collection of these boxes as an ensemble of statistical realizations; this means treating the amplitudes (or $\delta(\mathbf{x})$, for that matter) as random variables subject to a certain distribution. For example, if the $\{\delta_{\mathbf{k}}\}$ are all completely random, then the perturbation field will have Gaussian statistics. This is easy to understand: From (3.29), we see that $\delta(\mathbf{x})$ can be interpreted as the sum of several random variables; if these are mostly uncorrelated with each other, the central limit theorem guaranties that $\delta(\mathbf{x})$ will be Gaussian distributed. It turns out that Gaussian statistics are strongly motivated by inflation (although some deviations are expected).

To characterize the actual field statistics, we need to determine the several moments (or cumulants) of the distribution. The simplest way to do this is to consider a very large box. Assuming this über-box encompasses several statistically homogeneous realizations, we may invoke the ergodic hypothesis (this step involves many subtleties we won't go into here) and use spatial averages $\langle X \rangle$ as a substitute for

stochastic averages $E[X]$. It follows from this that $\langle\delta(\mathbf{x})\rangle = 0$ and

$$\begin{aligned}\sigma^2 \equiv \langle\delta^2(\mathbf{x})\rangle &= \frac{1}{V}\int \delta^2(\mathbf{x})\, d\mathbf{x} = V\int d\mathbf{x}\int \frac{d\mathbf{k}}{(2\pi)^3}\frac{d\mathbf{q}}{(2\pi)^3}\,\delta_{\mathbf{k}}\delta^*_{\mathbf{q}}\,\exp[i(\mathbf{k}-\mathbf{q})\cdot\mathbf{x}] \\ &= \frac{V}{(2\pi)^3}\int d\mathbf{k}\,|\delta_{\mathbf{k}}|^2 = \frac{V}{2\pi^2}\int_0^\infty |\delta_k^2|\,k^2 dk\,, \qquad (3.33)\end{aligned}$$

where in the last equality, we have assumed statistical isotropy (which seems to be a good approximation of the Universe, although we should keep an open mind). Higher moments $\langle\delta^n(\mathbf{x})\rangle$ are calculated in a similar fashion. Here, the quantity $P(k) \equiv |\delta_k|^2$ is usually called the *power spectrum* of the field; it measure the contribution of each scale to the overall dispersion. (For fields other than matter fields, however, a power spectrum per unit logarithm in k is normally preferred.) Most inflationary models predict a primordial power law $|\delta_k|^2 \propto k^n$, with current observations favoring $n \simeq 1$ (if n is exactly one, the spectrum is called Harrison-Zel'dovich).

Note, however, that the moments of a distribution may not converge, unless the power spectrum vanishes rapidly enough at the limit of either large or small scales. At large scales, we expect the spectrum to rapidly vanish (due to the Cosmological Principle), but not for small scales. For small scales, we have a lot of structure and so (3.33), for example, may not converge. To fix this, it is necessary to low-pass filter the density field: In Fourier space this is accomplished straightforwardly by introducing a new set of amplitudes

$$\delta_R(k) = \mathcal{W}(k,R)\delta_k\,, \qquad (3.34)$$

such that $\delta_R(k) \simeq \delta_k$ for $k^{-1} > R$ and $\delta_R(k) \simeq 0$ for $k^{-1} < R$, where R is some smoothing scale. This effectively wipes out most structures smaller than the smoothing scale while, at the same time, preserving larger ones. The variance of this filtered field

$$\delta_R(\mathbf{x}) = \int W(\mathbf{x}',R)\delta(\mathbf{x}-\mathbf{x}')d\mathbf{x}\,, \qquad (3.35)$$

is now given by

$$\sigma_R^2 \equiv \langle\delta_R^2(\mathbf{x})\rangle = \frac{V}{2\pi^2}\int_0^\infty P(k)\mathcal{W}^2(k,R)k^2 dk < \sigma^2\,, \qquad (3.36)$$

and can be made to converge. In (3.35), we have used the convolution theorem $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$ where $W = \mathcal{F}^{-1}(\mathcal{W})$ is called a window function. The actual filter doesn't matter very much; an ordinary step function (top hat) will do just fine. It also follows from the autocorrelation theorem $\mathcal{F}(f \star f) = |\mathcal{F}(f)|^2$ (also called Wiener-Khinchin theorem), that the power spectrum $|\delta_k|^2$ is the the Fourier transform of the so-called two-point correlation function $\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle$. This function measures how fluctuations in the density field are correlated (co-vary) with each other; thus, the typical size of overdense regions is determined by the first zero of this function. It can also be interpreted as the excess probability (over random) of finding two objects of the same class (say, galaxies) separated by $\mathbf{r}$. (Using galaxy surveys to estimate $\xi(\mathbf{r})$, however, is a complicated affair as galaxies are believed to form more frequently in high density regions than others; as a result, they may not constitute an unbiased sample of the matter distribution.)

Now, the shape of the primordial power spectrum doesn't survive unaltered to the present day because of the different way perturbations grow inside and outside the 'horizon'. The exact processing depends on the cosmological parameters (Ω, h) and the type (hot or cold) of non-baryonic dark matter. For HDM, fluctuations on scales smaller than the free streaming scale (typically the size of a cluster of galaxies) are wiped out and the final spectrum gets a well defined cutoff for small scales. This is not observed. On the other hand, for CDM, the primordial spectrum gets 'bent' at around $\lambda_{eq} = 14(\Omega_0 h^2)^{-1}\,\mathrm{Mpc}$, the size of the horizon at the time of matter-radiation equality. This happens because scales $\lambda < \lambda_{eq}$ (which enter the 'horizon' prior to equality) suffer a stagnation period due to the Meszaros effect [36] (basically, the background is expanding too rapidly for matter perturbations to grow inside the Hubble scale) when they enter. Scales outside, however, do not experience this stagnation period as they enter the horizon in the matter era. As a result, the spectrum preserves its form for large scales but losses power on smaller scales (but not nearly as much as in the HDM case). Writing $\delta_k(t_f) = T\delta_k(t_i)$, where T is called the transfer function, it follows that

$$T \simeq \begin{cases} 1 & k < k_{eq}\,, \\ (k_{eq}/k)^2 & k > k_{eq}\,. \end{cases} \tag{3.37}$$

Many accurate transfer function (obtained numerically) have been tabulated in the literature. A very popular parameterization is the BBKS transfer function due to Bardeen et al [101]

$$T = \frac{\ln(1+\epsilon_0\zeta)}{\epsilon_0\zeta}\left(\sum_i (\epsilon_i\zeta)^i\right)^{-1/4} \tag{3.38}$$

where $\epsilon = [2.34, 3.89, 16.1, 5.46, 6.71]$. Here, $[k] = h\,\mathrm{Mpc}^{-1}$ and $\zeta = k/\Gamma$, where $\Gamma = \Omega^0_{\mathrm{CDM}}\,h$ is called the shape parameter (this parameter controls where the bend occurs, hence the name). Recent work by [84] using data from the 2dF 100k galaxy survey has constrained Γ to be of the order of ~ 0.2, in agreement with preliminary SDSS results [102]. Writing the shape parameter as we did, however, ignores the fact that baryons are strongly coupled to photons (even in the matter era) for a long time (until recombination), which effectively prevents the growth of perturbations in the baryonic component on scales smaller than the sound horizon. To take their effect into account, we must apply Sugiyama's [103] empirical correction to the shape parameter

$$\Gamma_b = \Gamma \exp\left(-\Omega^0_b - \sqrt{2h}\,\Omega^0_b/\Omega^0_{\mathrm{CDM}}\right)\,. \tag{3.39}$$

Nowadays, CMBFAST [104] and, more recently, CMBEASY [105] are frequently used to obtain even more accurate transfer functions. Nonetheless, (3.38) coupled with Sugiyama's shape correction is already good enough for our purposes. In our model, however, we don't really have CDM; instead, we have a gCg component firmly behaving as CDM at equality (and also for a long time after). It is simple to show, as we did initially in [65], that the presence of this gCg only changes the shape parameter to $\widetilde{\Gamma} = \Omega^{0\,*}_{\mathrm{CDM}}\,h$ where

$$\Omega^{0\,*}_{\mathrm{CDM}} = \Omega^0_{\mathrm{CDM}} + \Omega^0_{\mathrm{gCg}}(1-\mathcal{A})^{1/1+\alpha}\,. \tag{3.40}$$

3.3.2 Results

We have taken a primordial scale-invariant Harrison-Zel'dovich spectrum and processed it using a BBKS transfer function (adapted for the gCg and corrected for

baryons). After equality and for a long time (essentially while the gCg behaves as matter) the shape of this power spectrum is preserved (for linear scales, that is). However, as the gCg transitions to a cosmological constant, linear perturbations are progressively erased by the ever increasing expansion rate of the background (in a way that is very similar to the Meszaros effect back in the radiation era). Nevertheless, linear perturbations in the baryonic component can still grow for a while (but not forever, of course). Hence, the question is, can baryons carry enough large scale structure to account for what is observe today? To find out, we evolved normalized initial conditions $[\delta_b, \delta_b', \delta_{cg}, \theta_{cg}]_0 = [1, 1, 1, 0]$ (recall that in the matter era $\delta \propto a$ either for baryons and the gCg, implying $\delta' \propto a$) from $z = 100$ to the present day (as in [86]) and obtained corresponding transfer functions to further process the spectrum that emerges from equality. But quite unlike [86], our transfer functions now come from the baryonic component, not the gCg, thus avoiding the violent oscillations (or exponential blowup) that are so sensitive to α.

We have used the matter power spectrum obtained by Tegmark et al [106] from the 2dF 100k redshift survey [84] to constrain the $(\alpha, \mathcal{A})$ parameter space assuming only reasonable priors coming from from WMAP [43], namely $\Omega_b^0 = 0.044$ and an Hubble constant $h = 0.71$. (Note that at the time of recombination, the Chaplygin gas would still firmly behave as CDM. Therefore, standard small scale CMB results are to be expected when one identifies $\Omega^0_{\rm CDM}$ with $\Omega^{0\,*}_{\rm CDM}$. The results could conceivably differ on very large scales, though here they would be competing against cosmic variance.) Also, as in Sandvik's work, we have discarded any 2dF data over $k > 0.3\,h\,{\rm Mpc}^{-1}$ so as to stay firmly grounded in the linear regime, where our analysis holds. Specifically, we have evaluated a $500 \times 100 \times 100$ data grid for A (the primordial spectrum amplitude), $\mathcal{A}$ and α, with $0 < \alpha < 1$ and $0 < \mathcal{A} < 1$. Corresponding probabilities were found and posteriorly summed over A. The resulting confidence regions are shown in Fig. 3.3, where two disjoint regions can be seen: one prominent, the other small (around $\alpha \sim 0$ and $\mathcal{A} \simeq 0.8$). We have also displayed the region of the parameter space corresponding to a value of the shape parameter of $\Gamma_b = 0.2 \pm 0.03$, as per [102]. The small area in the figure correspond to the ΛCDM limit of the gCg. On the other hand, there is also an entirely disjoint region at a

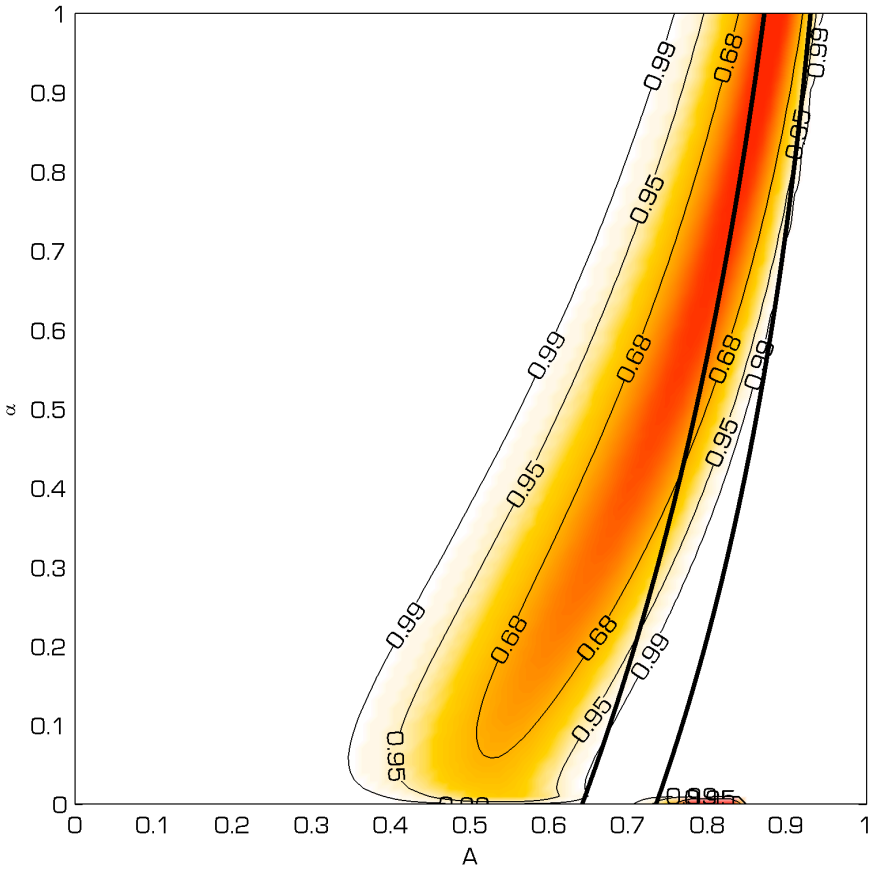


Figure 3.3: 68%, 95% and 99% likelihood contours in the $(\alpha, \mathcal{A})$ parameter space for a model of baryons plus a (generalized) Chaplygin gas, coming from the 2dF mass power spectrum. Note the minute ΛCDM region near $\alpha \sim 0$. The zone inside the solid lines corresponds to $\Gamma_b = 0.2 \pm 0.03$ [102].

very high confidence level. So, in fact, we have provided explicit evidence of a (generalized) Chaplygin gas not having to behave as ΛCDM in order to reproduce 2dF large-scale structure. This is an important result, no doubt, and provides some relief against Sandvik's constraint. Our result, however, still has to be jointly analyzed with other results. Fig. 3.4 shows the probability contours when we add the old 92 supernovae constraints from Chapter 2, while Fig. 3.5 shows the updated version

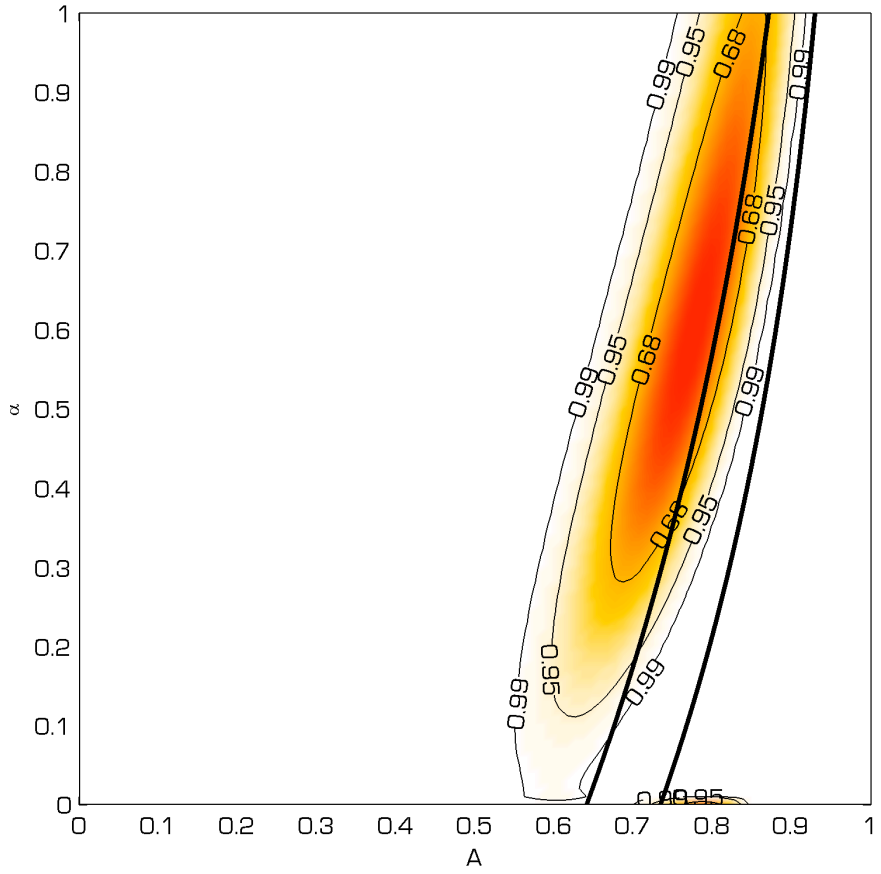


Figure 3.4: 68%, 95% and 99% likelihood contours resulting from a joint analysis of large-scale structure and type Ia supernovae (old sample) for a model of baryons plus a (generalized) Chaplygin gas.

using the new sample. With the updated version, we see a clear separation between the two areas that wasn't present in the old result. On the other hand, CMB constraints on the gCg have been obtained in a variety of papers [87, 71, 107, 108] and more recently [77]. Here, we only quote their main result: At 3σ, $\alpha < 0.2$. If follows, from Fig. 3.5, that adding baryons no longer significantly alleviates Sandvik's constraint on α (which was our strong conviction at the time we wrote [109]) and the gCg is indeed forced to behave very closely to a ΛCDM model.

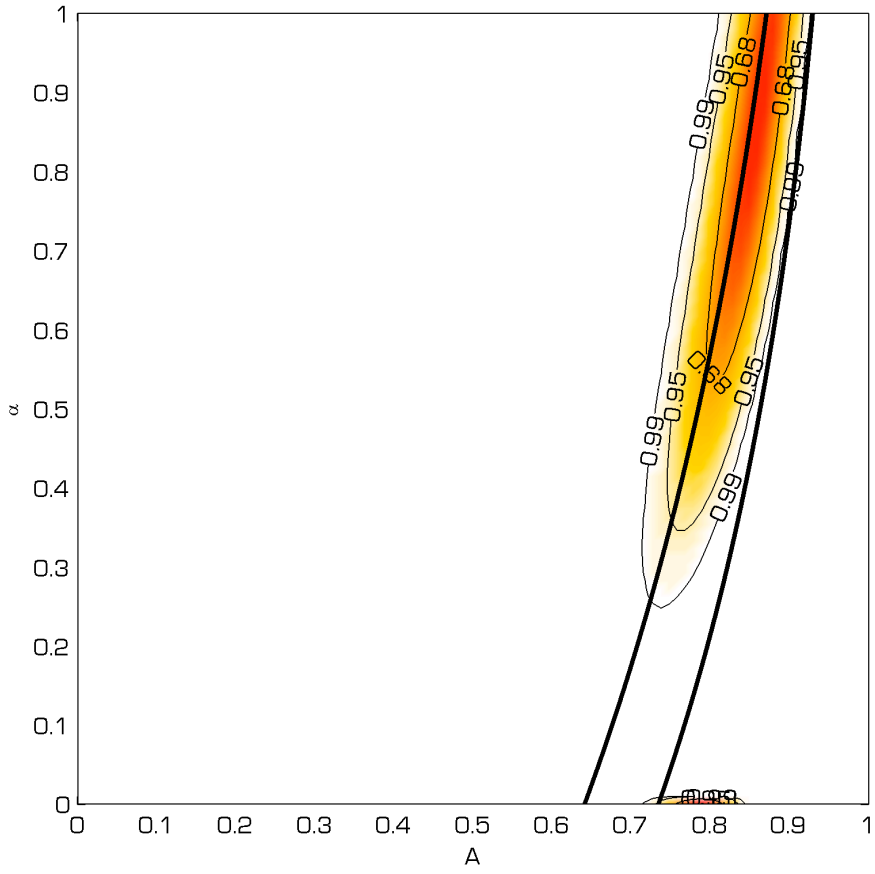


Figure 3.5: 68%, 95% and 99% likelihood contours resulting from a joint analysis of large-scale structure and type Ia supernovae (new sample) for a model of baryons plus a (generalized) Chaplygin gas. Note the separation between the two disjoint regions that wasn't present with the old sample. If we now take into consideration that CMB constraints require an $\alpha < 0.2$, we conclude that adding baryons is no longer able to alleviate Sandvik's constraint.

3.4 Conclusions

We have shown that baryons play an important role in the context of unified dark energy models. In a single gCg model, perturbations vanish completely as the gCg

transition to a cosmological constant, unless α is extraordinarily close to zero, in other words, to a ΛCDM model. Baryons, on the other hand, if present can carry the structure that gCg cannot; if we include them, α is no longer forced to be virtually zero; in fact, the entire spectrum $0 < \alpha < 1$ becomes possible. This, in itself, is a significant result; however, when analyzed together with other results from supernovae and CMB tests, we conclude that α, unfortunately, still has to be very close to zero. At the time [109] was published, a non-zero (albeit small) α value was still consistent with observations. With the new supernovae data, this is no longer possible and the gCg is indeed forced to behave very closely to ΛCDM. This result has been frequently advertised as the end of UDE. To be fair, however, this isn't quite true for the simple fact that, as we have already seen, ΛCDM is virtually indistinguishable (as far as gravity is concerned) from a unified gCg with $\alpha = 0$. Finally, we wish to emphasize the fact that the constraints obtained here (and those of most other researchers, in fact) have been obtained assuming that linear theory is a valid approximation for large scales. As we will discuss in detail in the next chapter, it turns out that we must be very careful about this assumption in the context of all unified dark energy models. There are, in fact, many non-trivial subtleties involved that make these models substantially more complicated to analyze than what first meets the eye.

Chapter 4

UDE — Non-Linear Regime

UDE models based on the gCg have been extensively tested against a wide variety of observations including high-z supernovae [65, 66, 67, 68, 109, 69, 70, 71, 72, 73, 74, 75], lensing [110, 69, 111, 112], high precision CMB [87, 71, 107, 108] and LSS [113, 114, 65, 86, 109, 115, 71]. The latest all-encompassing effort can be found in [77]. These tests, however, for better or worse, all bear a common factor: they originate from the same perturbative treatment of the gCg. In a nutshell, a zero-order homogeneous gCg background $\overline{\mathcal{M}}$ is assumed and subsequently (weakly) perturbed. However, there is a potential *caveat* to this whole way of treating quartessence that is the subject of this chapter.

4.1 All Quiet on the Western Front?

To see why problems may be lurking around the corner, let us start by recalling that the real Universe (denoted throughout by manifold $\mathcal{M}$) displays hierarchical structures like stars, galaxies, clusters of galaxies, and so on; it is far from being smooth. The dynamics in $\mathcal{M}$ is assumed to be entirely described by General Relativity (or some modified version of), as gravity is the only relevant force at work on cosmological scales. In practice, however, due to the highly non-linear nature of the field

equations, it is virtually impossible to solve the dynamics completely (even numerically), unless in a handful of high symmetry situations (see, for instance, [116]). On the other hand, several observations indicate that the Universe looks increasingly smooth, as larger and larger scales are considered (typically over 100 Mpc). This *average* background (representing the *global* behavior of $\mathcal{M}$) is routinely idealized as a completely featureless manifold, hereafter denoted by manifold $\langle\mathcal{M}\rangle$. As already discussed in the first chapter, the kinematics in $\langle\mathcal{M}\rangle$ is essentially contained in the cosmological principle, but what about its dynamics?

This is actually a tricky question. Normally, it is assumed that General Relativity applies just as well in $\langle\mathcal{M}\rangle$ as it does in $\mathcal{M}$. Yet, when we average $\mathcal{M}$ to obtain the background, we're also averaging complex non-linear interactions. Averaging linear terms in the field equations is no big deal, but averaging non-linear terms *is*; this is because they introduce back-reaction terms and hence new dynamics. The common expectation, however, is that back-reactions are negligible on cosmological scales, so that $\langle\mathcal{M}\rangle$ and $\mathcal{M}$ have effectively the same dynamics. In what follows, $\overline{\mathcal{M}}$ will denote the background when these back-reactions are ignored, to distinguish it from the real $\langle\mathcal{M}\rangle$ that includes them. Note that the Standard Model is built on top of $\overline{\mathcal{M}}$. Proving if back-reactions are negligible, of course, is nearly impossible; even checking it numerically is extremely difficult. As a matter of fact, 'averaging', in General Relativity, remains largely an unresolved and complicated problem [117, 118]. Thus, it is quite possible that the true background may evolve differently from $\overline{\mathcal{M}}$. In other words, $\langle\mathcal{M}\rangle$ should really be seen as a 'corrected' $\overline{\mathcal{M}}$ somehow. Recently, it has been speculated that such corrections might naturally give rise to the current acceleration of the Universe (as opposed to some weird dark energy component); see [119] and references therein. Appealing as this may sound, however, these corrections turn out to be extremely hard to quantify; currently, there is no convincing argument showing that they would be large enough for the job. The few times that such an analysis has been attempted (by studying high symmetry configurations such as closely spaced sheets of matter separated by voids, arguably not even a very good approximation of the Universe), have shown them to be small and hence incapable of producing the observed acceleration [120, 121]. Our personal opinion, is that

'averaging' alone will not explain the mystery behind dark energy.

Second, although in general we cannot determine the full evolution of $\mathcal{M}$, we can still solve the dynamics of 'small' perturbations. Indeed, when perturbations are small enough, the field equations can be fully linearized and, thus, numerically solved. Note, however, that treating $\mathcal{M}$ as a 'perturbed' $\overline{\mathcal{M}}$ implies knowing how the background evolves first. Here, we are aided by the fact that $\overline{\mathcal{M}}$, as a featureless manifold, is forced to behave as a perfect fluid (though not necessarily an isentropic one). Because of this symmetry, the field equations can be solved almost completely. Unfortunately, we still need to throw in the equation of state $\bar{p} = \bar{\omega}\bar{\varepsilon}$ of the background source. The question is, do we know $\bar{\omega}$? We do, of course, know the *local* equation $p = \omega\varepsilon$ at every point in $\mathcal{M}$ (this is what defines our model in the first place). The background equation, however, is a large scale average of the local one; consequently, a priori, $\bar{\omega} \equiv \langle\omega\rangle$ and ω should (functionally) differ from each other. It follows that we can't really know the background equation (and thus the background dynamics), unless by first solving $\mathcal{M}$. Needless to say, this is virtually impossible. Still, it is customary to assume that $\bar{\omega} = \omega$ (in the sense of being the same function). In fact, this is always true if the local pressure p is a *linear* function of the local density ε (as is the case of ordinary matter, radiation and some forms of quintessence, like the cosmological constant). But on every other case, it will only be true in certain regions of $\mathcal{M}$, where the anisotropies in the source distribution are 'small' enough. Once they grow 'big', though, the background equation and the local equation are no longer the same ($\bar{\omega} \neq \omega$) and linear theory is rendered invalid. The gCg is a good example of this: while the point to point local behavior in $\mathcal{M}$ is $p = -A\varepsilon^{-\alpha}$, the average one is not (unless $\alpha = 0$):

$$\langle p\rangle = -A\langle\varepsilon^{-\alpha}\rangle \neq -A\langle\varepsilon\rangle^{-\alpha}\,, \tag{4.1}$$

except, of course, if the perturbations are small ($\delta = \delta\varepsilon/\langle p\rangle \ll 1$), in which case

$$\begin{aligned}\langle p\rangle &= -A\langle\varepsilon^{-\alpha}\rangle \simeq -A\langle\varepsilon\rangle^{-\alpha}\langle 1-\alpha\delta\rangle \\ &\simeq -A\langle\varepsilon\rangle^{-\alpha}\,.\end{aligned} \tag{4.2}$$

Now, let us add to this the fact that perturbations have been evolving for a long time, since inflation to the present day, roughly speaking. In most *non*-UDE models, the dark matter component becomes highly clustered during this period (more specifically in the matter era), but because CDM always acts as *pressureless* mater, the average background pressure is not affected by the inhomogeneities in this component. On the other hand, the dark energy component normally evolves homogeneously enough in space, so as to look almost as if a cosmological constant. Thus, the average equation of state is also, typically, not very much affected by the inhomogeneities present in the dark energy distribution. Consequently, for most models and throughout this period, linear theory can be safely used.

However, in the case of UDE models, because dark energy and dark matter arise from the same underlying fluid, the local equation of state is forced to bridge Λ behavior and CDM behavior. This means that anisotropies in the quartessence distribution will most likely affect the average background equation of the Universe. Hence, *if* significant clustering (even if only at small scales) occurs early in the history of the Universe, then $\bar{\omega} \neq \omega$ (in the sense of being different functions) and linear theory is rendered useless (even on large scales); we can still use it, of course, but we'll be perturbing the *wrong* background. In principle, a second order treatment might bring perturbations closer to today than linear theory can, but probably not much closer. The point is that the small scale clustering in the quartessence component necessary to reproduce an equivalent CDM clustering, inevitably affects the equation of state of the average universe. The majority of the studies conducted so far on UDE, do not take this into consideration. Note that the potential breakdown of linear theory at late times is crucial for LSS tests but is of a lesser importance for CMB tests given its earlier occurrence in the history of the Universe. It also affects supernovae and other background results. (Identical worries have also been voiced in [15], but in the context of condensation and vacuum metamorphosis.) Therefore, it seems premature to judge the success of UDE models (as viable alternatives to ΛCDM) solely on traditional background and linear tests. These considerations highlight the fact that the whole concept of UDE is much harder to deal with (even phenomenologically) than originally anticipated.

4.2 Qualitative Approach

In this section, we aim to illustrate in the simplest possible manner how even small scale collapsed regions in a quartessence fluid can affect the evolution of the very large universe. For this purpose, we shall concentrate on two particular cases of unified dark energy fluids.

4.2.1 Case I: the gGg with $\bar{c}_s^2 > 0$

Here, one should bear in mind the fact that the gCg has a minimum density $\bar{\varepsilon}_\star$ it cannot go below. Let us now considering a spherical region in $\mathcal{M}$ of radius R with an average density $\langle\varepsilon\rangle$. If ε was *uniformly* distributed inside this region, then the average pressure would clearly be $\langle p\rangle = -A\langle\varepsilon\rangle^{-\alpha}$. In general, however, it will be something else. To make this more concrete, take the case of a prototypical collapsed region, where the envelope $R_1 < r < R$ has a smaller density than the core $r < R_1$. In fact, let us assume that $\varepsilon(r < R_1) = N\langle\varepsilon\rangle$, where N is some constant higher than one and $\varepsilon(R_1 < r < R) = \varepsilon_\star$ so that $p(R_1 < r < R) = p_\star = -\varepsilon_\star$. (We might consider smoothing the transition between the two areas, but for the type of qualitative argument we're making, this would be overkill.) Because the sum of the energy inside the two regions divided by the entire volume still has to be $\langle\varepsilon\rangle$, this implies that N is equal to

$$N = \left(\frac{R}{R_1}\right)^3 + \left[1 - \left(\frac{R}{R_1}\right)^3\right] \frac{\varepsilon_\star}{\langle\varepsilon\rangle}\,, \tag{4.3}$$

which links the size of the collapsed core to N. It is now straightforward to show that the average pressure is given by

$$\begin{aligned} \langle p\rangle &= \left(\frac{R_1}{R}\right)^3 (N\langle\varepsilon\rangle)^{-\alpha} - \left[1 - \left(\frac{R_1}{R}\right)^3\right] \varepsilon_\star\,, \\ &\simeq -\varepsilon_\star\,, \end{aligned} \tag{4.4}$$

where the approximation is valid for large N (small core). Thus, if $N \gg 1$, the average pressure $\langle p \rangle$, will be considerably larger (in modulus) than the pressure of the Chaplygin gas *in the absence of perturbations*. If we now recall that for the gCg (with $\alpha > 0$), the lower the density, the bigger the pressure, this is why the envelope dominates the entire region R. Hence, the non-linear collapse will work to make $\langle p \rangle = p_\star = -\varepsilon_\star$ early on, thus anticipating and slowing down the transition from dark matter to dark energy behavior.

It should be said, of course, that this is an oversimplified picture: we haven't taken into account the dynamical effects of pressure gradients. In high density regions the pressure will be significantly smaller (in modulus) than average. Still, we need to take into account that the gravitational collapse will only be effective on a given scale λ if $\lambda \gtrsim c_s H^{-1}$. However, once the perturbations become non-linear, the background pressure does not significantly influence any subsequent dynamics.

4.2.2 Case II: UDE with $\bar{c}_s^2 < 0$

Let us now consider a UDE fluid whose background mimics a two-component model of quintessence (with an *effective* $\bar{p}_Q = \bar{\omega}_Q \bar{\varepsilon}_Q$ where $\bar{\omega}_Q$ is a constant) plus pressureless CDM, in other words, a quartessence fluid with $\bar{p} = \bar{p}_Q$ and $\bar{\varepsilon} = \bar{\varepsilon}_Q + \bar{\varepsilon}_{\rm CDM}$. It is simple to check that this fluid does not have an *explicit* isentropic equation of state. Nevertheless, we can still calculate its sound speed indirectly by means of the following trick:

$$\bar{c}_s^2 \equiv \frac{d\bar{p}}{d\bar{\varepsilon}} = \frac{d\bar{p}_Q}{d\bar{\varepsilon}_Q}\,\frac{d\bar{\varepsilon}_Q}{d\bar{\varepsilon}} = \frac{\bar{\omega}_Q(1+\bar{\omega}_Q)\bar{\varepsilon}_Q}{(1+\bar{\omega}_Q)\bar{\varepsilon}_Q + \bar{\varepsilon}_{\rm CDM}}\,. \tag{4.5}$$

Thus, if $-1 \leqslant \bar{\omega}_Q < 0$, linear instabilities are expected to occur in the quartessence fluid. In this case, pressure does not hold the collapse of high-density regions (which tend to behave as pressureless matter), in fact, it contributes to it. On the other hand 'voids' will get increasingly emptier. There is, however, a major difference to the previous case: Now, there is no minimum density underdense regions cannot go below (except, of course, if $\bar{\omega}_Q = -1$), in other words, ε can arbitrarily approach zero.

This has the interesting consequence that the average pressure in the background may actually be very close to zero at all times, so that the large scale universe may never start to accelerate (despite the linear prediction), something which is clearly inconsistent with current observational evidence.

Again, we stress that this is a very simplified example that may not withstand closer inspection; in particular, note that the square sound speed in collapsed regions (albeit negative in the present case) is necessarily small, as these regions tend to behave as matter (this is a general feature of quartessence, after all). Thus, overdense regions may not get that enhanced by pressure, and consequently, the average pressure in $\mathcal{M}$ may not reach arbitrarily small values.

4.3 Quantitative Approach

In this section, we wish to study the onset of the non-linear regime in a more quantitative way. To this effect, we will study how a pure Chaplygin gas ($\alpha = 1$) plus baryons model processes the power spectrum that emerges from radiation-matter equality. Throughout the analysis we adopt priors in agreement with the WMAP first-year data release [5]. The parameters are an equivalent matter density $\Omega_m^\star = 1 - \mathcal{A} + \Omega_b = 0.29$, a baryon density $\Omega_b = 0.047$, an equivalent cosmological constant density $\Omega_\Lambda^\star = 1 - \Omega_m^\star = 0.71$, a Hubble parameter $h = 0.71$, a normalization $\sigma_8 = 0.9$ (using a top hat filter) and a primordial Harrison-Zel'dovich spectrum. We start by determining how linear density perturbations of the Chaplygin gas and baryon components evolve with time using the machinery of the previous chapter. Note that the use of linear theory so early in the matter era is actually a good approximation on large enough scales. This is because the effects of the breakdown of linear theory will only be important (on large cosmological scales) much latter, when a smooth transition from dark matter to dark energy domination would be naively expected.

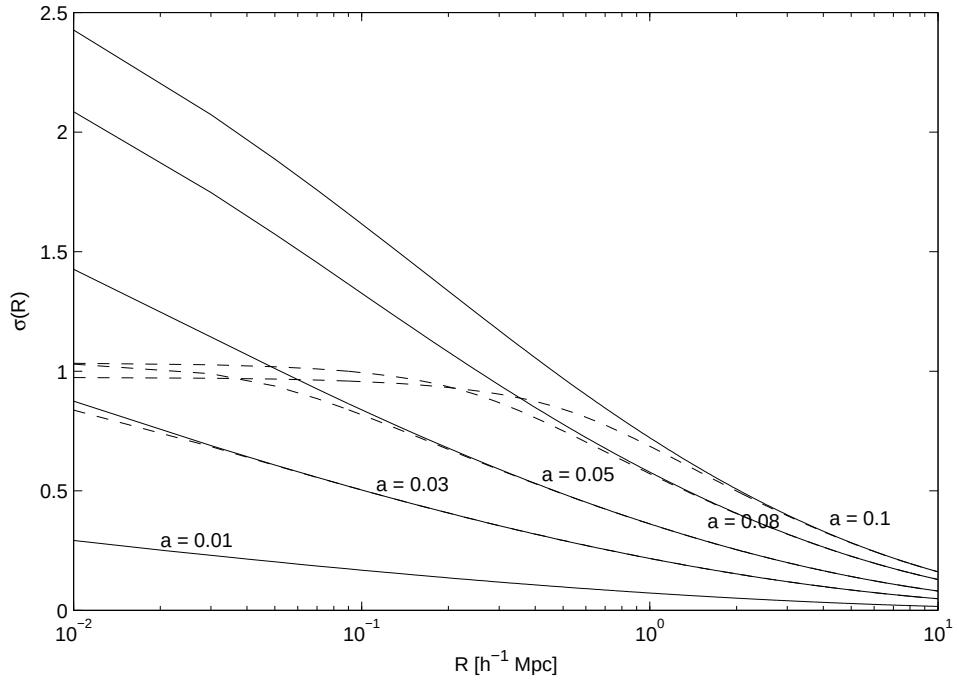


Figure 4.1: The linear evolution of the σ mass dispersion in the baryon (solid line) and Chaplygin gas (dashed lines) components as a function of R and a assuming $\alpha = 1$. Note that at early times the baryon and Chaplygin gas fluctuations evolve in tandem. Later on, pressure effects prevent the Chaplygin gas from collapsing further. However, baryon fluctuations can still keep growing (albeit at a slower pace).

4.3.1 Mass Dispersion

Therefore, we use linear theory in order to compute the value of the dispersion of the density fluctuations in the baryon and Chaplygin gas components $\sigma(R, a)$, as a function of R and a. This is plotted in Fig. 4.1 for the particular case of the original Chaplygin gas with $\alpha = 1$. We see that since the Chaplygin gas behaves as matter at early times, perturbations will grow proportionally to the scale factor, in tune with those in the baryonic component. Therefore, both fluids evolve in the same way early on and have approximately the same value of σ on all relevant scales. During this stage we have that $\sigma \propto a = (1 + z)^{-1}$. Later on, the pressure of the Chaplygin gas will have increased dramatically preventing its further collapse.

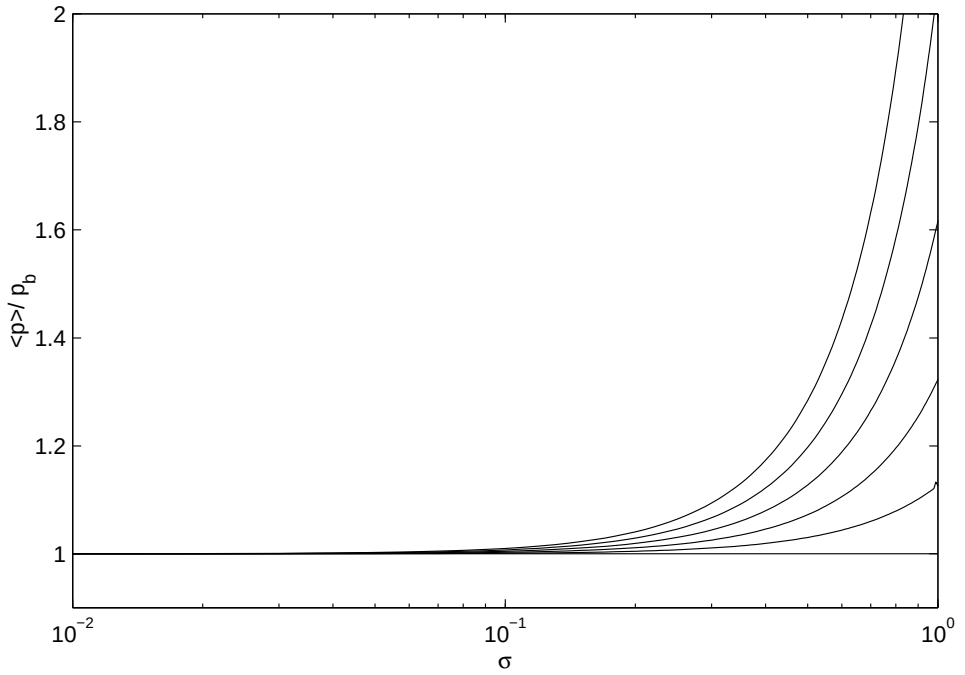


Figure 4.2: The ratio between the average value of p and and its zeroth order background value, $\langle p\rangle/\bar{p}$, as a function of the mass dispersion σ of the linear density fluctuations in the Chaplygin gas component for various values of $\alpha = [0, 0.2, 0.4, 0.6, 0.8, 1]$ (bottom up). For $\alpha > 0$ we clearly see that $\langle p\rangle/\bar{p}$ rapidly diverges from unity if σ is large enough.

However, the baryon fluctuations can still keep growing (at a slower pace). We also see in Fig. 4.1 that the Chaplygin gas component becomes non-linear on small scales very early in the matter era. It is clear that when this happens a significant fraction of the Chaplygin gas will have collapsed and decoupled from the background so that a transition from a dark matter-like to a dark energy-like stage (which necessarily requires lower densities) never happens in those regions.

4.3.2 $\langle p\rangle$ in the Non-Linear Regime

Using the Press-Schechter framework [122], we can show that for $\sigma = 1$ the fraction of the equivalent mass that is incorporated in collapsed objects is close to 0.1. In the

non-linear regime, an initial gaussian density field is better described by a lognormal one-point probability distribution function, $\mathcal{P}(\delta)$, given by (see, for example, [123] and references therein)

$$\mathcal{P}(\delta) = \frac{(1+\delta)^{-1}}{\sqrt{2\pi \ln(1+\sigma_{nl}^2)}} \exp\left(-\frac{\ln^2\left((1+\delta)\sqrt{1+\sigma_{nl}^2}\right)}{2\ln(1+\sigma_{nl}^2)} \right), \tag{4.6}$$

where $\sigma_{nl}^2 = \exp(\sigma^2) - 1$ and σ is computed using linear theory. We use (4.6) in order to estimate the ratio between the average pressure $\langle p \rangle$ and its zero-order background value $\bar{p} = -A/\bar{\varepsilon}^{\alpha}$,

$$\langle p \rangle / \bar{p} = \int_{-1}^{\infty} (1+\delta)^{-\alpha} \mathcal{P}(\delta) d\delta \,, \tag{4.7}$$

as a function of the dispersion of the density fluctuations in the generalized Chaplygin gas component, σ. (An objection may be raised here: Since the Chaplygin gas has a minimum non-null density, the lower limit in (4.7) cannot be exactly -1. However, if this minimum density is much smaller than the average density, using this lower limit doesn't create any major problem.) The results of this analysis are displayed in Fig. 4.2 for various values of α. We see that in all but one case (for $\alpha = 0$) the average value of the pressure strongly diverges from its zero-order background value as soon as σ becomes large enough. This inevitably causes the breakdown of linear theory. The magnitude of this effect becomes more pronounced at late times when the negative pressure starts to become dynamically important on all scales.

We thus conclude that weakly perturbing a homogenous Chaplygin gas does not take into account the effect that non-linearities have on the behavior of the very large universe. This caveat has crucial implications for the predicted observational consequences of the model given that linear theory breaks down at late times even on large cosmological scales (except in the $\alpha = 0$ case).

4.4 Conclusions and Future Prospects

Weakly perturbing a canonic UDE background fluid in the traditional way, where we simply assume that the background equation has the same functional form as the local one, effectively neglects the potential effect that collapsed regions an voids may have on the behavior of the average universe. We have argued in this chapter, both qualitatively and quantitatively, that outside the ΛCDM limit, non-linearities in UDE models cannot be safely ignored. This means that any dramatic conclusion on the fate of UDE solely based on background and linear tests, are premature at best. The average background and clustering properties of quartessence (treated as a single fluid) can only be definitely settled by solving the full non-linear Einstein equations. Obviously, this can only be accomplished in a handful of high symmetry situations; in the future, we intend to explore such configurations (like closely spaced sheets of 'matter' separated by voids) to see if some light can be shed on the background and clustering properties of quartessence. Preliminary work on this front has already begun.

4.5 A Possible Way Out?

Now, let us take a step back and consider the following: In the so-called concordance model of Cosmology, a range of observational data is used to postulate the existence of two dark fluids (dark matter and dark energy) for which so far there is no direct experimental evidence. The most common attitude towards dark energy and dark matter (in the context of General Relativity) is to model them as if two distinct minimally coupled fluids. The direct opposite to this, is to treat them as different manifestations of a single fluid (which has been the main focus of this thesis). An intermediate approach, on the other hand, is to treat them as two coupled fluids. In this case, however, if the coupling is very strong, we naturally expect the distinction of dark energy and dark matter as two different fluids to become somewhat blurred. In other words, if the coupling between them is very strong, then, to a certain extent,

we expect dark energy and dark matter to behave as if a single (quartessence) fluid. As far as we are aware today, this bridge between strongly coupled models and UDE has not been significantly explored by the community. To make this more concrete, let us take the case where dark energy and dark matter are coupled in the following way

$$\widehat{\mathcal{L}} = X - V(\phi) + h(\phi)\,\widehat{\mathcal{L}}_{\rm DM}\,, \tag{4.8}$$

where $\widehat{\mathcal{L}}_{\rm DM} \propto Y^n$ and

$$Y = -\frac{1}{2}\nabla^\mu\varphi\nabla_\mu\varphi\,. \tag{4.9}$$

(In this type of model, ϕ is normally called a 'Chameleon' field [124, 125, 126, 127].) Recall that a Lagrangian proportional to a power of the kinetic term of a given field φ describes a constant-ω isentropic fluid with $\omega = 1/(2n-1)$ as discussed in § 2.1. Thus, in the limit of large n, our $\widehat{\mathcal{L}}_{\rm DM}$ above describes pressureless (non-relativistic) dark matter. It follows that we can rewrite (4.8) in the form

$$\widehat{\mathcal{L}} = X - V(\phi) + g(\phi)\,\varepsilon_{\rm DM}\,, \tag{4.10}$$

where $g(\phi) = \omega_{\rm DM}h(\phi)$ is a rescaled coupling constant. It is easy to see (by varying the action in relation to φ) that the dark matter component evolves independently from the chameleon dark energy field ϕ. On the other hand, the evolution of ϕ is given by

$$\Box\phi = \frac{\partial V_{\rm eff}}{\partial\phi}\,, \tag{4.11}$$

where

$$V_{\rm eff} = V(\phi) - g(\phi)\,\varepsilon_{\rm DM}\,, \tag{4.12}$$

and therefore is affected by how dark matter is evolving. (Note here that although $V_{\rm eff}$ is almost V, its derivative can be very different from $\partial V/\partial\phi$). As for the energy-momentum tensor associated with (4.8), it is a simple matter to show (varying the action in relation to $g^{\mu\nu}$) that

$$\begin{aligned} T_{\mu\nu}(\phi,\varphi) = \nabla_\mu\phi\nabla_\nu\phi \;&+\; (X - V(\phi))g_{\mu\nu} + \\ &+\; h(\phi)\left[g_{\mu\nu}Y^n + nY^{n-1}\nabla_\mu\varphi\nabla_\nu\varphi\right]\,. \end{aligned} \tag{4.13}$$

Obviously, this energy-momentum tensor does not, in general, describe a perfect fluid. However, in the so-called *adiabatic regime* (described, in detail, in [128, 129]), it is assumed that the gradients of ϕ are negligible both in $T_{\mu\nu}$ and in the equation of motion (4.11). Thus, in this regime, (4.13) reduces to

$$T_{\mu\nu} \simeq (hY^n - V)g_{\mu\nu} + nhY^{n-1}\nabla_\mu\varphi\nabla_\nu\varphi\,, \tag{4.14}$$

which can be immediately rewritten in a perfect fluid form, if we make the following identifications

$$\begin{aligned} \varepsilon_{\text{eff}} &= (2n-1)hY^n + V\,, \\ &= h\,\varepsilon_{\text{DM}} + V\,, \\ p_{\text{eff}} &= hY^n - V\,, \\ &= g\,\varepsilon_{\text{DM}} - V \simeq -V\,, \end{aligned} \tag{4.15}$$

and $u_\mu = \nabla_\mu\varphi/\sqrt{2Y}$. Is it an isentropic fluid, though? Yes. Since the adiabatic regime is also characterized by the condition $\partial V_{\text{eff}}/\partial\phi = 0$, it follows from (4.12) that the value of ϕ is univocally related to ε_{DM}. Hence, p_{eff} only depends on the value of ε_{eff} and therefore the fluid is isentropic (although, in general, we won't have an explicit isentropic $p_{\text{eff}} = p_{\text{eff}}(\varepsilon_{\text{eff}})$ equation of state).

Now, the value of

$$m^2_{\text{eff}} \equiv \frac{\partial^2 V_{\text{eff}}}{\partial\phi^2}\,, \tag{4.16}$$

called the effective square mass of the chameleon field, determines the length scales for which the adiabatic approximation is valid. Specifically, this is the case for large scale perturbations with $L \gg m^{-1}_{\text{eff}}$, while for scales much smaller than this, the approximation is no longer valid. We thus conclude that above a certain scale, sufficiently coupled models behave as a single isentropic fluid but not below. Why is this relevant? It is relevant because such differentiated behavior above or below a certain scale may help solve the averaging problem that affects UDE models. Recall that in (canonic) UDE models, an isentropic fluid description is valid at all

scales and this is why quartessence is so susceptible to non-linearities. For the sake of argument, suppose the majority of the non-linear clustering occurs for scales smaller than $m_{\rm eff}^{-1}$; since now they are confined to a non-isentropic part of the fluid, it is possible that they may not affect the average background equation of state as before. If, on the other hand, significant non-linear clustering does extend beyond this scale, then non-linearities will still be a major problem in strongly coupled scenarios. A more detailed analysis is in order to determine if this turns out to be a successful solution to the averaging problem or not. (Note, however, that such large couplings are strongly constrained by several equivalence principle type experiments, so the cosmological relevance of these models remains unclear.)

As a final note, we would like to show how easy it is to obtain some key results regarding linear instabilities in strongly coupled models if we treat them as a single fluid. Say we start with ordinary quintessence plus a dark matter component. If they are minimally coupled, ϕ will roll down the potential V as usual. On the other hand, if we view ϕ as strongly coupled to dark matter, ϕ will roll down an effective $V_{\rm eff}$ that is very close to the original potential. Above a certain scale, this effective fluid behaves as an isentropic fluid. Additionally, in the absence of perturbations, it satisfies the usual

$$\frac{\dot{\varepsilon}_{\rm eff}}{\varepsilon_{\rm eff}} = -3H(1+\omega_{\rm eff})\,, \tag{4.17}$$

with $\omega_{\rm eff} > -1$. Since $\varepsilon_{\rm eff} > 0$, it follows trivially that $\dot{\varepsilon}_{\rm eff} < 0$. On the other hand, we know that $\dot{p}_{\rm eff} = -\dot{V}_{\rm eff} > 0$. Thus,

$$c_s^2 = \frac{\dot{p}_{\rm eff}}{\dot{\varepsilon}_{\rm eff}} < 0\,, \tag{4.18}$$

and instabilities are expected to occur. Linear instabilities such as these have been studied by a variety of authors [130, 131, 129, 128, 132], but here we have obtained them in a much simpler and straightforward manner.

Appendix A

Thesis X-Ray

Here, we summarize in bullet form the main results that have been obtained in the course of this work.

i. Canonic Implementation of UDE: In Chapter 2, a careful discussion was made relating to the nature of a *single* fluid. A single ('atomic') fluid is then defined as a fluid that can be described by a Lagrangian of the form $\widehat{\mathcal{L}}(X, \phi)$, where ϕ is a real scalar field. This field ϕ is formally equivalent to having a perfect fluid (though, not necessarily an isentropic one). It is argued that treating quartessence as a single fluid is the simplest possible way of implementing UDE, hence 'canonic' quartessence.

ii. Low-Level Implementation of the gCg as a scalar field obeying the Lagrangian (2.11) using a novel approach.

iii. Complete ΛCDM equivalence to the gCg when $\alpha \to 0$. Gravity alone does not distinguish the two. This result was established in a very straightforward manner, expanding the Lagrangian of the gCg and taking the limit $\alpha \to 0$. As far as we know, this demonstration has never been presented before in such a simple and direct way.

iv. Background Constraints on the gCg, *in the absence of perturbations*,

were obtained using the distance modulus of 192 Type Ia Supernovae. The parameter α is not significantly constrained by this analysis. On the other hand, the constrains on $\mathcal{A}$ depend on whether one lets the gCg coexist with a CDM component (in which case $\mathcal{A} \geqslant 0.95$) or a baryonic component ($0.7 < \mathcal{A} < 0.92$). In the former case, the gCg is forced to behave very close to a Λ cosmological constant (which is hardly surprising) while in the latter case, the Λ limit is strongly disfavored.

v. Sound Speed Link to how Fast the Quartessence Transition Occurs: if $c_s^2 > 0$ the background transition from DM to DE occurs faster than in ΛCDM (slower, if $c_s^2 < 0$). Linear instabilities are briefly discussed in Chapters 3 and 4.

vi. The Crucial Role of Baryons in the Formation of LSS: The gCg alone (unless $|\alpha| < 10^{-5}$) is not able to reproduce the 2dF mass power spectrum for large scales. In a nutshell, this is caused by the fact that the gCg attains very large sound speeds during the background transition. Thus, perturbations on large scales become heavily damped (or blow up exponentially) by the effect of pressure; the only way to avoid this is by having a very small α. If, on the other hand, baryons are added to the mixture, perturbations can still keep growing in the baryonic component (since baryons always have a very small sound speed), even when the gCg starts behaving differently from CDM. Although baryons are not that important for background studies, they are nonetheless crucial for LSS formation.

vii. Linear Constraints on α: if baryons are included, α is no longer forced to be very small. In fact, the entire interval $0 \leqslant \alpha \leqslant 1$ is consistent with the 2dF power spectrum (with $\mathcal{A} \simeq 0.8$). On the other hand, if we make a SN+LSS joint analysis (using the latest 192 supernovae sample), we find that only $\alpha \simeq 0$ and $0.2 < \alpha \leqslant 1$ are now possible. If to this we add the latest CMB constraint on α, i.e. $\alpha < 0.2$, we conclude that even with baryons present, the gCg is forced to behave very close to ΛCDM in order for it to simultaneously reproduce SN, CMB and LSS observations.

viii. The Averaging Problem in UDE: In Chapter 4, we argue both qualitatively and quantitatively that in the context of canonic quartessence, non-linear small scale clustering cannot be safely ignored, as these may have a significant impact on the background equation of state at late times. Since the local quartessence equation has to bridge dark matter and dark energy behavior, the average equation of state is inevitably affected by the anisotropies in this fluid. The problem is that once these become significant, we no longer know what background to perturb. To find out, we would have to solve the full Einstein field equations. The consequence of this is that the majority of background and linear tests have very shaky foundations. Thus, it is rather premature to make any dramatic conclusion on the fate of UDE, solely based on traditional tests, without first taking into account the effects of non-linearities. UDE models are therefore much more complicated to test than initially anticipated. One way that perhaps may shed some light on the matter is to study high symmetry configurations involving a gCg. These high symmetry configurations can be solved numerically many times without any approximations. Some preliminary work on this front has already begun.

ix. Strongly Coupled Models: The averaging problem in UDE models is rooted in the fact that quartessence is described by an isentropic fluid at every scale. On the other hand, strongly coupled models, as discussed at the end of Chapter 4, above a certain scale can be interpreted as an isentropic fluid but not below. Now, if small scale clustering occurs mainly at scales where the isentropic approximation is not valid, then it may be possible that the very large universe is not affected by them. On the other hand, if the clustering extends well beyond the critical scale into isentropic territory, then the average equation of state will still be, most likely, affected and strongly coupled models share the same problems as quartessence. Some significant work has yet to be done on this front which, a priori, seems promising.

Appendix B

Numerical Code

CHAPLYGIN.H: Glues all files together

```
1 #include <math.h>
2 #include <stdio.h>
3 #include <stdlib.h>
4 #include <gsl/gsl_errno.h>
5 #include <gsl/gsl_odeiv.h>
6
7 /* From model.c */
8 double H(double x, double omega_b0, double Abar, double alpha);
9 double w(double x, double Abar, double alpha);
10 double qui(double x, double omega_b0, double Abar, double alpha);
11 double omega_b(double x, double omega_b0, double Abar, double alpha);
12 double ips(double k, double g, double A);
13
14 /* From transfer.c */
15 int dydx (double x, const double y[], double dy[], void *params);
16 double transfer (double k, double omega_b0, double h, double Abar,
       double alpha);
```

MODEL.C: Model stuff

```
1 #include "chaplygin.h"
2
```

```
3 /* Hubble parameter */
4 double H(double x, double omega_b0, double Abar, double alpha) {
5
6         double a1 = exp(-3*x);
7         return sqrt(omega_b0*a1+(1-omega_b0)*pow(Abar+(1-Abar)*pow(a1,
              1+alpha), 1/(1+alpha)));
8 }
9
10 /* w for the gCg component, x=ln(a) */
11 double w(double x, double Abar, double alpha) {
12
13         return -1/(1+(1-Abar)*exp(-3*(1+alpha)*x)/Abar);
14 }
15
16 /* qui=H'/H where '=d/dx */
17 double qui(double x, double omega_b0, double Abar, double alpha) {
18
19         double   a1 = exp(-3*x),
20                  a2 = pow(a1, (1+alpha)),
21                  a3 = Abar+(1-Abar)*a2,
22                  num,
23                  den;
24
25         num = a1*omega_b0+(1-Abar)*(1-omega_b0)*a2*pow(a3,
              -alpha/(1+alpha));
26         den = a1*omega_b0+(1-omega_b0)*pow(a3, 1/(1+alpha));
27         return -1.5*num/den;
28 }
29
30 /* Baryon fraction */
31 double omega_b(double x, double omega_b0, double Abar, double alpha) {
32
33         double  a1 = exp(-3*x),
34                 a2 = omega_b0*a1;
35
36         return a2/(a2+(1-omega_b0)*pow(Abar+(1-Abar)*pow(a1, 1+alpha),
              1/(1+alpha)));
37 }
```

```
38
39 /* Power spectrum */
40 double ips(double k, double shape, double A) {
41
42         double  q = k/shape,
43                 t1, t2, t3, t4,
44                 tk;
45
46         t1 = 3.89*q;
47         t2 = pow(16.1*q, 2);
48         t3 = pow(5.46*q, 3);
49         t4 = pow(6.71*q, 4);
50
51         tk = pow((1 + t1 + t2 + t3 + t4), -0.25);
52
53         tk = tk*log(1+2.34*q)/(2.34*q);
54
55         /* CDM processed Harrison-Zeldovich A*k spectrum */
56         return A*k*pow(tk, 2);
57 }
```

TRANSFER.C: Perturbation Machinery

```
1 #include "chaplygin.h"
2
3 int dydx(double x, const double y[], double dy[], void *params) {
4
5         const double * const par=params;
6
7         double omega_b0 = par[0],
8                 Abar = par[1],
9                 alpha = par[2],
10                k = par[3],
11                a1 = w(x, Abar, alpha),
12                a2 = 1+a1, a3=exp(x)*H(x, omega_b0, Abar, alpha),
13                a4 = omega_b(x, omega_b0, Abar, alpha);
14
15        dy[0] = y[1];
```

```
16          dy[1] = 1.5*a4*y[0]-(2+qui(x, omega_b0, Abar,
                alpha))*y[1]+1.5*(1-a4)*(1-3*alpha*a1)*y[2];
17          dy[2] = a2*(y[1]-y[3]/a3)+3*(1+alpha)*a1*y[2];
18          dy[3] = -(alpha*a1*y[2]*k*k)/(a3*a2)-(1+3*alpha*a1)*y[3];
19
20          return GSL_SUCCESS;
21  }
22
23  double transfer(double k, double omega_b0, double h, double Abar,
        double alpha) {
24
25          const gsl_odeiv_step_type *T = gsl_odeiv_step_rk8pd;
26          gsl_odeiv_step *s = gsl_odeiv_step_alloc(T, 4);
27          gsl_odeiv_control *c = gsl_odeiv_control_y_new(1e-8, 0);
28          gsl_odeiv_evolve *e = gsl_odeiv_evolve_alloc(4);
29
30          double par[4] = {omega_b0, Abar, alpha, 2998*k};
31
32          gsl_odeiv_system sys = {dydx, NULL, 4, par};
33
34          double x = log(0.01),
35                 x1 = 0,
36                 h_istep = 1e-6,
37                 y[4] = {1, 1, 1, 0};
38
39          while (x < x1) {
40
41                 int status = gsl_odeiv_evolve_apply (e, c, s, &sys, &x,
                       x1, &h_istep, y);
42
43                 if (status != GSL_SUCCESS)
44                        break;
45          }
46
47          gsl_odeiv_evolve_free (e);
48          gsl_odeiv_control_free (c);
49          gsl_odeiv_step_free (s);
50
```

```
51          return y[0];
52  }
```

GO TRANSFER.C: Transfer Functions Driver

```
 1  #include "chaplygin.h"
 2
 3  /* Global Variable Declarations */
 4  const double k_2df[49], k_2df_windows[20][49], k_2df_power[20][5];
 5
 6  int main(void) {
 7
 8  FILE *fp;
 9  int i, j, k;
10
11  double  Abar[100],
12          alpha[100],
13          omega_b0=.044,
14          h=.71,
15          t[100][100][49];
16
17  extern const double   k_2df[49],
18                        k_2df_windows[20][49],
19                        k_2df_power[20][5];
20
21  /* Load k's for the window functions */
22  fp=fopen("2df_k.dat", "r");
23  for (i=0; i < 49; ++i)
24          fscanf(fp, "%e", &k_2df[i]);
25
26  /* Load window functions */
27  fp=fopen("2df_windows.dat", "r");
28  for (i=0; i < 20; ++i)
29  for (j=0; j < 49; ++j)
30          fscanf(fp, "%e", &k_2df_windows[i][j]);
31
32  /* Load 2df data */
33  fp=fopen("2df_power.dat", "r");
```

```
34 for (i=0; i < 20; ++i)
35         for (j=0; j < 5; ++j)
36                 fscanf(fp, "%e", &k_2df_power[i][j]);
37
38         fclose(fp);
39
40 /* Generate a 100*100 grid of [alpha, Abar] */
41 for (i=0; i < 100; ++i)
42         alpha[i]=.0 +i*1.0/99;        /* alpha=[0, ... , 1]     */
43
44 for (j=0; j < 100; ++j)
45         Abar[j]=.0 + j*.9999/99;      /* Abar=[0, ... , .9999] */
46
47 /* Calculate only once the transfer matrix t[i][j][k] (100x100x49) for
       a grid of models alpha[i], Abar[j] at k_2df[k] */
48
49 for (i=0; i < 100; ++i) {
50         for (j=0; j < 100; ++j) {
51                 for (k=0; k < 49; ++k) {
52
53                 t[i][j][k]=transfer(k_2df[k], omega_b0, h, Abar[j],
                       alpha[i]);
54                 printf("%i\t%i\t%i\t%e\n", i, j, k, t[i][j][k]);
55                 }
56         }
57 }
58 return 0;
59 }
```

GO CHI.C: 2dF χ^2 fitting

```
1 #include "chaplygin.h"
2
3 const double     trans[490000][4],
4                 k_2df[49],
5                 k_2df_windows[20][49],
6                 k_2df_power[20][5];
7
```

```
 8
 9 int main(void) {
10
11 FILE *fp;
12 int i, j, k, m, n;
13
14 double  Abar[100],
15         alpha[100],
16         A[500],
17         shape,
18         omega,
19         omega_b0=.044,
20         h=.71,
21         term,
22         t[100][100][49];
23
24 extern const double   k_2df[49],
25                       k_2df_windows[20][49],
26                       k_2df_power[20][5],
27                       trans[490000][4];
28
29 double  p[49],
30         Wp[20],
31         prob_matrix[500][100][100],
32         sum[20],
33         s,
34         norm=0,
35         prob_sum_A[100][100];
36
37
38 /* Load transfer matrix */
39 fp=fopen("transfer_matrix.dat", "r");
40
41         for (i=0; i < 490000; ++i)
42                 fscanf(fp, "%f %f %f %f", &trans[i][0], &trans[i][1],
                       &trans[i][2], &trans[i][3]);
43
44
```

```
45 /* Reconstruct 3D transfer matrix t(alpha, Abar, k_2df) (100x100x49) */
46 for (i=0; i < 490000; ++i)
47         t[(int) trans[i][0]][(int) trans[i][1]][(int)
               trans[i][2]]=trans[i][3];
48
49
50 /* Load k's for the window functions */
51 fp=fopen("2df_k.dat", "r");
52 for (i=0; i <= 48; ++i)
53         fscanf(fp, "%e", &k_2df[i]);
54
55 /* Load window functions */
56 fp=fopen("2df_windows.dat", "r");
57 for (i=0; i <= 19; ++i)
58 for (j=0; j <= 48; ++j)
59         fscanf(fp, "%e", &k_2df_windows[i][j]);
60
61 /* Load 2df data */
62 fp=fopen("2df_power.dat", "r");
63 for (i=0; i <= 19; ++i)
64         for (j=0; j <= 4; ++j)
65                 fscanf(fp, "%e", &k_2df_power[i][j]);
66
67         fclose(fp);
68
69 /* Generate a 500x100*100 grid of [A, alpha, Abar] */
70
71 for (i=0; i < 500; ++i)
72         A[i]=0 + i*3000./499;          /* A[i]=[0, ..., 3000] */
73
74 for (j=0; j < 100; ++j)
75         alpha[j]=.0 +j*1./99;          /* alpha=[0, ..., 1]      */
76
77 for (k=0; k < 100; ++k)
78         Abar[k]=.0 + k*.9999/99;      /* Abar=[0, ..., .9999]   */
79
80
81         term = sqrt(h/.5);
```

```
82
83 /* Probability Matrix (following Tegmark's notes) */
84
85 for (i=0; i < 500; ++i) {
86     for (j=0; j < 100; ++j) {
87         for (k=0; k < 100; ++k) {
88
89                     omega=omega_b0+(1-omega_b0)*pow(1-Abar[k],
                            1/(1+alpha[j]));
90                     shape=omega*h*exp(-omega_b0*(1+term/omega));
91                     for (m=0; m < 49; ++m) {
92                     p[m]=ips(k_2df[m], shape, A[i])*t[j][k][m]*t[j][k][m];
93                     }
94
95                     for (m=0; m < 20; ++m) {
96                             Wp[m]=0;
97                             for (n=0; n < 49; ++n)
98                                     Wp[m] += k_2df_windows[m][n]*p[n];
99
100                            sum[m] =
                                pow((k_2df_power[m][3]-Wp[m])/k_2df_power[m][4],
                                  2);
101                    }
102
103                    s=0;
104                    for (m=0; m < 20; ++m) {
105                    s += sum[m];
106
107                    }
108
109        prob_matrix[i][j][k]=exp(-s/2);
110        norm += prob_matrix[i][j][k];
111
112        }
113    }
114 }
115
116 /* Normalized probability matrix */
```

```
117 for (i=0; i < 500; ++i) {
118     for (j=0; j < 100; ++j) {
119         for (k=0; k < 100; ++k) {
120 
121         prob_matrix[i][j][k]=prob_matrix[i][j][k]/norm;
122 
123         }
124     }
125 }
126 
127 /* Probability Matrix summed over A */
128 
129 for (j=0; j < 100; ++j) {
130     for (k=0; k < 100; ++k) {
131         prob_sum_A[j][k]=0;
132 
133           for (i=0; i < 500; ++i)
134               prob_sum_A[j][k] += prob_matrix[i][j][k];
135 
136         if (k % 100 == 0) printf("\n");
137         printf("%e\t", prob_sum_A[j][k]);
138 
139     }
140 }
141 
142 return 0;
143 }
```

CONTOUR.C: Confidence Contours

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 
4 int cmp(const void *vp, const void *vq);
5 
6 /* To use with qsort: the compare function */
7 int cmp(const void *vp, const void *vq) {
8 
```

```
 9          const float *p = vp, *q = vq;
10          float dif = *p - *q;
11
12          return ((dif >= 0.0) ? ((dif > 0.0) ? +1 : 0) : -1);
13  }
14
15  int main(void) {
16
17          FILE *fp;
18
19          int i, j, k;
20          float list[100*100], contour[100][100], sum;
21
22          /* Load normalized probability matrix (prob_sum_A.dat) into
                contour[][] */
23          fp = fopen("prob_sum_A.dat", "r");
24
25          for (i=0; i < 100; ++i)
26              for (j=0; j < 100; ++j)
27                  fscanf(fp, "%e", &contour[i][j]);
28
29
30          /* Make a single row contaning all contour rows */
31          for (i=0; i < 100; ++i)
32              for (j=0; j < 100; ++j)
33                  list[i*100 + j] = contour[i][j];
34
35          /* Order list using qsort */
36          qsort(list, 10000, sizeof(float), cmp);
37
38          /* Accumulated probability (well, 1-it) */
39          for (i=0; i < 100; ++i) {
40              for (j=0; j < 100; ++j) {
41                  sum = 0;
42
43                      for (k=0; list[k] < contour[i][j]; ++k)
44                          sum += list[k];
45
```

```
46              contour[i][j] = 1 - sum;
47
48              if (j % 100 == 0) printf("\n");
49              printf("%e\t", contour[i][j]);
50              }
51          }
52  }
```

Bibliography

[1] S. Perlmutter et al. Measurements of Ω and Λ from 42 high-redshift supernovae. *Astrophys. J.*, 517:565–586, 1999.

[2] Adam G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.*, 116:1009–1038, 1998.

[3] John L. Tonry et al. Cosmological results from high-z supernovae. *Astrophys. J.*, 594:1–24, 2003.

[4] Adam G. Riess et al. Type Ia supernova discoveries at $z > 1$ from the HubbleSpace Telescope: Evidence for past deceleration and constraints on dark energy evolution. *Astrophys. J.*, 607:665–687, 2004.

[5] D. N. Spergel et al. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *Astrophys. J. Suppl.*, 148:175, 2003.

[6] D. N. Spergel et al. Wilkinson Microwave Anisotropy Probe (WMAP) three year results: Implications for cosmology. 2006.

[7] Sean M. Carroll. The cosmological constant. *Living Rev. Rel.*, 4:1, 2001.

[8] R. R. Caldwell, Rahul Dave, and Paul J. Steinhardt. Cosmological imprint of an energy component with general equation-of-state. *Phys. Rev. Lett.*, 80:1582–1585, 1998.

[9] Li-Min Wang, R. R. Caldwell, J. P. Ostriker, and Paul J. Steinhardt. Cosmic concordance and quintessence. *Astrophys. J.*, 530:17–35, 2000.

[10] C. Armendariz-Picon, Viatcheslav F. Mukhanov, and Paul J. Steinhardt. Essentials of k-essence. *Phys. Rev.*, D63:103510, 2001.

[11] R. R. Caldwell. A phantom menace? *Phys. Lett.*, B545:23–29, 2002.

[12] Sean Carroll, Mark Hoffman, and Mark Trodden. Can the dark energy equation-of-state parameter ω be less than -1? *Phys. Rev.*, D68:023509, 2003.

[13] J. S. Bagla, Hårvinder Kaur Jassal, and T. Padmanabhan. Cosmology with tachyon field as dark energy. *Phys. Rev.*, D67:063504, 2003.

[14] T. Padmanabhan. Accelerated expansion of the universe driven by tachyonic matter. *Phys. Rev.*, D66:021301, 2002.

[15] Bruce A. Bassett, Martin Kunz, David Parkinson, and Carlo Ungarelli. Condensate cosmology - dark energy from dark matter. *Phys. Rev.*, D68:043504, 2003.

[16] Bruce A. Bassett, Martin Kunz, Joseph Silk, and Carlo Ungarelli. A late-time transition in the cosmic dark energy? *Mon. Not. Roy. Astron. Soc.*, 336:1217–1222, 2002.

[17] Ruth Durrer and Roy Maartens. Dark Energy and Dark Gravity. 2007.

[18] P. P. Avelino and C. J. A. P. Martins. A supernova brane scan. *Astrophys. J.*, 565:661, 2002.

[19] Max Tegmark et al. Cosmological parameters from SDSS and WMAP. *Phys. Rev.*, D69:103501, 2004.

[20] Garry W. Angus, HuanYuan Shan, HongSheng Zhao, and Benoit Famaey. On the law of gravity, the mass of neutrinos and the proof of dark matter. *Astrophys. J.*, 654:L13–L16, 2007.

[21] Jacob D. Bekenstein. Relativistic gravitation theory for the MOND paradigm. *Phys. Rev.*, D70:083509, 2004.

[22] J. W. Moffat. Nonsymmetric gravitational theory. *Phys. Lett.*, B355:447–452, 1995.

[23] J. R. Brownstein and J. W. Moffat. Galaxy rotation curves without non-baryonic dark matter. *Astrophys. J.*, 636:721, 2006.

[24] J. W. Moffat. Gravitational lensing in modified gravity and the lensing of merging clusters without dark matter. 2006.

[25] Alexander Yu. Kamenshchik, Ugo Moschella, and Vincent Pasquier. An alternative to quintessence. *Phys. Lett.*, B511:265–268, 2001.

[26] Neven Bílic, Gary B. Tupper, and Raoul D. Viollier. Unification of dark matter and dark energy: The inhomogeneous Chaplygin gas. *Phys. Lett.*, B535:17–21, 2002.

[27] R. Jackiw. A particle field theorist's lectures on supersymmetric, non-abelian fluid mechanics and d-branes, 2000.

[28] Mokhtar Hassaine. Supersymmetric Chaplygin gas. *Phys. Lett.*, A290:157–164, 2001.

[29] M. C. Bento, O. Bertolami, and A. A. Sen. Generalized Chaplygin gas, accelerated expansion and dark energy-matter unification. *Phys. Rev.*, D66:043507, 2002.

[30] Steven Weinberg. *Gravitation and Cosmology: Principles and Practice of the General Theory of Relativity.* Wiley, 1972.

[31] Bernard Schutz. *Geometrical Methods of Mathematical Physics.* Cambridge University Press, 1999.

[32] Hans C. Ohanian and Remo Ruffini. *Gravitation and Spacetime.* Norton, 1994.

[33] Viatcheslav Mukhanov. *Physical Fundations of Cosmology.* Cambridge University Press, 2005.

[34] Sean Carroll. *An Introduction to General Relativity Spacetime and Geometry.* Pearson, Addison Wesley, 2004.

[35] Edward W. Kolb and Michael S. Turner. *The Early Universe.* Addison-Wesley, 1990.

[36] Peter Coles and Francesco Lucchin. *Cosmology, The Origin and Evolution of Cosmic Structure.* John Wiley & Sons, 1995.

[37] Adam G. Riess et al. The farthest known supernova: Support for an accelerating universe and a glimpse of the epoch of deceleration. *Astrophys. J.*, 560:49–71, 2001.

[38] Neal Jackson. The hubble constant. *Living Reviews in Relativity*, 10, 2007.

[39] Ray d'Inverno. *Introducing Einstein's Relativity.* Oxford University Press, 2002.

[40] Charles M. Misner, Kip S. Thorne, and John A. Wheeler. *Gravitation.* Freeman, 2000.

[41] Eric Poisson. *A Relativistic's Toolkit, The Mathematics of Black-Hole Mechanics.* Cambridge University Press, 2004.

[42] Robert M. Wald. *General Relativity.* The University of Chicago Press, 1984.

[43] Jens Kujat, Angela M. Linn, Robert J. Scherrer, and David H. Weinberg. Prospects for determining the equation of state of the dark energy: What can be learned from multiple observables? *Astrophys. J.*, 572:1–14, 2002.

[44] Irit Maor, Ram Brustein, and Paul J. Steinhardt. Limitations in using luminosity distance to determine the equation-of-state of the universe. *Phys. Rev. Lett.*, 86:6, 2001.

[45] Andrew R. Liddle and David H. Lyth. *Cosmological Inflation and Large-Scale Structure.* Cambridge University Press, 2000.

[46] Terry P. Walker, Gary Steigman, David N. Schramm, Keith A. Olive, and Ho-Shik Kang. Primordial nucleosynthesis redux. *Astrophys. J.*, 376:51–69, 1991.

[47] David N. Schramm and Michael S. Turner. Big-bang nucleosynthesis enters the precision era. *Rev. Mod. Phys.*, 70:303–318, 1998.

[48] Alan H. Guth. The inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev.*, D23:347–356, 1981.

[49] Andrei D. Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Phys. Lett.*, B108:389–393, 1982.

[50] Martin Lemoine, Jérome Martin, and Patrick Peter. *Inflationary Cosmology.* Springer, 2007.

[51] Gerard Jungman, Marc Kamionkowski, Arthur Kosowsky, and David N. Spergel. Cosmological parameter determination with microwave background maps. *Phys. Rev.*, D54:1332–1344, 1996.

[52] W. J. G. de Blok and A. Bosma. H alpha rotation curves of low surface brightness galaxies. *Astron. Astrophys.*, 385:816, 2002.

[53] Mordehai Milgrom. Mond–a pedagogical review. *Acta Phys. Polon.*, B32:3613, 2001.

[54] Douglas Clowe et al. A direct empirical proof of the existence of dark matter. *Astrophys. J.*, 648:L109–L113, 2006.

[55] Marusa Bradac et al. Strong and weak lensing united iii: Measuring the mass distribution of the merging galaxy cluster 1e0657-56. *Astrophys. J.*, 652:937–947, 2006.

[56] J. R. Brownstein and J. W. Moffat. The bullet cluster 1e0657-558 evidence shows modified gravity in the absence of dark matter. 2007.

[57] Constantinos Skordis, D. F. Mota, P. G. Ferreira, and C. Boehm. Large scale structure in bekenstein's theory of relativistic mond. *Phys. Rev. Lett.*, 96:011301, 2006.

[58] Anze Slosar, Alessandro Melchiorri, and Joseph Silk. Did boomerang hit mond? *Phys. Rev.*, D72:101301, 2005.

[59] Katherine Freese and Matthew Lewis. Cardassian expansion: a model in which the universe is flat, matter dominated, and accelerating. *Phys. Lett.*, B540:1–8, 2002.

[60] Katherine Freese and William H. Kinney. The ultimate fate of life in an accelerating universe. *Phys. Lett.*, B558:1–8, 2003.

[61] J. C. Fabris, S. V. B. Gonçalves, and R. de Sá Ribeiro. Generalized Chaplygin gas with $\alpha = 0$ and the ΛCDM cosmological model. *Gen. Rel. Grav.*, 36:211–216, 2004.

[62] P. P. Avelino, L. M. G. Beça, J. P. M. de Carvalho, and C. J. A. P. Martins. The ΛCDM limit of the generalized Chaplygin gas scenario. *JCAP*, 0309:002, 2003.

[63] Martin Kunz. The dark degeneracy: On the number and nature of dark components. 2007.

[64] M. C. Bento, O. Bertolami, and Anjan Ananda Sen. The revival of the unified dark energy - dark matter model? *Phys. Rev.*, D70:083519, 2004.

[65] P. P. Avelino, L. M. G. Beça, J. P. M. de Carvalho, C. J. A. P. Martins, and P. Pinto. Alternatives to quintessence model-building. *Phys. Rev.*, D67:023511, 2003.

[66] J. C. Fabris, S. V. B. Gonçalves, and P. E. de Souza. Fitting the supernova type Ia data with the Chaplygin gas. 2002.

[67] M. Makler, S. Quinet de Oliveira, and Ioav Waga. Constraints on the generalized Chaplygin gas from supernovae observations. *Phys. Lett.*, B555:1, 2003.

[68] J. S. Alcaniz, Deepak Jain, and Abha Dev. High-redshift objects and the generalized Chaplygin gas. *Phys. Rev.*, D67:043514, 2003.

[69] P. T. Silva and O. Bertolami. Expected constraints on the generalized Chaplygin equation of state from future supernova experiments and gravitational lensing statistics. *Astrophys. J.*, 599:829–838, 2003.

[70] R. Colistete, J. C. Fabris, S. V. B. Gonçalves, and P. E. de Souza. Bayesian analysis of the Chaplygin gas and cosmological constant models using the SNe Ia data. *Int. J. Mod. Phys.*, D13:669–694, 2004.

[71] Rachel Bean and Olivier Doré. Are Chaplygin gases serious contenders to the dark energy throne? *Phys. Rev.*, D68:023515, 2003.

[72] M. C. Bento, O. Bertolami, N. M. C. Santos, and A. A. Sen. Supernovae constraints on models of dark energy revisited. *Phys. Rev.*, D71:063501, 2005.

[73] O. Bertolami, Anjan Ananda Sen, S. Sen, and P. T. Silva. Latest supernova data in the framework of generalized chaplygin gas model. *Mon. Not. Roy. Astron. Soc.*, 353:329, 2004.

[74] Zong-Hong Zhu. Generalized Chaplygin gas as a unified scenario of dark matter/energy: Observational constraints. *Astron. Astrophys.*, 423:421–426, 2004.

[75] R. Colistete and J. C. Fabris. Bayesian analysis of the (generalized) Chaplygin gas and cosmological constant models using the 157 gold SNe Ia data. *Class. Quant. Grav.*, 22:2813–2834, 2005.

[76] Yun Wang. Flux-averaging analysis of type Ia supernova data. *Astrophys. J.*, 536:531, 2000.

[77] Tamara M. Davis et al. Scrutinizing exotic cosmological models using ESSENCE supernova data combined with other cosmological probes. *Astrophys. J.*, 666:716, 2007.

[78] Jesper Sollerman et al. Supernova cosmology and the essence project. *ESA Spec. Publ.*, 637:14.1, 2006.

[79] Pierre Astier et al. The supernova legacy survey: Measurement of Ω_M, Ω_Λ and w from the first year data set. *Astron. Astrophys.*, 447:31–48, 2006.

[80] Adam G. Riess et al. New hubble space telescope discoveries of type ia supernovae at $z > 1$: Narrowing constraints on the early behavior of dark energy. 2006.

[81] William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, third edition edition, 2007.

[82] Licia Verde. A practical guide to Basic Statistical Techniques for Data Analysis in Cosmology. 2007.

[83] P. P. Avelino, C. J. A. P. Martins, and P. Pinto. Modified median statistics and type Ia supernova data. *Astrophys. J.*, 575:989–995, 2002.

[84] Will J. Percival et al. The 2dF galaxy redshift survey: The power spectrum and the matter content of the universe. *Mon. Not. Roy. Astron. Soc.*, 327:1297, 2001.

[85] Rachel Bean and Alessandro Melchiorri. Current constraints on the dark energy equation of state. *Phys. Rev.*, D65:041302, 2002.

[86] Håvard Sandvik, Max Tegmark, Matias Zaldarriaga, and Ioav Waga. The end of unified dark matter? *Phys. Rev.*, D69:123524, 2004.

[87] Daniela Carturan and Fabio Finelli. Cosmological effects of a class of fluid dark energy models. *Phys. Rev.*, D68:103501, 2003.

[88] Wayne Hu. Structure formation with generalized dark matter. *Astrophys. J.*, 506:485–494, 1998.

[89] Shoba Veeraraghavan and Albert Stebbins. Causal compensated perturbations in cosmology. *Astrophys. J.*, 365:37–65, 1990. CfA-3088.

[90] Viatcheslav F. Mukhanov, H. A. Feldman, and Robert H. Brandenberger. Theory of cosmological perturbations. Part 1 Classical perturbations. Part 2 Quantum theory of perturbations. Part 3 extensions. *Phys. Rept.*, 215:203–333, 1992.

[91] Edmund Bertschinger. Cosmological dynamics: Course 1. 1993.

[92] Wayne Hu. Covariant linear perturbation formalism. 2004.

[93] Ruth Durrer. Cosmological perturbation theory. *Lect. Notes Phys.*, 653:31–70, 2004.

[94] Ian Aitchison and Anthony Hey. *Gauge Theories in Particle Physics, From Relativistic Quantum Mechanics to QED*, volume 1. Institute of Physics Publishing, 2003.

[95] John Baez and Javier P. Muniain. *Gauge Fields, Knots and Gravity.* World Scientific, 2003.

[96] Ruth Durrer. Gauge invariant cosmological perturbation theory: A General study and its application to the texture scenario of structure formation. *Fund. Cosmic Phys.*, 15:209, 1994.

[97] D. Lyth and E. Stewart. The evolution of density perturbations in the universe. *Astrophys. J.*, 361:343–353, 1990.

[98] George B. Arfken and Hans J. Weber. *Mathematical Methods for Physicists.* Elsevier Academic Press, 2005.

[99] T. Padmanabhan. *Structure Formation in the Universe.* Cambridge University Press, 1995.

[100] Andrea Gabrielli, Francesco Labini, Michael Joyce, and Luciano Pietronero. *Statistical Physics for Cosmic Structures*. Springer, 2005.

[101] James M. Bardeen, J. R. Bond, Nick Kaiser, and A. S. Szalay. The statistics of peaks of gaussian random fields. *Astrophys. J.*, 304:15–61, 1986.

[102] Scott Dodelson et al. The three-dimensional power spectrum from angular clustering of galaxies in early SDSS data. *Astrophys. J.*, 572:140–156, 2001.

[103] Naoshi Sugiyama. Cosmic background anistropies in CDM cosmology. *Astrophys. J. Suppl.*, 100:281, 1995.

[104] Uros Seljak and Matias Zaldarriaga. A line of sight approach to cosmic microwave background anisotropies. *Astrophys. J.*, 469:437–444, 1996.

[105] Michael Doran. Cmbeasy:: an object oriented code for the cosmic microwave background. *JCAP*, 0510:011, 2005.

[106] Max Tegmark, Andrew J. S. Hamilton, and Yong-Zhong Xu. The power spectrum of galaxies in the 2dF 100k redshift survey. *Mon. Not. Roy. Astron. Soc.*, 335:887–908, 2002.

[107] M. C. Bento, O. Bertolami, and A. A. Sen. WMAP constraints on the generalized Chaplygin gas model. *Phys. Lett.*, B575:172–180, 2003.

[108] L. Amendola, Fabio Finelli, C. Burigana, and D. Carturan. WMAP and the generalized Chaplygin gas. *JCAP*, 0307:005, 2003.

[109] L. M. G. Beça, P. P. Avelino, J. P. M. de Carvalho, and C. J. A. P. Martins. The role of baryons in unified dark matter models. *Phys. Rev.*, D67:101301, 2003.

[110] Abha Dev, Deepak Jain, and J. S. Alcaniz. Cosmological consequences of a Chaplygin gas dark energy. *Phys. Rev.*, D67:023515, 2003.

[111] M. Makler, S. Quinet de Oliveira, and Ioav Waga. Observational constraints on chaplygin quartessence: Background results. *Phys. Rev.*, D68:123521, 2003.

[112] Abha Dev, Deepak Jain, and Jailson S. Alcaniz. Constraints on Chaplygin quartessence from the CLASS gravitational lens statistics and supernova data. *Astron. Astrophys.*, 417:847–852, 2004.

[113] J. C. Fabris, S. V. B. Gonçalves, and P. E. de Souza. Density perturbations in an universe dominated by the Chaplygin gas. *Gen. Rel. Grav.*, 34:53–63, 2002.

[114] J. C. Fabris, S. V. B. Gonçalves, and P. E. De Souza. Mass power spectrum in a universe dominated by the Chaplygin gas. *Gen. Rel. Grav.*, 34:2111–2126, 2002.

[115] P. P. Avelino, L. M. G. Beça, J. P. M. de Carvalho, C. J. A. P. Martins, and E. J. Copeland. The onset of the non-linear regime in unified dark matter models. *Phys. Rev.*, D69:041301, 2004.

[116] Hans Stephani, Dietrich Krames, Malcolm MacCallum, Cornelius Hoenselaers, and Eduard Herlt. *Exact Solutions of Einstein's Field Equations*. Cambridge University Press, 2003.

[117] George F. R. Ellis and Thomas Buchert. The universe seen at different scales. *Phys. Lett.*, A347:38–46, 2005.

[118] John D. Barrow and Christos G. Tsagas. Averaging anisotropic cosmologies. *Class. Quant. Grav.*, 24:1023–1032, 2007.

[119] Thomas Buchert. Dark energy from structure - a status report. 2007.

[120] A. Gruzinov, M. Kleban, M. Porrati, and Michele Redi. Gravitational back-reaction of matter inhomogeneities. *JCAP*, 0612:001, 2006.

[121] Nan Li and Dominik J. Schwarz. Signatures of cosmological backreaction. 2007.

[122] W. H. Press and P. Schechter. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *Astrophys. J.*, 187:425, 1974.

[123] Varun Sahni and Peter Coles. Approximation methods for nonlinear gravitational clustering. *Phys. Rept.*, 262:1–135, 1995.

[124] Justin Khoury and Amanda Weltman. Chameleon cosmology. *Phys. Rev.*, D69:044026, 2004.

[125] Philippe Brax, Carsten van de Bruck, Anne-Christine Davis, Justin Khoury, and Amanda Weltman. Detecting dark energy in orbit: The cosmological chameleon. *Phys. Rev.*, D70:123518, 2004.

[126] Philippe Brax, Carsten van de Bruck, Anne-Christine Davis, and Anne M. Green. Small Scale Structure Formation in Chameleon Cosmology. *Phys. Lett.*, B633:441–446, 2006.

[127] David F. Mota and Douglas J. Shaw. Strongly coupled chameleon fields: New horizons in scalar field theory. *Phys. Rev. Lett.*, 97:151102, 2006.

[128] Rachel Bean, Eanna E. Flanagan, and Mark Trodden. Adiabatic instability in coupled dark energy-dark matter models. 2007.

[129] Rachel Bean, Eanna E. Flanagan, and Mark Trodden. The Adiabatic Instability on Cosmology's Dark Side. 2007.

[130] Niayesh Afshordi, Matias Zaldarriaga, and Kazunori Kohri. On the stability of dark energy with mass-varying neutrinos. *Phys. Rev.*, D72:065024, 2005.

[131] Manoj Kaplinghat and Arvind Rajaraman. Stable Models of superacceleration. *Phys. Rev.*, D75:103504, 2007.

[132] Ole Eggers Bjaelde et al. Neutrino Dark Energy – Revisiting the Stability Issue. 2007.